OG-SAM: Enhancing Multi-Organ Segmentation with Organogenesis-Based Adaptive Modeling

Xidong Wu[1,3,*], Hao Chen[2,*], Zhuoyuan Li[1,3], and Chao Li[1,2]

[1] School of Science and Engineering, University of Dundee
[2] Department of Clinical Neurosicence, University of Cambridge
[3] College of Medicine and Biological Information Engineering , Northeastern University

Abstract. The Segment Anything Model (SAM) excels in image segmentation. yet is challenged in multi-organ segmentation, due to the inherent similarities between organ tissues and the substantial variability in organ size, structure, and texture. This paper proposes to guide the adaptation of SAM for multi-organ segmentation, introducing biological priors of organogenesis, where organs arise from specific germ layers and develop with shared early-stage pathways before divergence into unique structures. We present *OG-SAM* (Organogenesis SAM), a new paradigm that enables organ-wise adaptation. First, we present *OrganAdapt* (Organ Adaptation) that integrates a biologically inspired hierarchical adaptation module into SAM, where parameter sharing and specialization follow the developmental trajectory of organs. Second, to effectively address variations in organ size, we propose *GoF* (Generalized Organ-feature Fusion), a mechanism that facilitates organ-specific multiscale feature pyramid fusion, thereby enhancing segmentation accuracy and robustness. OG-SAM functions as a query-based plug-in, seamlessly integrating with SAM. Experiments show that OG-SAM outperforms competing methods, particularly for challenging organ boundaries.

Keywords: Multi-Organ Segmentation · Organ-specific Adaptation · Segment Anything Model.

1 Introduction

Background. Accurate multi-organ segmentation enables a systematic assessment of human health and is critical for various clinical applications [26,21,13]. Conventional multi-organ segmentation techniques, such as thresholding [11,23], edge detection [29,22,17], and K-means clustering [27,31], rely on manually designed features and parameter tuning. Further, these methods lack adaptability and struggle to achieve robust performance across varying anatomical structures.

Deep learning has transformed image segmentation, enabling significant advancements in accuracy and efficiency [19,3]. Convolutional neural network (CNN)-based architectures, such as U-Net [24], and its subsequent refinements, in-

*Equal contribution. Corresponding email: cl647@cam.ac.uk

cluding nnU-Net [12] and UX-Net [15], have demonstrated remarkable performance. More recently, Transformer-based models, such as nnFormer [30] and UNETR [25], have further enhanced segmentation by leveraging self-attention mechanisms. However, despite these advancements, existing methods often struggle with generalizability across diverse datasets and tasks. Additionally, their reliance on large volumes of annotated training data poses a significant challenge for real-world medical applications, where data annotation is costly and labor-intensive.

The emerging Segment Anything Model (SAM), a large-scale foundational model, has significantly advanced image segmentation. Evidence has shown that SAM promises to effectively capture shape, structure, and advanced semantic features for segmentation. Previous studies further demonstrated that only limited annotated data is needed for fine-tuning SAM, achieving better performance over professional models on comprehensive datasets. For example, Chen *et al.* [2] proposed MA-SAM, which devises a parameter-efficient fine-tuning strategy introducing a 3D adapter into the image encoder to integrate volume or time information, effectively adapting the 2D network backbone to the 3D medical images. In addition, Gong *et al.* [6] developed a 3D SAM adapter that extends SAM from 2D to 3D through overall architecture modifications to support volume input while retaining most pre-training parameters.

Despite these advancements, developing robust models for multi-organ segmentation remains challenging, with two critical obstacles yet to be addressed:

1. *Nuanced Relationships Between Organs.* A major challenge in multi-organ segmentation stems from the fact that certain organs have both shared and unique structural and textural features, due to their shared embryonic origins and subsequent divergence in development. For example, the liver and pancreas, both from the endoderm, share glandular traits but differ in textural characteristics. This variability complicates segmentation, as a single set of parameters struggles to generalize effectively across organs. Explicitly modelling organ relationships is crucial for capturing robust features.
2. *Distinctive Morphological Characteristics.* Multi-organ segmentation is inherently challenging due to the substantial variations in organ morphology, size, and structural complexity. These differences make it particularly difficult to accurately identify and delineate individual organs, especially smaller structures such as the adrenal glands. Furthermore, the anatomical proximity of abdominal organs complicates segmentation, often resulting in blurred boundaries, organ overlaps, or missed structures. For example, distinguishing the gallbladder from the liver remains a persistent challenge due to their adjacent positioning and similar tissue characteristics.

Contribution. To tackle these challenges, this paper proposes *OG-SAM* (Organogenesis SAM), a novel adaptation paradigm that introduces organ development modelling as a biological prior to enhance the accuracy and robustness of SAM for multi-organ segmentation. It is established in developmental biology that organs arise from common germ layers (ectoderm, mesoderm, and endoderm) and follow divergent developmental pathways. This hierarchical organogenesis

informs our design of a hierarchical branching adaptation architecture in our proposed OrganAdapt (Organ Adaptation), where earlier layers capture fundamental structural features shared across multiple organs, while later layers gradually refine these representations to organ-specific features.

To address organ size variability, we propose GoF (Generalized Organ-feature Fusion) module, which extracts a feature pyramid from different layers of the segmentation model, aggregating features at multiple levels of abstraction adaptively. By leveraging a hierarchical representation, GoF allows organs of different sizes to selectively utilize relevant feature levels. It incorporates organ-specific parameters, enabling dynamic adaptation that optimizes feature fusion from different layers based on organ characteristics. Our contributions include:

- The key component of OG-SAM, OrganAdapt, is inspired by embryonic development and regulates parameter sharing and specialization among organs. By aligning model adaptation with organ development pathways, it dynamically adjusts shared and organ-specific representations, ensuring anatomically coherent segmentation with enhanced accuracy and robustness.
- The GoF module extracts a multi-scale feature pyramid from different layers of the segmentation model. Using an organ-specific parameter p, it adaptively fuse features based on each organ's unique structure and morphology.
- We formulate OG-SAM, as a query-based plug-in, seamlessly integrating with SAM, where the organ class serves as a query to dynamically adjust parameters for each organ following a unique trajectory. This enables a gated adaptation mechanism, allowing the model to effectively function as a specialized model with enhanced performance, particularly for challenging cases.

2 Methodology

2.1 Preliminaries: SAM with Parameter-Efficient Adaptation

The SAM is a foundational vision model for general-purpose segmentation based on a given prompt. For adaptation, we utilize SAM's image encoder, a *ViT-B* architecture [5] with 14×14 windowed attention and four evenly distributed global attention blocks. We also employ SAM's mask decoder, consisting of a modified transformer decoder block followed by a dynamic mask prediction head. Given an input image I, SAM generates a segmentation mask M, defined as:

$$M = d_\delta(f_\theta(I)), \tag{1}$$

where f_θ and d_δ represent the image encoding and decoding functions, respectively, both parameterized by deep neural network weights θ and δ.

Training large-scale vision models from scratch incurs substantial computational costs. To improve efficiency, Parameter-Efficient Adaptation (PEA) methods, such as Low-Rank Adaptation (LoRA) [10] and Adapters [9] , are commonly employed. LoRA constrains the weight update $\Delta\theta$ to a low-rank decomposition:

$$\Delta\theta = AB^T, \quad A \in \mathbb{R}^{d\times r}, B \in \mathbb{R}^{d\times r}, \tag{2}$$

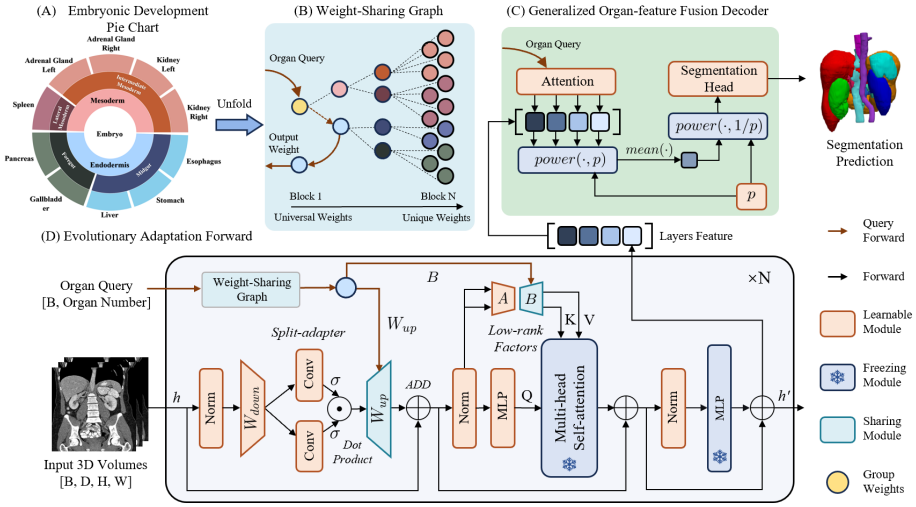

Fig. 1: Overview of OG-SAM. (A) Illustration of the organogenesis development stages and organ specialization. (B) A weight-sharing graph designed based on (A), specifying the shared weights among organs at different layers. (C) Generalized fusion mechanism incorporating a learnable parameter p to hierarchically integrate features based on a given organ query. (D) Adaptation of the SAM framework using OrganAdapt, where the organ query first determines the shared weights, which are subsequently applied to the adaptation modules.

where $r \ll d$, ensuring a reduction in the parameter space while maintaining expressive capacity. In contrast, Adapters modify the model output from h to h' rather than altering its primary weight parameters, achieved by introducing a learnable bottleneck structure:

$$h' = h + \mathcal{H}(h) = h + W_{\text{up}}(\sigma(W_{\text{down}}h)), \tag{3}$$

where $\mathcal{H}(h)$ consists of a sequence of operations: a down-projection $W_{\text{down}} \in \mathbb{R}^{d\times r}$, a non-linearity $\sigma(\cdot)$ (*e.g.*, GELU [8]), and an up-projection $W_{\text{up}} \in \mathbb{R}^{r\times d}$.

For model adaptation, given the initial parameters θ_0 of a particular layer i in SAM, the adaptation process is formulated as:

$$h_{i+1} = f^i_{\Theta}(h_i + \mathcal{H}(h_i)), \quad \text{where} \quad \Theta = \theta_0 + AB^T \tag{4}$$

where maintaining a small rank r allows PEA to significantly reduce memory overhead while enabling SAM to efficiently specialize in new segmentation tasks.

2.2 Overview of Organogenesis SAM

As shown in Figure 1, OG-SAM is structured into two modules: Organ Adaptation (OrganAdapt) (§2.3) and Generalized Organ-feature Fusion (GoF) (§2.4).

Our approach follows a query-based paradigm, where each query encodes the interest in specific organs or their absence.

2.3 Organ Adaptation

Figure 1A presents the hierarchical development of organs where an embryo undergoes sequential partitioning, gradually developing into distinct organs. Analogous to this process, our weight-sharing graph (Figure 1B) follows a progressive approach. At the initial stage, all organs share a common set of weights, which gradually diverge, forming specialized clusters until each organ acquires its own distinct set of weights, ultimately enabling independent functionality.

Our Organ Adaptation (Figure 1D) consists of three key components, each split by additive operations: the Split-Adapter, Low-Rank Factors, and a MLP. The Organ Adaptation begins with an *Organ Query*, a binary indicator specifying the organs of interest within the corresponding input 3D volume. This query is first processed through Weight-Sharing Graph, which dynamically determines the weight W_{up} and B based on both the depth of the current block and the organ types. At each layer, a given input feature h is first processed by Split-Adapter, which applies a downscaling MLP before partitioning the feature into two distinct convolutional branches, followed by a dot-product computation. The resulting representations are projected through an upsampling MLP:

$$h_a = W_{up}\left(\sigma(\text{Conv}_1(\tilde{h})) \odot \sigma(\text{Conv}_2(\tilde{h}))\right) + h, \quad \text{where } \tilde{h} = W_{down}\text{N}(h), \quad (5)$$

where the $\text{N}(\cdot)$ represent the Norm module. Subsequently, the transformed feature h_a is fed into the Low-Rank Factors module, where the Key (K) and Value (V) representations are learned through a Low-rank factors before passing it into a multi-head self-attention mechanism. The Query (Q) is separately obtained through another MLP with weight W_q.

$$Q = W_q(\text{N}(h_a)), \quad [K, V] = BA(\text{N}(h_a)), \quad (6)$$

where the weight matrix B expands the channel dimension to twice that of the Q before being evenly split into the K and V. The Q, K, and V are then fed into the multi-head self-attention mechanism to compute the feature for next step.

Finally, the output undergoes a nonormalizationrm layer followed by an MLP transformation. Notably, while the normalization layer originates from the pretrained network and is typically frozen, our experiments demonstrate that making it learnable enhances adaptability.

2.4 Generalized Organ-feature Fusion

Existing segmentation models are challenged by the significant variability in organ size and shape, leading to misdetections and blurred boundaries. A unified feature utilization decoder may not address the heterogeneity of organs and their

scale variations. Therefore, we propose Generalized Organ-feature Fusion within the Decoder (Figure 1C) to extract feature outputs from each block. Given an Organ Query Q_o, we first compute a soft attention scoring each feature:

$$S = W_2(\sigma(W_1 Q_o)), \tag{7}$$

where W_1 and W_2 are the weight matrices of two MLP layers, and S represents the computed attention scores. Subsequently, a weighted aggregation operation is applied, followed by a feature-wise generalized mean pooling [28] to fuse the feature representations into one:

$$f' = (G)^{\frac{1}{p}}, \quad \text{where} \quad G = (\sum_i (S_i f_i)^p)/(\sum_i S_i), \tag{8}$$

where f_i is the feature at layer i, and p is a learnable parameter. Then the output fused feature f' will go through the segmentation head to predict the mask.

3 Experiments and Results

3.1 Setup and Implementation

The BTCV Challenge dataset [14], is a widely used multi-organ segmentation benchmark in abdominal CT scans. It contains a total of 30 CT volumes, with 13 abdominal organs manually annotated. We only include parenchymatous organs and remove blood vessels due to the focus on modelling organ structure and texture. Each CT scan contains 85 to 198 slices. Slice thickness ranges from 2.5-5.0 mm. All scans are 512 × 512 pixels. In-plane resolution ranges from 0.54×0.54 mm^2 to 0.98×0.98 mm^2 .

Metrics: To assess model performance, we use two metrics: Dice Similarity Coefficient (Dice) and 95% Hausdorff Distance (HD95). The Dice measures the overlap between the predicted mask and the ground truth, with a higher score indicating a higher accuracy. Meanwhile, HD95 evaluates the maximum discrepancy between the predicted and actual boundaries within the 95th percentile.

Implementation: We implemented models in PyTorch [20] and MONAI [1] frameworks. All experiments and comparisons are based on the SAM-B model, using ViT-B [5] as the backbone of the image encoder. The model was trained on NVIDIA A6000 GPU using AdamW optimizer [18] and equipped with a linear scheduler for 200 epochs of training.

3.2 Model Comparisons

We extensively compared our method with existing 3D medical segmentation models, including CNN-based, Transformer-based, and SAM-based fine-tuning methods, *i.e.*, nnU-net, UNETR, swinUNETR, nnFormer, UX-net, SAM-Adapter, LoRaMedNet, and 3DSAM-adapter (Table 1). The results show that our method achieves the highest average performance in multi-organ segmentation, outperforming the best competing model by at least 1.0% in Dice score (UX-net) and

Table 1: Comparison of quantitative results on the BTCV dataset.

Dice [%] ↑	Spleen	R.Kd	L.Kd	GB	Eso.	Liver	Sto.	Pan.	AG	Avg
nnU-Net[12]	88.1	93.0	**94.0**	73.4	89.1	79.8	90.2	73.0	69.7	83.3
UNETR[25]	93.3	91.5	92.2	68.8	95.5	83.3	90.2	65.8	55.3	81.7
SwinUNETR[7]	93.5	92.9	92.6	64.6	**96.1**	87.0	91.3	75.3	66.1	84.3
nnFormer[30]	95.0	93.1	92.9	72.1	**96.1**	84.1	89.9	69.2	63.1	83.9
UX-net[15]	**95.2**	92.9	93.3	64.4	**96.1**	84.9	**91.6**	**77.4**	64.8	84.5
Sam-Adapter[4]	85.9	83.5	83.2	55.7	91.4	76.3	83.7	51.2	40.1	72.3
LoRaMedNet[16]	78.8	65.8	79.9	56.0	78.4	44.6	74.1	46.3	35.6	62.2
3DSAM-adapter[6]	92.4	90.5	90.8	67.2	95.1	86.3	89.6	55.9	55.6	80.4
OG-SAM(Ours)	94.4	**93.2**	92.6	**80.1**	**96.1**	**88.9**	82.7	70.9	**70.3**	**85.5**
HD95 ↓	Spleen	R.Kd	L.Kd	GB	Eso.	Liver	Sto.	Pan.	AG	Avg.
nnU-Net[12]	1.94	**1.54**	**1.47**	7.39	52.23	7.07	1.99	**1.66**	3.91	8.80
UNETR[25]	2.23	2.23	2.00	10.44	2.44	9.05	1.41	3.00	4.47	4.14
SwinUNETR[7]	2.00	2.23	2.00	8.72	**2.23**	10.24	3.00	1.73	**3.00**	3.91
nnFormer[30]	46.01	3.03	3.93	17.32	62.58	25.33	25.82	3.60	11.70	22.14
UX-net[15]	**1.41**	2.00	1.73	9.48	2.44	10.44	**1.00**	1.73	120.53	16.75
SAM-Adapter[4]	2.00	86.26	20.33	6.08	3.74	128.88	1.41	2.23	118.00	40.99
LoRaMedNet[16]	105.66	177.87	132.31	7.49	135.69	163.51	154.03	140.66	6.70	113.77
3DSAM-adapter[6]	2.23	70.89	5.00	8.06	2.82	16.40	1.41	3.16	140.20	27.80
OG-SAM(Ours)	**1.41**	3.16	3.60	**3.00**	**2.23**	**4.00**	3.60	2.23	**3.00**	**2.91**

R.Kd: Right kidney, L.Kd: Left kidney, GB: Gall bladder, Eso.: Esophagus, Sto: Stomach, Pan: Pancreas, AG: Adrenal gland.

reduces the HD95 score (SwinUNETR) by 1.00, suggesting our balanced model performance due to our GOF module that aggregates features at multiple levels, maintaining anatomical consistency.

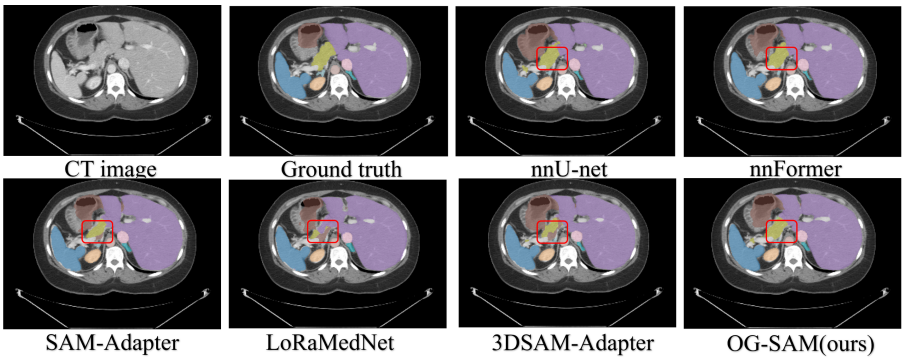

Fig. 2: Qualitative segmentation visualization on the BTCV dataset.

Further, our method also surpasses competing methods in achieving the best performance in more organs, measured by both Dice and HD95 scores, demon-

Table 2: Ablation studies on the different components

Dice [%] ↑	Spleen	R.Kd	L.Kd	GB	Eso.	Liver	Sto.	Pan.	AG	Avg
w/o GoF	93.4	92.8	**92.8**	77.0	95.8	88.7	**90.6**	68.1	66.0	85.0
w/o Weight-sharing	87.0	92.1	92.1	**81.1**	94.1	88.2	89.0	**71.7**	66.2	84.6
w/o Split-adapter	84.4	86.6	91.0	74.5	93.5	80.1	81.6	68.1	63.1	80.3
Ours	**94.4**	**93.2**	92.6	80.1	**96.1**	**88.9**	82.7	70.9	**70.3**	**85.5**
HD95 ↓	Spleen	R.Kd	L.Kd	GB	Eso.	Liver	Sto.	Pan.	AG	Avg.
w/o GoF	2.00	**2.44**	**2.44**	3.31	**2.23**	4.69	**1.41**	3.14	5.19	2.98
w/o Weight-sharing	91.86	22.09	114.28	7.56	134.31	129.81	**1.41**	**2.00**	**5.00**	56.48
w/o Split-adapter	111.62	113.58	60.88	6.16	85.47	132.13	166.76	2.23	5.83	76.07
Ours	**1.41**	3.16	3.60	**3.00**	**2.23**	**4.00**	3.60	2.23	**3.00**	**2.91**

strating its balanced performance across diverse anatomical structures. Of note, our model performs the best in gallbladder and liver segmentation, consistently achieving the highest Dice and HD95 scores, suggesting our model could separate anatomically adjacent structures through the weight-sharing strategy. As SAM is trained on natural images, medical images with complex tissue structures, with noise and artefacts, make SAN less effective for organs with blurred boundaries and low contrast. Compared with these SAM-based methods, our method has better boundary segmentation performance, indicated by our low HD95 scores in Table 1. The visual comparison of all models is shown in Figure 2.

3.3 Ablation studies

To further explore the effectiveness of each component: (1) GoF; (2) Weight-sharing; (3) Split-adapter, we ablate them and compare the model overall performance. The experimental results in Table 2 show the performance decreases after the ablation, demonstrating the effectiveness of each proposed module. Notably, the model performance improves with a 73.16 increase in HD95 when the Organ Adaptor is ablated, compared to the GoF and Weight-sharing modules, suggesting the effectiveness of our Adaptor.

4 Conclusion

We introduce OG-SAM, a biologically inspired multi-organ segmentation framework, with OrganAdapt inspired by embryonic development, regulating parameter sharing and specialization among multiple organs, ensuring anatomically coherent segmentation with enhanced accuracy and robustness. The GoF module extracts multiscale feature pyramids from different layers and adaptively fuses features using organ-specific parameters, capturing organ-unique features. As a query-based plug-in, OG-SAM dynamically adjusts parameters for each organ, enabling a gated adaptation mechanism to enhance model performance. Our experiments show that our method surpasses competing methods and demonstrates its ability to dynamically adapt to the unique features of organs while

maintaining robustness across organs. Further, unlike conventional methods that require full retraining for new organs, our approach is inherently scalable and efficient, leveraging shared parameters to transfer knowledge across organs of common origin and development. Our model offers a new perspective on biologically inspired adaptation.

References

1. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
2. Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., et al.: Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. Medical Image Analysis **98**, 103310 (2024)
3. Chen, H., Zhang, H., Chan, U.W., Yin, R., Wang, X., Li, C.: Domain game: Disentangle anatomical feature for single domain generalized segmentation. In: International Workshop on Computational Mathematics Modeling in Cancer Analysis. pp. 41–51. Springer (2024)
4. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: Sam-adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023)
5. 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)
6. Gong, S., Zhong, Y., Ma, W., Li, J., Wang, Z., Zhang, J., Heng, P.A., Dou, Q.: 3dsam-adapter: Holistic adaptation of sam from 2d to 3d for promptable tumor segmentation. Medical Image Analysis **98**, 103324 (2024)
7. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
8. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
9. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
11. Huang, T., Yin, H., Huang, X.: Improved genetic algorithm for multi-threshold optimization in digital pathology image segmentation. Scientific Reports **14**(1), 22454 (2024)
12. 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)
13. La Macchia, M., Fellin, F., Amichetti, M., Cianchetti, M., Gianolini, S., Paola, V., Lomax, A.J., Widesott, L.: Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiation Oncology **7**, 1–16 (2012)

14. 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)
15. Lee, H.H., Bao, S., Huo, Y., Landman, B.A.: 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. arXiv preprint arXiv:2209.15076 (2022)
16. Li, K., Rajpurkar, P.: Adapting segment anything models to medical imaging via fine-tuning without domain pretraining. In: AAAI 2024 Spring Symposium on Clinical Foundation Models (2024)
17. 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)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D.: Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence **44**(7), 3523–3542 (2021)
20. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
21. Pfister, D.G., Spencer, S., Adelstein, D., Adkins, D., Anzai, Y., Brizel, D.M., Bruce, J.Y., Busse, P.M., Caudell, J.J., Cmelak, A.J., et al.: Head and neck cancers, version 2.2020, nccn clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network **18**(7), 873–898 (2020)
22. Ray, B., Mukhopadhyay, S., Hossain, S., Ghosal, S.K., Sarkar, R.: Image steganography using deep learning based edge detection. Multimedia Tools and Applications **80**(24), 33475–33503 (2021)
23. Ren, L., Zhao, D., Zhao, X., Chen, W., Li, L., Wu, T., Liang, G., Cai, Z., Xu, S.: Multilevel thresholding segmentation for pathological images: Optimal performance design of a new modified differential evolution. Computers in Biology and Medicine **148**, 105910 (2022)
24. 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)
25. Shaker, A.M., Maaz, M., Rasheed, H., Khan, S., Yang, M.H., Khan, F.S.: Unetr++: delving into efficient and accurate 3d medical image segmentation. IEEE Transactions on Medical Imaging (2024)
26. Van Ginneken, B., Schaefer-Prokop, C.M., Prokop, M.: Computer-aided diagnosis: how to move from the laboratory to the clinic. Radiology **261**(3), 719–732 (2011)
27. Wang, H., Qiao, X., Ding, W., Chen, G., Miao, Y., Guo, R., Zhu, X., Cheng, Z., Xu, J., Li, B., et al.: Robust and generalizable artificial intelligence for multi-organ segmentation in ultra-low-dose total-body pet imaging: a multi-center and cross-tracer study. European Journal of Nuclear Medicine and Molecular Imaging pp. 1–15 (2025)
28. Wang, Z., Li, Z., Sun, J., Xu, Y.: Selective convolutional features based generalized-mean pooling for fine-grained image retrieval. In: 2018 IEEE Visual Communications and Image Processing (VCIP). pp. 1–4. IEEE (2018)

29. Yu-Qian, Z., Wei-Hua, G., Zhen-Cheng, C., Jing-Tian, T., Ling-Yun, L.: Medical images edge detection based on mathematical morphology. In: 2005 IEEE engineering in medicine and biology 27th annual conference. pp. 6492–6495. IEEE (2006)
30. Zhou, H.Y., Guo, J., Zhang, Y., Yu, L., Wang, L., Yu, Y.: nnformer: Interleaved transformer for volumetric segmentation. arXiv preprint arXiv:2109.03201 (2021)
31. Zhou, T., Li, L., Bredell, G., Li, J., Unkelbach, J., Konukoglu, E.: Volumetric memory network for interactive medical image segmentation. Medical Image Analysis **83**, 102599 (2023)