MICCAI

SynPo: Boosting Training-Free Few-Shot Medical Segmentation via High-Quality Negative Prompts

Yufei Liu[1], Haoke Xiao[2,4], Jiaxing Chai[1], Yongcun Zhang[1], Rong Wang[2], Zijie Meng[3], and Zhiming Luo[1,2(✉)]

[1] Department of Artificial Intelligence, Xiamen University, Xiamen, China
[2] Institute of Artificial Intelligence, Xiamen University, Xiamen, China
[3] School of Software and Microelectronics, Peking University, Beijing, China
[4] Kuaishou Technology, Beijing, China
✉Correspondence: `zhiming.luo@xmu.edu.cn`

Abstract. The advent of Large Vision Models (LVMs) offers new opportunities for few-shot medical image segmentation. However, existing training-free methods based on LVMs fail to effectively utilize negative prompts, leading to poor performance on medical images with low-contrast. To address this issue, we propose **SynPo**, a training-free few-shot method based on LVMs (e.g., SAM), with the core insight: ***improving the quality of negative prompts***. To select point prompts in a more reliable confidence map, we design a novel Confidence Map Synergy Module by combining the strengths of DINOv2 and SAM. Based on the confidence map, we select the top-k pixels as the positive points set and choose the negative points set using a Gaussian distribution, followed by independent K-means clustering for both sets. Then, these selected points are leveraged as high-quality prompts for SAM to get the segmentation results. Extensive experiments demonstrate that **SynPo** achieves performance comparable to state-of-the-art training-based few-shot methods. Project page: `https://liu-yufei.github.io/synpo-project-page/`.

Keywords: Few-Shot Segmentation · Training Free · Foundation Model.

1 Introduction

Few-shot medical image segmentation focuses on generating accurate segmentations with a limited number of annotated samples per class or structure. Recently, few-shot medical image segmentation methods can be classified into two categories: prototype-based methods [2,9,11,22,4,20] and LVMs-based methods [1,16,21], the latter including training-free and fine-tuning methods. A representative example of such LVMs is the Segment Anything Model (SAM) [6]. Whether on natural images or medical images [12,17,19,1], the process of using SAM for training-free few-shot segmentation task can be summarized in the Fig. 1 (1). With a support-query image pair and support mask as input, this process firstly extracts feature maps for support-query image pair using a strong pre-trained visual encoder (**e.g.,** DINO), then computes the 2-D confidence map
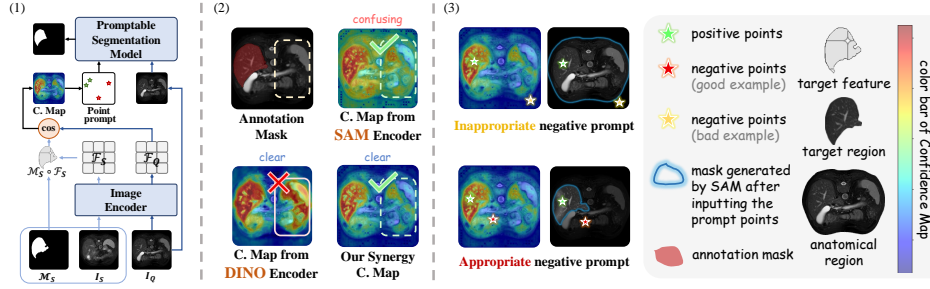
Fig. 1. Challenges in Training-Free Few-Shot Segmentation. (1) The universal process of point promptable segmentation model for training-free few-shot. (2) Different Confidence Maps (C. Map) vs. Ground Truth. In the confidence map generated by DINOv2 features, irrelevant regions on the right are mistakenly identified as "similar". The confidence map from SAM-ViT features exhibits less clear differentiation in values. Our synergy confidence map leverages the strengths of both while mitigating their respective weaknesses. (3) Pilot Experiment. Positioning negative prompts outside the anatomical region, even when using identical positive prompt locations, yields worse segmentation performance compared to placing negative prompts inside.

between query features and the target feature by cosine distance. The maximum value on the confidence map is selected as the point prompt and is used to guide SAM in performing interactive segmentation on the query image.

Although this paradigm achieves good results in the medical field, there is still room for improvement in the generation of high-quality point prompts, which is crucial for it. First, the utility of confidence maps can still be enhanced. DINOv2 is widely adopted as the feature extractor in this paradigm [1,16], excelling at extracting semantic features [18]. However, its employment of relative positional encoding [14] compromises the awareness of absolute spatial localization, potentially causing incorrect positive point selection, shown in the bottom left of Fig. 1(2). This impedes the leveraging of anatomical priors for organ positioning, which may result in segmentation's anatomical inconsistency. As for SAM-ViT, which is also widely employed, using absolute position encoding with better location awareness can effectively avoid this situation [3], as shown in the green box of the image in the top right of Fig. 1(2). Second, the selection strategy of negative prompts is coarse, resulting in segmentation performance degradation. Current methods identify pixels in the confidence map with minimal similarity to the target region as negative points, causing most negative points to cluster in background areas rather than in the anatomical region. The pilot experiment, shown in Fig. 1(3), demonstrates that this approach is counterintuitive and, therefore, leads to poor results. In medical imaging (**e.g.**, MRI/CT), such background regions are readily distinguishable from targets, rendering negative prompts redundant for SAM guidance, which leads to over-segmentation, as also noted in [13].

To address these limitations, we propose **SynPo**, a novel training-free method with Confidence Map **Syn**ergy Module and **Po**int Selection Module. First, drawing from SAM's ability to capture precise low-level spatial information [6,10], which can be a supplement of features from DINOv2, we introduce the Confidence Map Synergy Module. This module combines high-level semantic features from DINOv2 with SAM-ViT's absolute spatial cues, leveraging both to enhance anatomical structure capture and refine segmentation boundaries. Second, the Point Prompt Strategy Module, heuristically selecting negative prompts in the anatomical region, improves the selection of informative negative points, optimizing prompt guidance for segmentation and reducing redundancy prompt information. Furthermore, we incorporate a Noise-aware Refine Module that utilizes standard morphology and SAM to refine coarse masks. Experiments conducted on both CT and MRI abdominal datasets, focusing on four major organs, demonstrate that our method achieves performance on par with the state-of-the-art few-shot segmentation approaches.

2   Methodology

2.1   Overview

SynPo, as shown in Fig. 2(1), consists of three key components: Confidence Map Synergy Module (CMSM), Point Selection Module (PSM), and Noise-aware Refine Module (NRM). Given a support-query pair, first extract zero-shot visual features using pre-trained vision models (SAM-ViT and DINOv2). In CMSM, the feature maps along with the support mask $\mathcal{M}_S \in \mathbb{R}^{H \times W}$, are used to calculate the synergy map $SynMap \in \mathbb{R}^{H \times W}$ and model the negative confidence distribution P_{neg}, which are used to support the generation of prompts. In PSM, the pixels in the synergy map are sorted by their confidence score into a ranked list, which, along with the confidence distribution, are decisive factors in the selection of point prompts. Finally, the generated point prompt and Query Image I_Q are fed into the SAM to predict the naive mask $\mathcal{M}_{coarse} \in \mathbb{R}^{H \times W}$. Additional NRM is designed for refining the $\mathcal{M}_{coarse}$.
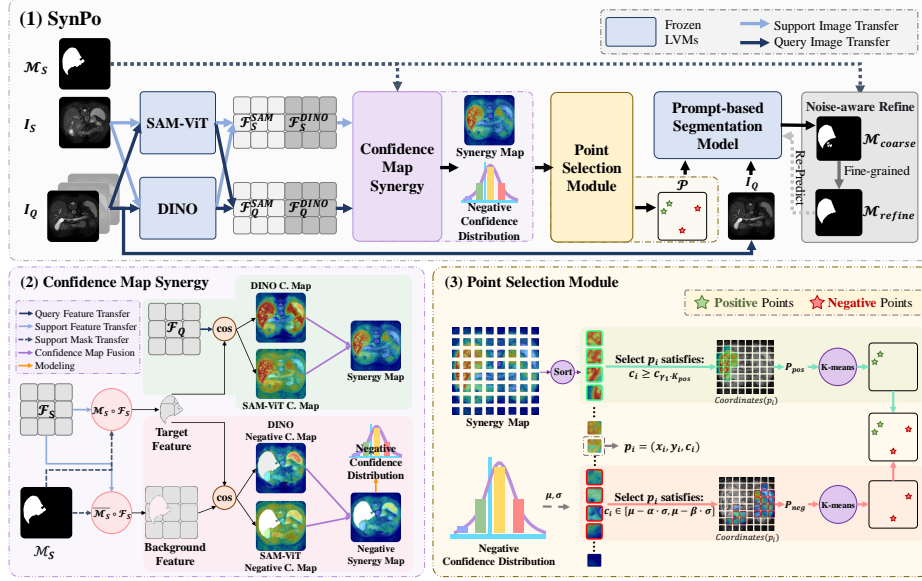
2.2   Confidence Map Synergy Module

As shown in Fig. 2 (2), CMSM is an innovative method with two branches to generate the synergy map, accompanied by the negative confidence distribution as a by-product. For the upper branch, the support features extracted corresponding to the foreground pixels within the visual concept from $\mathcal{F}_s$ using $\mathcal{M}_s$:

$$\left\{ \mathcal{T}_S^i \right\}_{i=1}^n = \{x_i | \ where \ x_i \in \mathcal{M}_s \circ \mathcal{F}_s \ and \ x_i \neq 0\}, \tag{1}$$

where $T_S^i \in \mathbb{R}^{1 \times c}$ and $\circ$ denotes spatial-wise multiplication. Subsequently, we calculate n confidence maps for each foreground pixel i by the cosine similarity between $\mathcal{T}_S^i$ and query feature $\mathcal{F}_Q$:

$$\left\{ \mathcal{S}^i \right\}_{i=1}^n = \left\{ \mathcal{F}_q \times \left(\mathcal{T}_S^i \right)^{\mathbf{T}} \right\}_{i=1}^n, \quad where \quad \mathcal{S}^i \in \mathbb{R}^{h \times w}. \tag{2}$$

Fig. 2. (1) Overview of SynPo Architecture. SynPo integrates through three key modules: Confidence Map Synergy Module (CMSM), Point Selection Module (PSM), and Noise-aware Refine Module (NRM). (2) Illustration of Confidence Map Synergy. In the Synergy Map module, features extracted from both the support and query images, using SAM-ViT and DINO, are jointly processed through a series of computations to produce the synergy map and the negative confidence distribution. (3) Illustration of Point Selection Module. Positive and negative prompts are selected in this module based on the generated synergy map and negative confidence distribution.

Next, we adopt an average pooling to aggregate all n local maps to obtain the overall confidence map $\mathcal{S} \in \mathbb{R}^{h \times w}$ for the target organ.

For the lower branch generating negative confidence map, we crop the support features of background pixels within the visual concept from $\mathcal{F}_S$ by $\overline{\mathcal{M}}_S$, where $\overline{\mathcal{M}}_S$ represents the inverse of $\mathcal{M}_S$:

$$\left\{\mathcal{B}_S^i\right\}_{i=1}^{H \times W - n} = \overline{\mathcal{M}}_s \circ \mathcal{F}_s, \tag{3}$$

where $\mathcal{B}_S^i \in \mathbb{R}^{1 \times c}$ and $\circ$ denotes spatial-wise multiplication. Afterward, we treat $\mathcal{B}_S$ as $\mathcal{F}_q$ and compute the negative confidence map $\mathcal{S}_{neg}$ in the same way.

Subsequently, we combine $\mathcal{S}^{SAM}, \mathcal{S}^{DINO}$ to generate $SynMap$:

$$SynMap = \delta_{S\text{-}D} \cdot \left(\mathcal{S}^{SAM} \odot \mathcal{S}^{DINO}\right) + \delta_S \cdot \mathcal{S}^{SAM} + \delta_D \cdot \mathcal{S}^{DINO}, \tag{4}$$

where $\odot$ denotes to Hadamard product and $\delta_{S\text{-}D} + \delta_S + \delta_D = 1$. The first term captures the non-linear interaction between the two matrices, effectively amplifying extreme values in the resultant matrix. This increases the sensitivity

to significant deviations in both confidence maps while reducing the impact of more neutral or moderate values. Furthermore, the subsequent terms provide controlled weights to the contribution of each matrix. Next, the same procedure is applied to $\mathcal{S}_{neg}^{SAM}$ and $\mathcal{S}_{neg}^{DINO}$, resulting in the fused representation, which is then flattened to obtain $SynMap_{neg} \in \mathbb{R}^{(H \times W - n) \times 1}$:

$$SynMap_{neg} = \text{Flatten}\left(\delta_{\text{S-D}} \cdot \left(\mathcal{S}_{neg}^{SAM} \odot \mathcal{S}_{neg}^{DINO}\right) + \delta_S \cdot \mathcal{S}_{neg}^{SAM} + \delta_D \cdot \mathcal{S}_{neg}^{DINO}\right). \tag{5}$$

For $SynMap_{neg}$, we model each pixel value p_i in it using a Gaussian probability density function:

$$P_{neg}(p_i^B|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(p_i^B - \mu\right)^2}{2\sigma^2}\right), \tag{6}$$

where μ and σ are derived through maximum likelihood estimation.

2.3   Point Selection Module

We sort each pixel $p_i = (x_i, y_i, c_i)$ in the synergy map by the confidence score of the pixel c_i in descending order, where x_i, y_i are the coordinates of the pixel.

For **positive** points, we selected the top $\gamma_1 \cdot K_p$ of p_i, where γ_1 is a *clustering scaling factor*[*] and K_p represents the desired number of positive points. The set of coordinates for the selected points can be written as:

$$R_{pos} = \{Coordinates\,(p_i)\,|i \leq \gamma_1 \cdot K_p\}, \tag{7}$$

where $p_i \in \mathbb{R}^2$. Next, we perform K-means clustering on these coordinates and select K_p centroid points, yielding $\mathcal{P}_{pos}$:

$$\mathcal{P}_{pos} = \text{KMeans}(R_{pos}, K_p), \tag{8}$$

where $\mathcal{P}_{pos} \in \mathbb{R}^{K_p \times 2}$ represents the set of coordinates.

For **negative** points, we select $\gamma_2 \cdot K_n$ pixels at most from the confidence region that are more relevant to the ROI, as defined by the following equation:

$$P'_{neg} = \{Coordinates(p_i)|c_i \in [\mu - \alpha \cdot \sigma, \mu - \beta \cdot \sigma]\}, \tag{9}$$

where the μ and σ derived from $SynMap_{neg}$, α and β are user-defined constants that control the bounds of the confidence interval. Subsequently, $\gamma_2 \cdot K_n$ pixels are randomly selected:

$$R_{neg} = \begin{cases} P'_{neg}, & \text{if } |P'_{neg}| \leq \gamma_2 \cdot K_n, \\ \text{Random}(P'_{neg}, \gamma_2 \cdot K_n), & \text{Otherwise.} \end{cases} \tag{10}$$

---

[*] The clustering scaling factor is to find points that are γ times larger than the range of all search points for clustering.

Then, the clustering similarly is applied to the positive points, yielding $\mathcal{P}_{neg}$:

$$\mathcal{P}_{neg} = \text{KMeans}(R_{neg}, K_n). \tag{11}$$

Finally, we take the intersection of two sets and assign labels to the points, resulting in the set

$$\mathcal{P} = \{(x,y,1)|(x,y) \in \mathcal{P}_{pos}\} \cup \{(x,y,0)|(x,y) \in \mathcal{P}_{neg}\}, \tag{12}$$

which is fed into SAM to generate the segmentation result $\mathcal{M}_{coarse}$.

2.4   Noise-aware Refine Module

The module firstly refines the naive coarse mask $\mathcal{M}_{coarse}$ generated by the PSM through an initial erosion operation to remove small noise, followed by a dilation step to restore the main structural regions. Let $\mathcal{M}_j$ denote the mask obtained for the $\mathcal{M}_{coarse}$'s j-th connected region C_j, we segment query feature by $\mathcal{M}_j$:

$$\left\{T_{Q,C_j}^i\right\}_{i=1}^{|C_j|} = \mathcal{M}_j \circ \mathcal{F}_Q, \tag{13}$$

where $|C_j|$ denotes the total number of pixels in C_j. Afterward, we treat T_{Q,C_j} as F_q and compute the confidence score of C_j in the same way, yeiled $\mathcal{S}_{C_j}$. Thereafter, the mean value is calculated in each connected domain:

$$score(C_j) = \frac{1}{|C_j|}\mathcal{S}_{C_j}. \tag{14}$$

We select the connected region with the highest score as $\mathcal{M}_{refine}$, and feed into the PSM as mask prompts. Combined with the point prompts, this facilitates further refinement of the segmentation. The output is then processed again through the same steps to produce the final segmentation $\mathcal{M}_{final}$.

3   Experiments and Results

3.1   Experimental Settings

Datasets & Evaluation Metrics: To assess the generalizability of our method, we conducted 1-shot segmentation experiments on Synapse-CT [7] and CHAOS-MRI [5] datasets, which contains 30 3D abdominal CT scans and 20 3D T2-SPIR MRI scans respectively. Dataset splits followed the protocols used by ADNet [4] and PerSAM [19]. We use the mean Dice score to assess performance, and report the standard deviation of Dice scores across 5-fold cross-validation.

Implementation Details: Image preprocessing follows the method described in [9], converting images into ROI of size 256×256. Features are extracted using DINOv2 [8] (model: Sinder [15]), resulting in a feature map of spatial size 64×64. SAM [6] (model: Sam2.1 Hiera large image prediction [10]) is applied to obtain a feature map of the same spatial size 64×64 for height and width. All experiments are conducted on an NVIDIA RTX-3090. As for Hyper-parameters, $\delta_{S\text{-}D} = 0.8$, $\delta_S = 0.1$, and $\delta_D = 0.1$ are fixed for Synapse-CT and CHAOS-MRI.

Table 1. Comparison with SOTA Methods. Except PerSAM and SynPo, the datas are either sourced from the original or other related articles.

Training Method			CHAOS-MRI					Synapse-CT				
			Upper		Lower		Mean	Upper		Lower		Mean
			Spleen	Liver	LK	RK		Spleen	Liver	LK	RK	
w	SSL-ALPNet[ECCV20] [9]		67.02	73.05	73.63	78.39	73.02	60.25	73.65	63.34	54.82	63.02
	ADNet[MIA22] [4]		75.92	80.81	75.28	83.28	78.82	63.48	77.24	72.13	79.06	72.97
	Q-Net[IntelliSys23] [11]		75.99	81.74	78.36	87.98	81.02	74.86	71.21	75.26	74.79	74.03
	RPT[MICCAI23] [22]		76.37	82.86	80.72	89.82	82.44	79.13	82.57	77.05	72.58	77.83
	GMRD[TMI24] [2]		76.09	81.42	83.96	90.12	82.90	78.31	79.60	81.70	74.46	78.52
w/o	PerSAM[ICLR24] [19]	μ	69.14	42.44	64.84	71.36	61.12	65.03	65.55	58.47	60.31	62.34
		$\pm\sigma$	±8.64	±3.09	±2.44	±5.66	±3.00	±9.33	±3.91	±4.30	±4.56	±1.78
	ProtoSAM[arxiv24] [1]	μ	76.51	81.94	71.46	81.43	77.83	65.50	87.84	69.44	71.04	73.45
		$\pm\sigma$	±7.97	±6.83	±6.85	±5.16	±6.85	±7.72	±4.71	±10.21	±10.52	±10.21
	SynPo(ours)	μ	80.30	77.32	77.32	83.04	81.15	83.76	81.32	75.00	79.63	79.91
		$\pm\sigma$	±7.63	±4.50	±3.10	±2.04	±3.21	±3.46	±2.16	±9.78	±5.39	±3.20
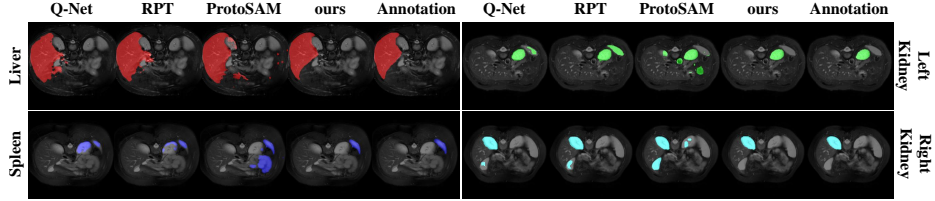
Fig. 3. Quality Results of Different Methods.

3.2   Comparison with SOTA Methods

Performance on Synapse-CT. As shown in Table 1, SynPo achieves the highest average Dice score (**79.91**) among training-free methods, surpassing Proto-SAM (73.45) and PerSAM (62.34) by a large margin. Notably, it outperforms some training-based methods in spleen (83.76) and right kidney (79.63), validating the effectiveness of our confidence map synergy and point prompt strategy.

Performance on CHAOS-MRI. SynPo achieves an average Dice score of **81.15**, outperforming training-free methods like ProtoSAM (77.83) and PerSAM (61.12), while performing competitively with the top training-based methods like GMRD (82.90). It also excels in spleen segmentation (80.30) and maintains competitive performance on other organs, demonstrating strong generalization.

Qualitative Comparison. As shown in Fig. 3, SynPo achieves segmentation results that closely align with the ground truth. In contrast, Q-Net, RPT, and ProtoSAM tend to produce over-segmented regions. SynPo effectively captures organ boundaries, reduces false positives, and preserves anatomical consistency,

Table 2. Modules' Ablation Study.

SAM	DINO	PSM	NRM	Dice	Gain
✓				57.48	-
✓	✓			66.70	+9.22
✓	✓	✓		78.74	+12.04
✓	✓	✓	✓	**79.93**	+1.19

Table 3. Ablation Study of Negative Prompt Strategy on CHAOS.

Model	Dice
PerSAM	56.72
+Negative Point Selection	**68.90**
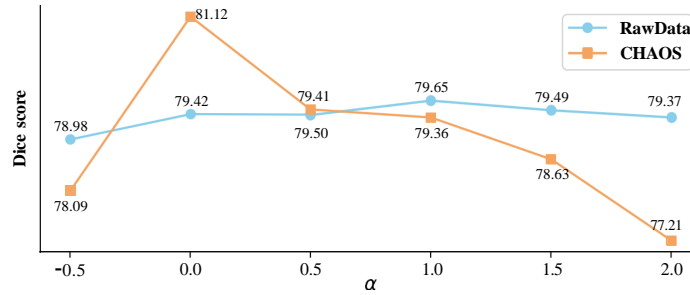
Fig. 4. Parameter Analysis on α and β (with $\beta = \alpha - 1.5$).

benefiting from our Confidence Map Synergy Module and Point Prompt Strategy. These results further demonstrate the superiority of SynPo in few-shot medical image segmentation.

3.3 Ablation study

Module Ablation To evaluate the contribution of each module, we conducted an ablation study on the Synapse-CT dataset (Table 2). The results show that adding DINO to SAM-ViT improves performance by +**9.22**, and further incorporating the PSM provides an additional +**12.04** gain. The NRM further boosts performance by +**1.19**, demonstrating the effectiveness of our synergy-based design.

Ablation Study of Negative Prompt Strategy Additionally, we replaced the negative point selection strategy in PerSAM to evaluate its impact on CHAOS-MRI (Table 3). This replacement yields a significant Dice improvement of +**13.18**, highlighting the importance of effective negative prompt selection in transferring natural image knowledge to medical.

Parameter Experiments Regarding the confidence interval bounds controller α and β are also conducted, shown in Fig. 4. The x-axis represents the value of α (with $\beta = \alpha - 1.5$) and the y-axis denotes the Dice score. The figure shows

that in CHAOS and RawData, the highest dice score was achieved under the confidence interval $[\mu, \mu + 1.5\sigma]$ and the confidence interval $[\mu - \sigma, \mu + 0.5\sigma]$. The figure proves low Dice scores at high-value $large$, validating our proposed negative point selection strategy that prioritizes not least but less similar points.

4   Conclusion

Our model SynPo addresses the limitations such as anatomical consistency of existing training-free few-shot medical segmentation methods by improving negative prompt quality through a Confidence Map Synergy Module and a Point Selection Module. Experiments on CT and MRI datasets demonstrate that SynPo achieves performance comparable to state-of-the-art methods and effectively transfers knowledge from LVMs trained on natural images to the medical image domain. Our work highlights the importance of high-quality negative prompts for few-shot medical image segmentation.

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

References

1. Ayzenberg, L., Giryes, R., Greenspan, H.: Protosam: One-shot medical image segmentation with foundational models. arXiv preprint arXiv:2407.07042 (2024)
2. Cheng, Z., Wang, S., Xin, T., Zhou, T., Zhang, H., Shao, L.: Few-shot medical image segmentation via generating multiple representative descriptors. IEEE Transactions on Medical Imaging **43**(6), 2202–2214 (2024)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
4. Hansen, S., Gautam, S., Jenssen, R., Kampffmeyer, M.: Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Medical Image Analysis **78**, 102385 (2022)
5. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical image analysis **69**, 101950 (2021)
6. 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)

7. Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In: Proc. MICCAI multi-atlas labeling beyond cranial vault—workshop challenge. vol. 5, p. 12. Munich, Germany (2015)
8. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
9. 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: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. pp. 762–780. Springer (2020)
10. 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)
11. Shen, Q., Li, Y., Jin, J., Liu, B.: Q-net: Query-informed few-shot medical image segmentation. In: Proceedings of SAI Intelligent Systems Conference. pp. 610–628. Springer (2023)
12. Tang, L., Jiang, P.T., Xiao, H., Li, B.: Towards training-free open-world segmentation via image prompt foundation models. International Journal of Computer Vision **133**(1), 1–15 (2025)
13. Tang, L., Xiao, H., Li, B.: Can sam segment anything? when sam meets camouflaged object detection. arXiv preprint arXiv:2304.04709 (2023)
14. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
15. Wang, H., Zhang, T., Salzmann, M.: Sinder: Repairing the singular defects of dinov2. In: European Conference on Computer Vision. pp. 20–35. Springer (2024)
16. Wang, R., Yang, Z., Song, Y.: Osam-fundus: A training-free, one-shot segmentation framework for optic disc and cup in fundus images. Biomedical Signal Processing and Control **100**, 107069 (2025)
17. Xiao, H., Tang, L., Li, B., Luo, Z., Li, S.: Zero-shot co-salient object detection framework. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4010–4014. IEEE (2024)
18. Zhang, J., Herrmann, C., Hur, J., Polania Cabrera, L., Jampani, V., Sun, D., Yang, M.H.: A tale of two features: Stable diffusion complements dino for zero-shot semantic correspondence. Advances in Neural Information Processing Systems **36** (2024)
19. Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Ma, X., Dong, H., Gao, P., Li, H.: Personalize segment anything model with one shot. arXiv preprint arXiv:2305.03048 (2023)
20. Zhao, Z., Gao, J., Luo, Z., Li, S.: Cpnet: Cross prototype network for few-shot medical image segmentation. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 135–149. Springer (2024)
21. Zhao, Z., Wang, R., Gao, J., Luo, Z., Li, S.: Mask matching network for self-supervised few-shot medical image segmentation. In: 2024 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2024)
22. Zhu, Y., Wang, S., Xin, T., Zhang, H.: Few-shot medical image segmentation via a region-enhanced prototypical transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 271–280. Springer (2023)