# DSFC: Deformation-Aware Learning Strategy via Self-sustaining Feedback Cycle for Medical Vision Foundation Model Domain Adaptation

Jie Lin[1], Hengyi Jiang[1], Hong Liu[1], and Liansheng Wang[1(✉)]

School of Informatics, Xiamen University
{jaylin, liuhong}@stu.xmu.edu.cn, harryj156@163.com, lswang@xmu.edu.cn

**Abstract.** Vision foundation model, despite strong segmentation capabilities enabled by pretraining on large-scale data, remain underexplored in specific medical visual concept segmentation tasks. Medical imaging presents unique challenges: pixel intensity differences between target regions and surrounding structures are often subtle, and significant variations in the shape, size, and location of anatomical structures limit the effectiveness of traditional pixel-similarity-based alignment strategies. This paper proposes a Deformation-Aware Learning Strategy via Self-sustaining Feedback Cycle (DSFC) for medical image segmentation. The framework introduces a dual-deformation perturbation mechanism, combining global gaussian-distributed deformations and target-focused local deformations, to preserve anatomical patterns while capturing non-rigid variations. Hard Example Adaptive (HEA) loss is proposed to enhance training stability and mask accuracy. DSFC establishes a closed-loop training process, alternately optimizing the segmentation model and destroyer to improve anatomical understanding. Our extensive experiments on public datasets with various dimensions, organs demonstrate that DSFC significantly enhances model performance in fully supervised training settings without the need for increasing the samples. and its components are effective. Our code is available at: https://github.com/jaylinio/DSFC.

**Keywords:** Medical image segmentation · SAM · Deformation aware augmentation · Foundation models adaptation.

## 1 Introduction

Pretrained vision foundational models [7, 16, 19, 20, 1] have demonstrated immense potential in medical image segmentation due to their powerful zero-shot segmentation capabilities. Among them, SAM series [7, 16] excels by achieving target segmentation without requiring task-specific pretraining, leveraging user-provided visual prompts such as points, boxes, or masks. This capability is particularly advantageous in data-scarce medical scenarios. However, when applied to medical imaging, SAM may yield suboptimal results [24].

Jie Lin and Hengyi Jiang——Contributed equally


**Fig. 1. SAM demonstrates incomplete segmentation even when provided with correct prompts.** From left to right are the original image, the prompt initialized from GT, zero-shot, one-shot, few-shot, and fully supervised segmentations.

To address the challenges in medical image segmentation, numerous studies have proposed improvements to SAM, aiming to enhance its segmentation performance in medical image tasks through few-shot auto-prompting [21, 23, 4, 22], fine-tuning [25, 18, 11, 27], and training framework modification [26, 9, 2]. Although these methods have shown advantages in reducing user intervention and the number of training samples, as shown in Figure 1, segmentation results still have deficiencies [12], especially in tasks that require high precision, such as tumor resectability assessment [13] and neoadjuvant therapy evaluation [6, 14].

To this end, this paper proposes a learning framework called **D**eformation-Aware Learning Strategy via **S**elf-sustaining **F**eedback **C**ycle (DSFC), which introduces a perturbation process during adaptation of the foundation model (e.g., MedSAM-2 [27]). This process generates soft global and target-focused local deformation disturbance signals, the former is sampled with a gaussian distribution sampled from foreground organs in image, while the latter is generated by our proposed self-sustaining local destroyer model. Those deformations have high global and local diversity, and follow the original distribution pattern (i.e., the size, shape, related location of organs) of the segmented targets, encourage model to capture the geometric non-rigid deformation differences among different patients. In addition, to address the challenge of learning from difficult pixels, we propose Hard Example Adaptive (HEA) loss function to further stabilize the training process and improve the completeness and accuracy of the mask prediction. This forms a closed-loop training process. The refiner (i.e., the segmentation model) and destoryer are optimized alternatley during training, As the training gradually converges, the base model is able to more accurately understand the spatial geometric features of real anatomical structures, resulting in denser masks that are biologically plausible. Experimental results on three medical datasets demonstrate the superior performance of DSFC in medical image segmentation tasks, showcasing excellent adaptability and generalizability. Our core contributions can be summarized as follows:

1. We introduce a deformation-aware destroy process to model geometric deformation differences between patients, enhancing the biological plausibility of medical image segmentation.
2. We propose to optimize the DSFC framework with a cyclic training strategy, significantly improving adaptation of foundation model (i.e., MedSAM-2) without additional samples.

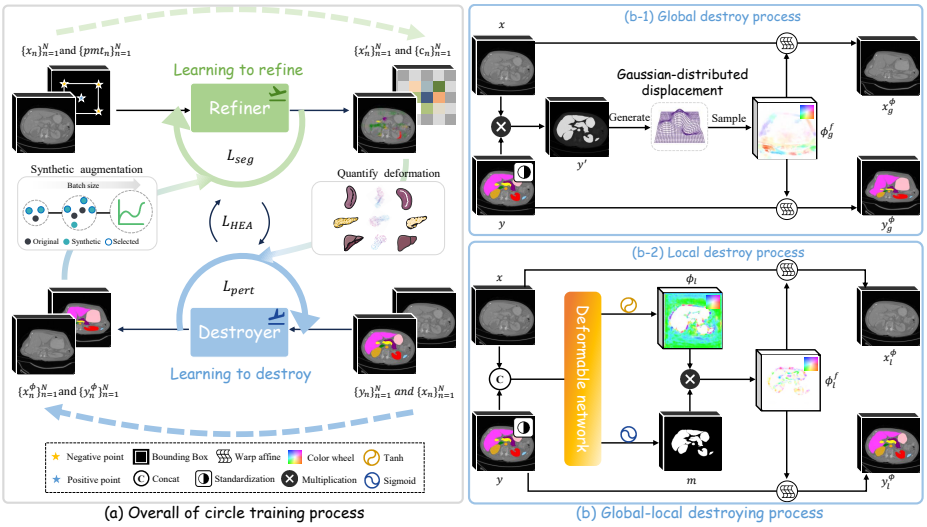


**Fig. 2. Overview of the proposed DSFC pipeline.**

3. We extensively evaluate our method on both 2D and 3D public datasets. The results demonstrate that our method outperforms state-of-the-art (SOTA) techniques and highlights the effectiveness of its components.

## 2 Method

The DSFC framework consists of a refiner (e.g., MedSAM-2 [27]), a destroyer and an alternating training mechanism. The refiner processes images $\{x_n\}_{n=1}^N$ and prompts $\{pmt_n\}_{n=1}^N$ to generate coarse predictions $\{x'_n\}_{n=1}^N$ and confidence matrices $\{c_n\}_{n=1}^N$. These are used by the destroyer to compute spatial discrepancies with ground truths $\{y_n\}_{n=1}^N$, driving it to learn a dense deformation field $\phi$ for self-supervised backpropagation. The destroyer then perturbs the images and ground truths using $\phi$, creating augmented data $\{x_n^\phi\}_{n=1}^N$ and $\{y_n^\phi\}_{n=1}^N$ for the refiner. This iterative process is stabilized by Hard Example Adaptive (HEA) loss, forming an end-to-end cyclic training loop.

**Refiner** The proposed DSFC incorporates MedSAM-2 [27] as the Refiner, an extended version of SAM2 [16] specifically designed for medical imaging tasks. The architecture consists of several key modules that work together to achieve precise segmentation of medical images. First, the Image Encoder ($E_{\text{img}}$) encodes each frame of the medical image $x_t$ into a feature embedding $f_t = E_{\text{img}}(x_t)$, providing the foundational features for subsequent processing. Simultaneously, the Prompt Encoder ($E_{\text{pmt}}$) processes the user-provided prompt $pmt_t$ to generate a corresponding embedding $q_t = E_{\text{pmt}}(pmt_t)$, which guides the execution of the segmentation task. The Self-Sorting Memory Bank ($\mathcal{M}_t^{\text{sort}}$) dynamically stores

and updates feature embeddings from previous time steps, ensuring high information content and diversity through an update mechanism based on confidence and dissimilarity. Next, the Memory Attention Mechanism ($A$) combines the current frame's feature embedding $f_t$, the updated memory bank $\tilde{\mathcal{M}}_t^{\text{sort}}$, and the prompt embedding $q_t$ to generate weighted information through attention computation. Finally, the Mask Decoder ($D$) predicts the segmentation mask $y_t$ based on the attention output and produces a segmentation confidence matrix $c_t = D_{\text{conf}}(A(f_t, \tilde{\mathcal{M}}_t^{\text{sort}}, q_1))$, where $c_t$ represents the category-specific confidence values of the segmentation mask, providing a quantitative measure of the reliability of the segmentation results. The segmentation process is mathematically expressed as Eq. 1 follows:

$$\begin{aligned} y_t &= D(A(f_t, \tilde{\mathcal{M}}_t^{\text{sort}}, q_1)), \quad t = 1, \ldots, T \\ c_t &= D_{\text{conf}}(A(f_t, \tilde{\mathcal{M}}_t^{\text{sort}}, q_1)), \quad t = 1, \ldots, T \end{aligned} \tag{1}$$

where $f_t = E_{\text{img}}(x_t)$, $q_1 = E_{\text{pmt}}(p_1)$, and $\tilde{\mathcal{M}}_t^{\text{sort}}$ is the resampled feature embedding from the self-sorting memory bank $\mathcal{M}_t^{\text{sort}}$. The segmentation confidence matrix $c_t$ is also produced by the mask decoder, which provides a category-wise confidence value for each predicted segmentation.

**Destroyer** The Destroyer is composed by a global destroy process and a local destroy process, the former is sampled with a gaussian distribution sampled from foreground organs in image, while the latter is generated by our proposed self-sustaining local destroyer model.

*Global destroy process.* The goal of global destroy process is to generate a globally random spatial transform. Specifically, we firstly initial $\phi_g$ by assigning its each position $(x, y)$ a two-dimensional vector $\phi_g(x, y) = (d_x, d_y)$, representing the displacement at that position. Each component $d_x$ and $d_y$ of the deformation field is randomly sampled from a gaussian distribution, as shown in Eq. 2.

$$d_x \sim \text{Norm}(0, \sigma_D^2), \quad d_y \sim \text{Norm}(0, \sigma_D^2) \tag{2}$$

where $\sigma_D$ denotes the standard deviation of the gaussian distribution, controlling the intensity of the deformation field. Notably, to preserve the organ's spatial consistency, we compute $\sigma_D = (\sigma_x + \sigma_y)/2$ from the mask region $y'$.

Then, a gaussian filtering $G_\sigma$ is adopt to ensure the smoothness of $\phi_g$. Additionally, to limit the magnitude of the deformation field, the smoothed deformation field is clipped to a predefined range $[-a, a]$, producing the final deformation field $\phi_g^f$, as described in Eq. 3.

$$\phi_g^f = \text{Clip}(\phi_g * G_\sigma, -a, a) \tag{3}$$

where Clip denotes the clipping function, and $a$ is the predefined range. Finally, the deformation field $\phi_g^f$ is applied to the original image $x$ and the ground truth label $y$, yielding the distorted image $x_g^\phi$ and the distorted label $y_g^\phi$.

*Local destroy process.* The global destroy process provide a globally destroyed version of the input data, while the lack of locally variation limiting the performance of the model, which is difficult to simulate by random deformation field. In this section, we propose to train a local destroyer model, a lightweight U-shape network to output a local focused deformation field. Since the original segmentation results usually differ from the annotations only in local areas but are generally consistent on a global scale (i.e., the size, shape, and relative positions of organs), we can consider the original segmentation results as new images that are locally deformed versions of the labels. Thus, we proposed to guide the local destroyer to learn the deformation between the original segmentation and the label. Specifically, given the ground truth $y^s$ (with standardization) and the image $x$ as input, the model output a deformation field $\phi_l$ and a alpha matte $m$. The final deformation field $\phi_l^f$ is obtained by multiply the $m$ and $\phi_l$. Following [8], we use a binary cross entropy loss to supervise the learning of $m$, formally:

$$\mathcal{L}_{pert} = -\frac{1}{N}\sum_{i=1}^{N}[y_i^s \log(m_i) + (1 - y_i^s)\log(1 - m_i)] \tag{4}$$

where $N$ represents the number of pixels, $y_i$ is the value of the $i$-th pixel, and $m_i$ is the value of the $i$-th pixel, thus enabling self-supervised learning for the deformation field $\phi_l$. Its gradient jointly optimizes the deformation field, enabling unsupervised alignment.

**Training Strategy** The DSFC method enhances training stability through online augmentation by doubling the size of each batch, which includes both globally and locally augmented data. During refiner training, the augmentation probability $p$ is applied per sample within a batch, yielding $\lfloor N \cdot p \rfloor$ augmented samples from $N$ total augmentations. The destroyer is thus optimized on a combined set of original and augmented samples. During training, the destroyer and refiner are alternately updated in odd and even iterations respectively, while keeping the other module frozen.

We employ Hard Example Adaptive (HEA) loss helps the model focus on the hard-to-classify pixels within each mini-batch during training. Hard examples are pixels where the conditional probability $P(c \mid v)$ of class $c$ given pixel value $v$ is below a threshold $\tau$, i.e., $P(c \mid v) < \tau$. HEA loss further restrict pixels within the ground truth foreground region, formally as Eq. 5:

$$\mathcal{L}_{HEA} = \sum_{v=1}^{V}\sum_{c=1}^{C} \mathbf{y^s} \cdot \mathbb{I}\{P(c \mid v) < \tau\} \cdot \log P(c \mid v), \tag{5}$$

where $C$ is the total number of classes, $V$ is the total number of pixels in a mini-batch, $\tau \in (0, 1]$ is the confidence threshold and $\mathbb{I}\{\cdot\}$ is an indicator function that returns 1 if the condition is true and 0 otherwise. In this paper, we set $\tau = 0.5$ and $M = \max(y^s > 0)$, where $M$ corresponds to the minimum number of hard-to-classify pixels used in each mini-batch. The Destroyer $D$ is updated using the gradient $-\Delta_D(\mathcal{L}_{pert})$, and the Refiner $R$ is updated using the gradient $-\Delta_R(\mathcal{L}_{seg})$ and $-\Delta_R(\mathcal{L}_{HEA})$, $\mathcal{L}_{seg}$ follows the same approach as used in MedSAM2 [27].

The detailed training procedure of the DSFC method is shown in Algo. 1.

**Algorithm 1** DSFC

**Require:** Coarse predictions $\{x_n^{'}\}_{n=1}^{N}$, ground truths $\{y_n\}_{n=1}^{N}$.
**Ensure:** Destroyer $D$ aligns $y$ with $x^{'}$; Refiner $R$ predicts $x^{'}$.
Augment $\{y_n\}_{n=1}^{N}$ to $\{y_n^{\phi}\}_{n}^{N}$ samples via Destroyer.
Initialize $D, R$.
**for** each epoch $\leq$ maximum epoch **do**
  **for** each iteration $\leq$ maximum iterations **do**
    **if** Destroyer training is true **then**
      Sample mini-batches $y, x, x^{'}$ from $y_n, x_n, x_n^{'}$.
      Compute $\mathcal{L}_{pert}$.
      Update $D \overset{t}{\leftarrow} -\Delta_D(\mathcal{L}_{pert})$ (gradient update).
    **else**
      Load source data $(x, y, pmt)$ and augmented data $(x^{\phi}, y^{\phi}, pmt^{\phi})$.
      Compute $\mathcal{L}_{seg}$ and $\mathcal{L}_{HEA}$.
      Update $R \overset{t}{\leftarrow} -\Delta_R(\mathcal{L}_{HEA})$ (gradient update).
      Update $R \overset{t}{\leftarrow} -\Delta_R(\mathcal{L}_{seg})$ (gradient update).
    **end if**
  **end for**
**end for**

## 3 Experiments

**Datasets** We conducted multi-organ segmentation experiments on three publicly available BTCV [3], Synapse [5] and JSRT [17] dataset. The BTCV [3] consists of 24 training samples and 6 test samples, each with around 130 valid slices, and features labels for 13 abdominal organs. The Synapse [5] comprises 18 training samples and 12 test samples, with each sample containing approximately 70 valid slices, and includes labels for 8 abdominal organs. The JSRT [17] contains 197 training samples and 50 test samples, with each sample having approximately 1 valid slice, and provides labels for 1 chest organ. During the training process, the resolution for all datasets is standardized to 512 $\times$ 512.

**Implementation Details** The experiments are conducted with PyTorch [15] using 8 NVIDIA-A800 GPUs. The optimizer employed was AdamW [10] ($\beta_1 = 0.9, \beta_2 = 0.999$), with a learning rate of 1e-4. We set the warm-up ratio to 0.001 and use the cosine decay schedule after warm-up. The weight decay was set to 0.1. Following [27], we evaluate the model's performance using task-specific prompts, with the prompt types including point, box, and mask.

**Comparison with SOTA SAM-based Frameworks** The experimental results demonstrate the superior performance of our proposed method (both 2D and

**Table 1. Quantitative Comparison of Medical Images Segmentation Performance**. We show the comparison of DSFC with SAM-based methods over BTCV [3], Synapse [5] and JSRT [17] evaluated by dice score.

| Method | BTCV | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Spleen | R.Kid | L.Kid | Gall. | Eso. | Liver | Stom. | Aorta | IVC | Veins | Panc. | Adre. | Ave |
| SAM[7] | 0.368 | 0.522 | 0.621 | 0.116 | 0.156 | 0.446 | 0.401 | 0.589 | 0.462 | 0.137 | 0.165 | 0.158 | 0.345 |
| SAM2[16] | 0.517 | 0.621 | 0.669 | 0.224 | 0.338 | 0.615 | 0.593 | 0.647 | 0.489 | 0.221 | 0.135 | 0.132 | 0.433 |
| SAMed[25] | 0.862 | 0.71 | 0.798 | 0.677 | 0.735 | 0.944 | 0.766 | 0.874 | 0.798 | 0.775 | 0.579 | 0.79 | 0.776 |
| SAM-Med3D[18] | 0.873 | 0.884 | 0.932 | 0.795 | 0.79 | 0.943 | 0.889 | 0.872 | 0.796 | 0.813 | 0.779 | 0.797 | 0.847 |
| BLO-SAM[26] | 0.527 | 0.661 | 0.775 | 0.614 | 0.39 | 0.51 | 0.429 | 0.508 | 0.335 | 0.352 | 0.464 | 0.59 | 0.513 |
| SAMUS[9] | 0.868 | 0.776 | 0.834 | 0.69 | 0.71 | 0.922 | 0.805 | 0.863 | 0.844 | 0.782 | 0.611 | 0.78 | 0.79 |
| One-Prompt[21] | 0.801 | 0.789 | 0.814 | 0.816 | 0.818 | 0.791 | 0.808 | 0.737 | 0.729 | 0.75 | 0.813 | 0.77 | 0.786 |
| FSSP-SAM[22] | 0.862 | 0.909 | 0.893 | 0.691 | 0.551 | 0.822 | 0.588 | 0.873 | 0.779 | 0.517 | 0.456 | 0.456 | 0.699 |
| MedSAM[11] | 0.722 | 0.81 | 0.835 | 0.746 | 0.701 | 0.851 | 0.805 | 0.812 | 0.723 | 0.751 | 0.74 | 0.713 | 0.767 |
| MedSAM-2[27] | 0.918 | 0.951 | 0.954 | 0.92 | 0.923 | 0.945 | 0.909 | 0.919 | 0.859 | 0.875 | 0.771 | 0.86 | 0.9 |
| Ours 2D | 0.948 | 0.937 | 0.94 | 0.846 | 0.885 | 0.952 | 0.925 | **0.934** | 0.926 | 0.684 | 0.782 | 0.721 | 0.873 |
| Ours 3D | **0.951** | **0.965** | **0.965** | **0.955** | **0.928** | **0.957** | **0.941** | 0.911 | **0.93** | **0.855** | **0.844** | **0.927** | **0.927** |

| Method | Synapse | | | | | | | | | JSRT | **AVE** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Spleen | R.Kid | L.Kid | Gall. | Liver | Stom. | Aorta | Panc. | Ave | Ave | |
| SAM[7] | 0.173 | 0.452 | 0.265 | 0.231 | 0.32 | 0.53 | 0.372 | 0.506 | 0.359 | 0.221 | 0.308 |
| SAM2[16] | 0.596 | 0.433 | 0.428 | 0.256 | 0.332 | 0.553 | 0.445 | 0.495 | 0.441 | 0.463 | 0.446 |
| SAMed[25] | 0.871 | 0.862 | 0.894 | 0.344 | 0.938 | 0.771 | 0.83 | 0.429 | 0.742 | 0.809 | 0.776 |
| SAM-Med3D[18] | 0.895 | 0.901 | 0.915 | 0.491 | **0.952** | 0.788 | 0.868 | 0.542 | 0.794 | 0.845 | 0.829 |
| BLO-SAM[26] | 0.686 | 0.861 | 0.661 | 0.533 | 0.7162 | 0.536 | 0.623 | 0.44 | 0.632 | 0.856 | 0.667 |
| SAMUS[9] | 0.718 | 0.762 | 0.76 | 0.705 | 0.63 | 0.724 | 0.741 | 0.701 | 0.718 | 0.877 | 0.828 |
| One-Prompt[21] | 0.752 | 0.727 | 0.729 | 0.562 | 0.557 | 0.769 | 0.689 | 0.575 | 0.67 | 0.614 | 0.69 |
| FSSP-SAM[22] | 0.728 | 0.449 | 0.483 | 0.199 | 0.728 | 0.299 | 0.146 | 0.113 | 0.393 | 0.367 | 0.486 |
| MedSAM[11] | 0.799 | 0.813 | 0.82 | 0.351 | 0.924 | 0.522 | 0.731 | 0.294 | 0.657 | 0.763 | 0.729 |
| MedSAM-2[27] | 0.886 | 0.861 | 0.861 | 0.733 | 0.816 | 0.836 | 0.923 | 0.894 | 0.851 | 0.845 | 0.865 |
| Ours 2D | 0.895 | 0.901 | 0.915 | 0.791 | **0.952** | 0.788 | 0.868 | 0.842 | 0.869 | 0.968 | 0.903 |
| Ours 3D | **0.944** | **0.925** | **0.932** | **0.877** | 0.937 | **0.946** | **0.931** | **0.945** | **0.929** | **0.971** | **0.942** |

3D variants) compared to state-of-the-art (SOTA) models across three medical imaging datasets: BTCV [3], Synapse [5] and JSRT [17]. On BTCV [3], our method achieves the highest dice scores for 10 out of 12 anatomical structures, including significant improvements in challenging regions such as the gallbladder (95.5% vs. 92.0% from MedSAM-2) and pancreas (84.4% vs. 77.1% from SAM-Med3D). On Synapse [5], our method outperforms all competitors with an average dice of 92.9%, showcasing robustness in multi-organ segmentation. For JSRT [17], our method achieved the best segmentation performance with a dice score of 97.1%, surpassing SAM-based models. Overall, the proposed method attains the highest average dice (94.5%) across all datasets, highlighting its generalization capability and effectiveness in both 2D and 3D medical image segmentation tasks. We present a qualitative comparison of different segmentation methods on BTCV [3] in Fig. 3. The comparative trends are similar to Table. 1.

**Effect of Sub-Module** To validate the efficacy of the proposed modules in this paper, we conducted a comprehensive ablation study on the BTCV [3], Synapse [5], and JSRT [17], comparing the performance of several variants of our model in Fig. 4 (left). Since the trends in the ablation experiments were similar across the three datasets, we selected the BTCV [3] for detailed

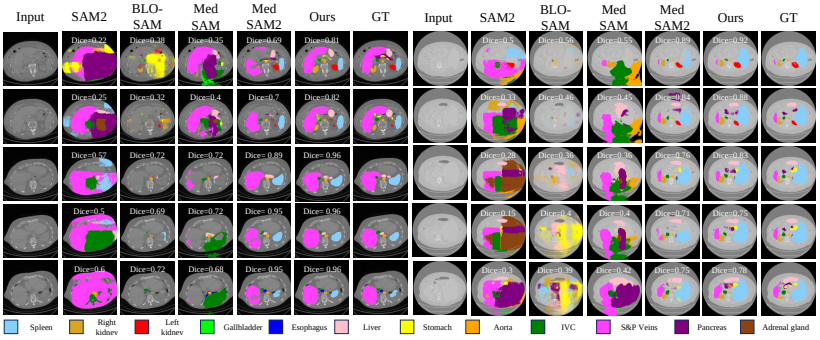


**Fig. 3. Qualitative Visualisation on Medical Image Segmentation.** Comparison of SAM-based methods, our DSFC and GT on the BTCV [3].

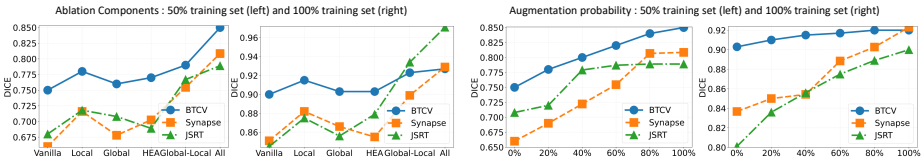


**Fig. 4. Ablation Study of DSFC.** We evaluate each of components of DSFC and each of hyper-parameters on the BTCV [3], Synapse [5] and JSRT [17].

textual illustration. First, the baseline result, which is the original segmentation output of Vanilla (MedSAM [27]), exhibited the worst performance, highlighting its insufficient capability on specific medical image datasets. Second, all three variants incorporating our proposed modules (Local-deform, Global-deform, and HEA loss) showed considerable improvements compared to the baseline, thereby validating their effectiveness. Third, the combination of Local-deform and Global-deform further enhanced the model's performance, suggesting that they are compatible and effective in augmenting data diversity in global and local image patterns, respectively. Finally, our proposed model, which integrates all three modules, achieved the best performance. These results collectively demonstrate the effectiveness of each submodule and the importance of their collaborative interaction.

**Effect of Hyper-Parameter** We investigate the impact of $p$ (augmentation probability) in Fig. 4 (right). As the probability of enhancement increases, the results of the network on the test sets of all datasets also show a corresponding improvement. Eventually we select $p = 1$ for performance evaluation and comparison with other methods.

## 4 Conclusion

This paper presents a Deformation-Aware Learning Strategy via Self-sustaining Feedback Cycle (DSFC) for domain adaptation of medical vision foundation models. By introducing global and local deformation perturbation strategies, DSFC generates samples that conform to real anatomical rules, thereby aiding the model in learning complex and varied organ patterns. Additionally, we propose Hard Example Adaptive (HEA) loss function to enhance training stability with limited samples. Experiments on various public datasets demonstrate that DSFC significantly improves model performance in multi training settings without the need for additional samples.

**Acknowledgments.** This work was supported by National Natural Science Foundation of China (Grant No. 62371409) and Fujian Provincial Natural Science Foundation of China (Grant No. 2023J01005).

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

## References

1. Bar, A., Gandelsman, Y., Darrell, T., Globerson, A., Efros, A.: Visual prompting via image inpainting. Advances in Neural Information Processing Systems **35**, 25005–25017 (2022)
2. Chai, S., Jain, R.K., Teng, S., Liu, J., Li, Y., Tateyama, T., Chen, Y.w.: Ladder fine-tuning approach for sam integrating complementary network. Procedia Computer Science **246**, 4951–4958 (2024)
3. Fang, X., Yan, P.: Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction. IEEE Transactions on Medical Imaging **39**(11), 3619–3629 (2020)
4. He, X., Hu, Y., Zhou, Z., Jarraya, M., Liu, F.: Few-shot adaptation of training-free foundation model for 3d medical image segmentation. arXiv preprint arXiv:2501.09138 (2025)
5. Igelsias, J., Styner, M., Langerak, T., Landman, B., Xu, Z., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge (2015)
6. Jeon, S.K., Lee, J.M., Lee, E.S., Yu, M.H., Joo, I., Yoon, J.H., Jang, J.Y., Lee, K.B., Lee, S.H.: How to approach pancreatic cancer after neoadjuvant treatment: assessment of resectability using multidetector ct and tumor markers. European Radiology **32**(1), 56–66 (2022)
7. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
8. Lin, J., Wu, D., Huang, L.: Self-supervised bi-directional mapping generative adversarial network for arbitrary-time longitudinal interpolation of missing data. Biomedical Signal Processing and Control **105**, 107514 (2025)

9. Lin, X., Xiang, Y., Zhang, L., Yang, X., Yan, Z., Yu, L.: Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation. arXiv preprint arXiv:2309.06824 (2023)
10. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
11. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
12. Magg, C., Kervadec, H., Sánchez, C.I.: Zero-shot capability of sam-family models for bone segmentation in ct scans (2024), https://arxiv.org/abs/2411.08629
13. Mahmoudi, T., Kouzahkanan, Z.M., Radmard, A.R., Kafieh, R., Salehnia, A., Davarpanah, A.H., Arabalibeik, H., Ahmadian, A.: Segmentation of pancreatic ductal adenocarcinoma (pdac) and surrounding vessels in ct images using deep convolutional neural networks and texture descriptors. Scientific Reports **12**(1), 3092 (2022)
14. Miranda, J., Causa Andrieu, P., Nincevic, J., Gomes de Farias, L.d.P., Khasawneh, H., Arita, Y., Stanietzky, N., Fernandes, M.C., De Castria, T.B., Horvat, N.: Advances in mri-based assessment of rectal cancer post-neoadjuvant therapy: a comprehensive review. Journal of Clinical Medicine **13**(1), 172 (2023)
15. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. arXiv preprint arXiv:1912.01703 (2019)
16. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
17. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.i., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. American journal of roentgenology **174**(1), 71–74 (2000)
18. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., et al.: Sam-med3d: towards general-purpose segmentation models for volumetric medical images. arXiv preprint arXiv:2310.15161 (2023)
19. Wang, X., Wang, W., Cao, Y., Shen, C., Huang, T.: Images speak in images: A generalist painter for in-context visual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6830–6839 (2023)
20. Wang, X., Zhang, X., Cao, Y., Wang, W., Shen, C., Huang, T.: Seggpt: Segmenting everything in context. arXiv preprint arXiv:2304.03284 (2023)
21. Wu, J., Xu, M.: One-prompt to segment all medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11302–11312 (2024)
22. Wu, Q., Zhang, Y., Elbatel, M.: Self-prompting large vision models for few-shot medical image segmentation. In: MICCAI workshop on domain adaptation and representation transfer. pp. 156–167. Springer (2023)
23. Xu, J., Li, X., Yue, C., Wang, Y., Guo, Y.: Sam-mpa: Applying sam to few-shot medical image segmentation using mask propagation and auto-prompting. arXiv preprint arXiv:2411.17363 (2024)
24. Yuan, C., Jiang, J., Yang, K., Wu, L., Wang, R., Meng, Z., Ping, H., Xu, Z., Zhou, Y., Song, W., et al.: Is segment anything model 2 all you need for surgery video segmentation? a systematic evaluation. arXiv preprint arXiv:2501.00525 (2024)

25. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
26. Zhang, L., Liang, Y., Zhang, R., Javadi, A., Xie, P.: Blo-sam: Bi-level optimization based finetuning of the segment anything model for overfitting-preventing semantic segmentation. In: Forty-first International Conference on Machine Learning
27. Zhu, J., Qi, Y., Wu, J.: Medical sam 2: Segment medical images as video via segment anything model 2. arXiv preprint arXiv:2408.00874 (2024)