Improving Medical Image Segmentation with Implicit Representation and Noisy Label Robustness

Suruchi Kumari, Harshdeep Singh, and Pravendra Singh ⋆

Indian Institute of Technology Roorkee

Abstract. Medical image segmentation plays a vital role in healthcare by identifying and delineating specific structures, such as organs, tumors, or lesions, from medical images. While deep learning has significantly advanced this field, existing methods face two major challenges. First, they rely on pixel-wise discrete representations, which lead to difficulties in scaling to different input sizes and create ambiguity in fine boundary delineation. Second, the presence of noisy labels in medical datasets hinders model accuracy. To address these challenges, we propose a novel approach that leverages continuous representations and incorporates three key components: the Hierarchical Channel-Attention Encoder (HCAE), the Three-Stage Implicit Decoder with Noise-Based Index Selector (NBIS), and the High-Frequency Noise Modulator (HFNM). HCAE enhances feature extraction by capturing both fine and coarse details through hierarchical attention mechanisms. NBIS refines segmentation by identifying stable and unstable feature indices, improving performance in challenging regions. Meanwhile, HFNM selectively introduces noise to high-frequency components, helping the model mitigate the effects of label noise. This comprehensive solution demonstrates improved segmentation accuracy, particularly in the presence of noisy labels, making it a promising approach for medical image analysis.

Keywords: Medical image segmentation · Hierarchical Channel-Attention Encoder · Implicit representations · Noisy Label.

1 Introduction

Medical image segmentation is a crucial task in healthcare, aimed at delineating and identifying specific structures or regions of interest in medical images, such as organs, tumors, or lesions. Deep learning (DL) has revolutionized this field by automating feature extraction and enabling models to learn complex patterns [14]. Various deep learning methods have been proposed for this task [20, 3, 15], achieving strong performance. However, two main challenges remain with these methods. The **first** challenge is that they primarily focus on pixel-wise predictions, that is, on discrete representations. These discrete representations

⋆ Corresponding author: pravendra.singh@cs.iitr.ac.in

lack spatial continuity, leading to discretization artifacts and limited flexibility when dealing with arbitrary input sizes or fine-grained boundary details. This is critical in medical image analysis, as precise boundaries help distinguish between different tissues or anatomical structures. The **second** challenge is the prevalence of noisy labels in medical datasets [21, 11], a widespread issue often caused by human bias or inconsistencies. The direct application of supervised learning methods to data with noisy labels consistently leads to a decline in model performance [1].

To address the first challenge, researchers have explored continuous representations as an alternative to discrete predictions [27, 22, 9]. Continuous representations leverage Implicit Neural Representations (INRs) [17] to convert discrete segmentation outputs into a continuous space. Several approaches achieve this by learning a mapping between encoded image features and grid coordinates, allowing for adaptability across different output resolutions. However, despite their effectiveness, these methods often rely on features extracted from a single resolution and do not sufficiently capture both global context and local details, leading to poor segmentation performance in complex structures and boundary regions. Furthermore, their performance declines when faced with noisy labels. To address both of these issues, we propose HierachSAM. HierachSAM utilizes three components: Hierarchical Channel-Attention Encoder (HCAE), Three stage implicit decoder with Noise-Based Index Selector (NBIS), and High-Frequency Noise Modulator (HFNM).

To efficiently capture global context and local details, we utilize a hierarchical encoder based on SAM [10] to extract both fine-grained and coarse-level features. These features are then combined and passed through the Channel-Wise Hierarchical Attention (CWHA) mechanism, enabling the model to focus on the most informative features across both channels and spatial resolutions. Additionally, to further enhance performance in boundary regions, the NBIS module, which is integrated within the implicit decoder, selects both stable and unstable indices based on the variance between the original and noisy feature representations, identifying parts of the feature vector that require further refinement. This process enhances segmentation in challenging regions. In the HFNM module, the image is first decomposed using wavelet transforms to separate it into different frequency components. Controlled perturbations are then introduced specifically to the high-frequency components, which primarily represent edges, boundaries, and textures. Since label noise often appears in regions with complex boundaries, introducing noise to these high-frequency features during training encourages the model to learn more robust representations. By leveraging the HFNM module, the model becomes better equipped to handle label noise and improve feature extraction in high-frequency regions.

The main contributions of our work are as follows: (1) We introduce HCAE to enhance implicit segmentation by effectively capturing hierarchical features and focus on the most informative features across both channels and spatial resolutions. While three-stage implict decoder refines feature representations by using NBIS. (2) We propose HFNM to mitigate the impact of label noise by introducing

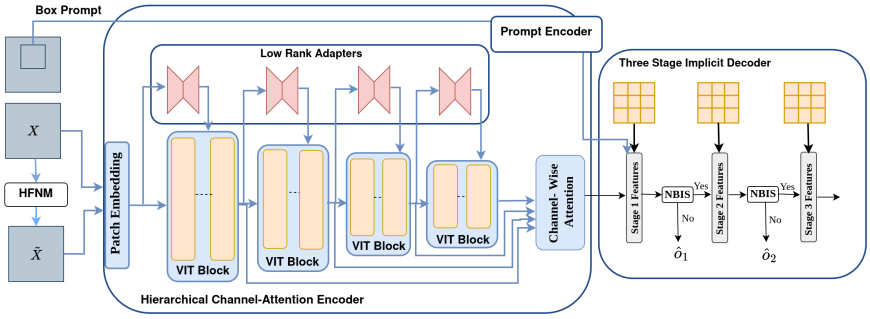

Fig. 1: Overview of our proposed framework: HCAE is utilized for improved feature learning, while NBIS within the implicit decoder refines segmentation in challenging regions, and HFNM enhances robustness against noisy labels.

controlled perturbations in high-frequency regions, improving model robustness. (3) Our approach outperforms existing methods on two medical datasets. Additionally, we conduct experiments under noisy conditions to demonstrate the effectiveness of our method.

2 Method

2.1 Preliminaries

In line with other approaches, we utilize the same convention as [22]. In conventional discrete segmentation, we have N medical images $X \in \mathbb{R}^{H\times W\times 3}$, which are mapped to class probability maps $O \in \mathbb{R}^{H\times W\times K}$ while maintaining the original resolution, where H and W represent the height and width of the image, respectively and K is the number of classes. In contrast, implicit image segmentation takes each pixel coordinate $p_i = (x, y)$, where $x, y \in [-1, 1]$, and maps it along with the corresponding image X_i to class probabilities $\hat{o}_i \in \mathbb{R}^K$ using a neural network N_θ. This is formulated as $N_\theta : (p_i, X_i) \rightarrow \hat{o}_i$, where θ represents the network parameters. Unlike discrete segmentation, this method integrates spatial coordinates directly, enabling flexible predictions at arbitrary resolutions. This allows transformations from $X \in \mathbb{R}^{H\times W\times 3}$ to $O \in \mathbb{R}^{H'\times W'\times K}$, adapting the segmentation output to different scales.

2.2 Hierarchical Channel-Attention Encoder

In medical images, structures can vary greatly in size, shape, and position. Hierarchical features help the model understand different scales of features, from coarse (high-level structures) to fine (detailed anatomical boundaries). To efficiently capture these hierarchical features, we utilize both a feature encoder and a prompt encoder based on the SAM [10]. SAM has demonstrated strong

potential as a backbone for various segmentation tasks. Consequently, several works employ diverse adapters using parameter-efficient fine-tuning (PEFT) to adapt SAM for medical image segmentation [23, 24]. However, in contrast to fine-tuning all parameters in the image encoder, we leverage the Low-Rank Adapter (LoRA) to update only a small fraction of parameters, which allows us to efficiently adapt SAM to medical images [22]. Furthermore, we introduce our novel Channel-Wise and Hierarchical Attention (CWHA) mechanism, to focus on the most informative features across both channels and spatial resolutions.

Combining the embeddings from multiple layers helps capture features at different levels. However, combining them simply does not yield improvements. To efficiently combine these embeddings, we utilize CWHA mechanism, Specifically, If the embeddings of the encoder at layer l are denoted as E_l, given four sets of embeddings with shapes (C, H, W)—where C represents the channel dimension and $H \times W$ represents the spatial dimensions—the CWHA mechanism applies a series of transformations to compute attention weights for each hierarchical level. Each set of embeddings is first interpolated to match the spatial dimensions of the original image. Once all embeddings are of same size, we apply channel averaging as given below:

$$\mathbf{e}_l = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} E_l[:, h, w] \tag{1}$$

where $l \in \{1, 2, 3, 4\}$ denotes the embedding levels. Next, we introduce a randomly initialized vector, $\mathbf{v} \in \mathbb{R}^C$, which will weight each channel. The dot product between $\mathbf{e}_l$ and $\mathbf{v}$ is passed through a ReLU activation to ensure non-negativity, giving us a channel-weighted vector as given below:

$$\mathbf{w}_i = \text{ReLU}(\mathbf{e}_l \cdot \mathbf{v}) \tag{2}$$

The channel-weighted vector $\mathbf{w}_i$ is then used to reweight the original embeddings E_l, emphasizing the most relevant channels:

$$E'_l = E_l \odot \mathbf{w}_i \tag{3}$$

where $\odot$ denotes element-wise multiplication. Further, we perform a 3D average pooling on the reweighted embeddings across the hierarchical levels, followed by a softmax operation applied across these four levels to compute the attention weights. Let $\mathbf{a} \in \mathbb{R}^4$ represent the vector of attention weights for each level:

$$\mathbf{a_l} = \text{softmax}\left(\text{pooling}(\mathbf{E'}_l)\right.$$

Finally, we compute the combined embedding as a weighted sum of the embeddings from each hierarchical level:

$$\sum_{i=1}^{4} a_l \cdot E'_l$$

2.3 Three Stage Implicit Decoder with Noise-Based Index Selector

In implicit decoder [22], image and prompt features are interpolated from the source to the target resolution and concatenated with target coordinates, p, which are normalized to $[-1, 1]$. To avoid bias from direct coordinate usage, a high-frequency positional encoding is applied to the coordinates [18]. The encoded coordinates, along with the interpolated image and prompt features, are then concatenated and passed into the decoder.

To further refine the features, we utilize a Noise-Based Index Selector (NBIS), which samples important positional indices from the output feature vector based on the variance between original and noisy representations of the feature. This helps identify both stable and unstable indices, highlighting which parts of the feature vector require further refinement. Specifically, we create a noisy representation of the feature by adding speckle noise denoted by $\mathbf{n}_s \sim \mathcal{N}(1, \sigma_s^2)$, to get: $\tilde{\mathbf{f}}_s = \mathbf{f} \cdot \mathbf{n}_s$. Next we add Gaussian noise $\mathbf{n}_g \sim \mathcal{N}(0, \sigma_g^2)$ to $\tilde{\mathbf{f}}_s$ to obtain the final noisy features:

$$\tilde{\mathbf{f}} = \tilde{\mathbf{f}}_s + \mathbf{n}_g$$

Now, we pass both $\mathbf{f}$ and $\tilde{\mathbf{f}}$ through the linear layers to obtain the output vectors $\mathbf{o}$ and $\tilde{\mathbf{o}}$, respectively. The variance $\mathbf{v}$ between these two outputs is computed element-wise as:

$$\mathbf{v} = (\mathbf{o} - \tilde{\mathbf{o}})^2$$

We select the top k indices based on the highest values in $\mathbf{v}$, corresponding to the indices with the greatest variance. These indices represent the parts of the feature vector that are less stable and may require additional refinement.

To effectively refine the indices that are uncertain, we utilize the Three-Stage Implicit Decoder. In this decoder, at each subsequent stage, the top k_1 and k_2 unstable coordinates are selected. The decoder consists of three stages, where each stage involves a combination of linear layers and outputs a segmentation mask. Specifically, the features f are passed through the first-stage layers, producing the mask M_1. Stable and unstable coordinates are then selected using the NBIS mechanism. The top k_1 unstable coordinates are forwarded to the second stage for further refinement. This process is repeated in the third stage, where the remaining unstable coordinates undergo additional refinement. This noise-based sampling strategy helps the model focus on both stable and unstable indices, ensuring that features needing further attention are identified and refined.

2.4 High-Frequency Noise Modulator

Noisy labels are a frequent obstacle when training deep neural networks effectively. To address this issue, previous research has introduced various techniques to enhance noise tolerance and improve model performance. In this work, we utilize the power of frequency domain to lessen the impact of noisy labels.

We convert the input images into the wavelet space using discrete wavelet transformations (DWT) with Haar wavelets. This process allows us to separate

Table 1: Comparison of segmentation performance across different methods.

Binary Polyp Segmentation			Multi-class Organ Segmentation		
Method	Dice (%)↑	HD Distance ↓	Method	Dice (%)↑	HD Distance ↓
Discrete Approaches					
U-Net [20]	63.89±1.30	31.30	U-Net	74.47±1.57	6.50
PraNet [5]	82.56±1.08	-	UNETR	81.14±0.85	-
Res2UNet [6]	81.62±0.97	-	Res2UNet	79.23±0.66	-
nnUNet [7]	82.97±0.89	-	nnUNet	85.15±0.67	-
MedSAM [16]	82.88±0.55	21.53	MedSAM	85.85±0.81	10.62
ConDSeg [13]	89.1	-	-	-	-
-	-	-	TDFormer [4]	90.1	5.7
Implicit Approaches					
OSSNet [19]	76.11±1.14	-	OSSNet	73.38±1.65	-
IOSNet [9]	78.37±0.76	51.57	IOSNet	76.75±1.37	21.46
SwIPE [27]	85.05±0.82	-	SwIPE	81.21±0.94	-
I-MedSAM [22]	91.49±0.52	11.59	I-MedSAM	89.91±0.68	5.95
Ours	92.68±0.32	9.74	(ours)	91.67±0.24	5.23

an image into its low-frequency components, representing broad structural information, and high-frequency components, capturing detailed variations. For a 2D image, we obtain four distinct frequency components, which can be expressed as:

$$LL, \{LH, HL, HH\} = DWT(X), \tag{4}$$

where LL captures low-frequency information, while LH, HL, HH encode high-frequency details. We introduce random perturbations to the high-frequency coefficients, with details provided in the supplementary material. The perturbed coefficients, $\tilde{LH}, \tilde{HL}, \tilde{HH}$, are then reconstructed using the Inverse Discrete Wavelet Transform (IDWT):

$$\tilde{X} = IDWT(LL, \tilde{LH}, \tilde{HL}, \tilde{HH}). \tag{5}$$

We utilize the $\tilde{X}$ in the same way as X. This approach is useful in combating label noise, as noisy annotations often affect fine-grained structures like edges and textures. By training on augmented images where high-frequency details are altered but low-frequency structures remain stable, the model becomes less sensitive to mislabeled or uncertain regions, improving its ability to generalize and make consistent predictions despite noisy labels.

2.5 Training

For training, the pre-trained SAM's image encoder Enc_I is kept frozen, LoRA and prompt encoder Enc_P, and INRs remain trainable. We leverage SAM's image encoder with LoRA to extract features from medical images X, while the prompt encoder processes coarse bounding box P features for target segmentation, as shown in Figure 1. The bounding box P is preprocessed following previous work [26]. Finally, the extracted features, along with mapped coordinate values, are concatenated and decoded using the proposed three-stage INR decoder.

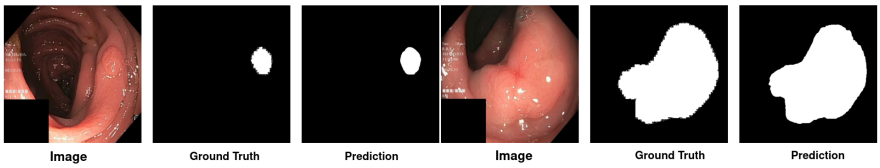

Fig. 2: Qualitative results of the proposed method.

With three-stage INR decoder, we obtain three INRs (Figure 1), which provide three point-wise segmentation probabilities $\{\hat{o}_1, \hat{o}_2, \hat{o}_3\}$, combined as $\hat{o}$. For training optimization, we adopt the conventional segmentation loss, formulated as:

$$L_{\text{seg}}(o_i, \hat{o}_i) = L_{\text{CeDice}}(o_i, \hat{o}_i) \tag{6}$$

where $L_{\text{CeDice}} = 0.5(L_{\text{ce}}+L_{\text{dice}})$ and L_{ce} and L_{dice} represent the Cross Entropy loss and Dice loss, respectively. Similarly, we pass the $\tilde{X}$ to the encoder and decoder and get the final output as $\hat{o'}$.

$$L_{\text{seg\_HFNM}}(o_i, \hat{o}'_i) = L_{\text{CeDice}}(o_i, \hat{o}'_i) \tag{7}$$

The final loss is defined as follows:

$$L_{\text{final}} = L_{\text{seg}}(o_i, \hat{o}_i) + \eta \cdot L_{\text{seg\_HFNM}}(o_i, \hat{o}'_i) \tag{8}$$

3 Experiments

Datasets. We conducted experiments on two datasets. The Kvasir-Sessile dataset [8] is a challenging dataset for binary polyp segmentation, containing 196 RGB images of small sessile polyps. The second dataset, BCV [12], is used for multi-organ segmentation and includes 30 CT scans annotated with 13 organs. To evaluate the model's generalization ability, we tested the pre-trained model—trained on the Kvasir-Sessile dataset—directly on the CVC-ClinicDB dataset [2], which consists of 612 images from 31 colonoscopy sequences.

3.1 Experimental Results

Segmentation Comparisons. We compare our approach with both discrete and implicit methods, as shown in Table 1. For the smaller polyp dataset, we observe significant improvements over the best-performing implicit and discrete methods, as shown in Table 1. In the case of multi-organ segmentation on the BCV dataset, we also achieve notable performance improvements compared to the leading implicit and discrete methods. The qualitative results are shown in Figure 2.

Table 2: Cross-domain and Cross-resolution results on binary polyp datasets.

Method	Kvasir-Sessile	Kvasir-Sessile → CVC	384 → 128	384 → 896
nnUNet [7]	63.89	84.91	73.97	83.56
MedSAM [16]	82.88	74.59	82.37	83.32
IOSNet [9]	78.37	70.10	76.18	78.01
I-MedSAM [22]	91.49	88.83	91.45	91.33
Ours	92.68	92.07	92.8	92.43

Table 3: Experiments with noisy labels on binary polyp segmentation.

	Noise (0.7, 0.03, 180)		Noise (0.3, 0.05, 200)		Noise (0.8, 0.05, 200)	
Method	Dice (%)↑	HD (%)↓	Dice (%)↑	HD (%)↓	Dice (%)↑	HD (%)↓
I-MedSAM	86.2	18.33	83.87	19.49	84.37	18.71
I-MedSAM v2	87.1	17.39	85.1	18.45	85.86	17.89
Ours (without HFNM)	89.21	14.21	88.43	15.61	87.93	15.42
Ours (with HFNM)	90.05	12.61	89.47	13.72	89.1	13.31

Cross-Resolution and Cross-Domain Comparisons. We compare the robustness across two different resolutions: 128×128 for lower resolutions and 896 × 896 for higher resolutions. As shown in Table 2, our method achieves the highest performance across both output resolutions. Secondly, we investigate the robustness of model performance across different datasets for the same task. In the binary-class polyp segmentation task, all methods are pre-trained on the Kvasir-Sessile dataset and evaluated directly on the CVC dataset. As shown in Table 2, our method outperforms existing methods.

Segmentation Comparison on Noisy Annotations. Following [25], we simulate annotation noise, with specific details provided in the supplementary materials. We evaluate our method under three different types of noise, as shown in Table 3. First, we compare our approach with I-MedSAM, the state-of-the-art (SOTA) method for implicit medical image segmentation. Our method significantly outperforms I-MedSAM. However, since I-MedSAM is not explicitly designed for noisy annotations, we further enhance its robustness to noise using the approach given in [3], which we refer to as I-MedSAM v2. Despite this adaptation, as shown in Table 3, our method still achieves a significantly higher performance margin.

4 Ablation Studies

Component-wise ablations. To evaluate the effectiveness of each component, we perform a component-wise ablation study on the Kvasir-Sessile dataset, as shown in Table 4. First, we report the Dice score for the baseline, which is defined as using the pre-trained SAM model with LoRA and INR for segmentation. Next, we introduce the CWHN component, resulting in a 2.27% improvement in Dice score performance. Furthermore, utilizing a three-stage decoder without NBIS yields an improvement of 0.8%, while incorporating NBIS further enhances the results by an additional 0.8%.

Three-stage decoder ablations. We first experimented with a two-stage decoder, which achieved a Dice score of 91.4. Incorporating NBIS improved the

Table 4: Effectiveness of each component of the pipeline. We evaluate the Dice metric for both cross-domain and cross-resolution tasks.

Baseline	CWHN	Three-Stage decoder	NBIS	Kvasir-Sessile
✓				88.8
✓	✓			91.1
✓	✓	✓		91.9
✓	✓	✓	✓	92.7

score to 92.0. However, switching to a three-stage decoder with NBIS further improved the Dice score by 0.7% compared to the two-stage version.

5 Conclusion

In this work, we address two key challenges in medical image segmentation: the limitations of discrete pixel-wise representations and the adverse effects of noisy labels. To overcome these issues, we propose a novel framework that leverages implicit neural representations and enhanced feature extraction mechanisms. Specifically, Our method integrates hierarchical channel-attention based encoding (HCAE) for improved feature learning, a noise-based index selector (NBIS) within the three-stage implicit decoder to refine segmentation in challenging regions, and a high-frequency noise modulator (HFNM) to enhance robustness against noisy labels. Experimental results demonstrate that our method outperforms existing approaches across two datasets and proves effective in clean and noisy label scenarios.

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

References

1. Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., et al.: A closer look at memorization in deep networks. In: International conference on machine learning. pp. 233–242. PMLR (2017)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
3. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
4. Du, H., Dong, Q., Xu, Y., Liao, J.: Tdformer: Top-down token generation for 3d medical image segmentation. IEEE Journal of Biomedical and Health Informatics (2025)

5. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 263–273. Springer (2020)
6. Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.: Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence **43**(2), 652–662 (2019)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Jha, D., Smedsrud, P.H., Johansen, D., De Lange, T., Johansen, H.D., Halvorsen, P., Riegler, M.A.: A comprehensive study on colorectal polyp segmentation with resunet++, conditional random field and test-time augmentation. IEEE journal of biomedical and health informatics **25**(6), 2029–2040 (2021)
9. Khan, M.O., Fang, Y.: Implicit neural representations for medical imaging segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 433–443. Springer (2022)
10. 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)
11. Kumari, S., Singh, P.: Data efficient deep learning for medical image analysis: A survey. arXiv preprint arXiv:2310.06557 (2023)
12. 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 (2015)
13. Lei, M., Wu, H., Lv, X., Wang, X.: Condseg: A general medical image segmentation framework via contrast-driven feature enhancement. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 4571–4579 (2025)
14. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
15. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
16. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024)
17. Molaei, A., Aminimehr, A., Tavakoli, A., Kazerouni, A., Azad, B., Azad, R., Merhof, D.: Implicit neural representation in medical imaging: A comparative survey. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2381–2391 (2023)
18. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: International conference on machine learning. pp. 5301–5310. PMLR (2019)
19. Reich, C., Prangemeier, T., Cetin, Ö., Koeppl, H.: Oss-net: memory efficient high resolution semantic segmentation of 3d medical data. British Machine Vision Conference (2021)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)

21. Shi, J., Zhang, K., Guo, C., Yang, Y., Xu, Y., Wu, J.: A survey of label-noise deep learning for medical image analysis. Medical Image Analysis **95**, 103166 (2024)
22. Wei, X., Cao, J., Jin, Y., Lu, M., Wang, G., Zhang, S.: I-medsam: Implicit medical image segmentation with segment anything. In: European Conference on Computer Vision. pp. 90–107. Springer (2024)
23. Wei, X., Zhang, R., Wu, J., Liu, J., Lu, M., Guo, Y., Zhang, S.: Nto3d: Neural target object 3d reconstruction with segment anything. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20352–20362 (2024)
24. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
25. Yao, J., Zhang, Y., Zheng, S., Goswami, M., Prasanna, P., Chen, C.: Learning to segment from noisy annotations: A spatial correction approach. arXiv preprint arXiv:2308.02498 (2023)
26. Zhang, K., Liu, D.: Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785 (2023)
27. Zhang, Y., Gu, P., Sapkota, N., Chen, D.Z.: Swipe: Efficient and robust medical image segmentation with implicit patch embeddings. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 315–326. Springer (2023)