



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

A Unified Few-Shot Framework for Multi-Atlas Neuroimaging Segmentation Leveraging Vision Foundation Models

Bocheng Guo^{1*}, Xiyuan Zhang^{1*}, Wei Zhang¹, Ofer Pasternak², Lauren O'Donnell², and Fan Zhang^{1**}

¹ University of Electronic Science and Technology of China
guobocheng9@gmail.com, mail_Xiyuan_Zhang@126.com, zhangwei@uestc.edu.cn, zhangfanmark@gmail.com

² Harvard Medical School, Boston, USA
{ofer, odonnell}@bwh.harvard.edu

Abstract. Accurate brain segmentation across diverse atlas protocols is essential for neuroimaging, yet it is hindered by the high cost of generating large-scale ground truth labels and the prohibitive overhead of re-training models for new protocols. While vision foundation models offer powerful representations, their seamless adaptation to 3D multi-atlas segmentation in few-shot scenarios remains challenging. In this study, we propose a unified few-shot framework that adapts large-scale vision models to neuroimaging via an efficient "pre-train, fine-tune and inference" paradigm. By coupling anatomically-prioritized self-supervision with an $O(N)$ Linear Transformer, our model learns a universal, semantically consistent latent space for dense semantic label transfer. Instead of exhaustive re-training, we introduce a "Lock-and-Finetune" strategy that exclusively updates a lightweight refinement module to recover fine structural boundaries. Experiments on the Human Connectome Project (HCP) dataset show our approach achieves over 95% of fully supervised nnU-Net performance using less than 5% of labeled data. Notably, the model adapts to unseen atlases using only a few samples, providing an efficient solution for neuroimaging segmentation.

Keywords: Multi-Atlas · Brain Segmentation · Few-shot Learning.

1 Introduction

Quantitative analysis of brain MRI serves as a cornerstone of modern neuroscience and clinical neurology [1]. Accurate anatomical segmentation of brain structures in structural T1-weighted (T1w) data is critical for disease diagnosis, surgical planning, and longitudinal morphometry [2–4]. While atlas-based methods like FreeSurfer [5] have been the mainstream standard for decades, their

* These authors contributed equally to this work.

** The corresponding author of this manuscript is Fan Zhang.

reliance on iterative deformable models is slow and computationally expensive, limiting their application in large-scale cohorts and real-time pipelines [6].

Deep learning methods, such as FastSurfer [7] and nnU-Net [8], have largely accelerated the inference process, yet they still face challenges when handling diverse protocols. SynthSeg [11] improves robustness to MRI contrast and resolution variations, but remains tied to a predefined anatomical label protocol. Neuroimage segmentation relies on a vast array of brain atlases, each defined by unique topological rules [9]. Adapting fully supervised models to a new atlas requires constructing large-scale ground truth datasets [10] and incurring prohibitive retraining costs, which severely bottlenecks large-scale, multi-protocol neuroimaging analysis.

Vision foundation models such as SAM [12, 13] and DINO [14, 15] show powerful 2D feature representation capabilities, offering a promising pathway to break this bottleneck. For instance, MedDINOv3 [16] successfully adapted DINOv3 for medical image segmentation through large-scale pre-training on 3.87 million CT slices. However, transferring 2D models to 3D neuroimaging faces a dimension gap: slice-by-slice inference disrupts Z-axis spatial context and topological continuity, causing inconsistent inter-slice segmentations [17].

To address these challenges, we propose a 3D few-shot brain segmentation framework that adapts vision foundation models to T1w images. Specifically, our model utilizes an $O(N)$ linear transformer matcher to efficiently propagate anatomical labels from just a few support samples to query images via dense voxel-wise correlation. To bridge the dimensional gap between 2D foundation models and 3D neuroimaging, we integrate anatomically-prioritized spatial-aware masked autoencoding (SA-MAE) with a 3D ConvNeXt Adapter. Finally, to ensure practical multi-atlas deployment, our resource-efficient "Lock-and-Finetune" strategy freezes the heavy backbone and matching networks to exclusively update a lightweight Super-Resolution Refiner. This enables efficient adaptation to unseen atlases using minimal labeled data under low computational overhead.

2 Methodology

2.1 Overall Architecture

To enable few-shot segmentation across diverse neuroimaging atlases, our framework reformulates the task from atlas-specific voxel classification to a generalizable cross-subject feature matching process (Fig. 1). The workflow operates through three distinct phases:

1) Two-Stage Joint Pre-training: A 2D ConvNeXt encoder [18] first learns 3D structural priors via *Spatial-Aware MAE* and semantic anchoring. Subsequently, a 3D Adapter and a *Linear Transformer Matcher* are trained to perform voxel-wise label transfer with $O(N)$ complexity.

2) Resource-Efficient Fine-tuning: For a new atlas, the output channels of the segmentation head inside the *Super-Resolution Refiner* are reset according to the target label count, while the backbone and matcher remain frozen.

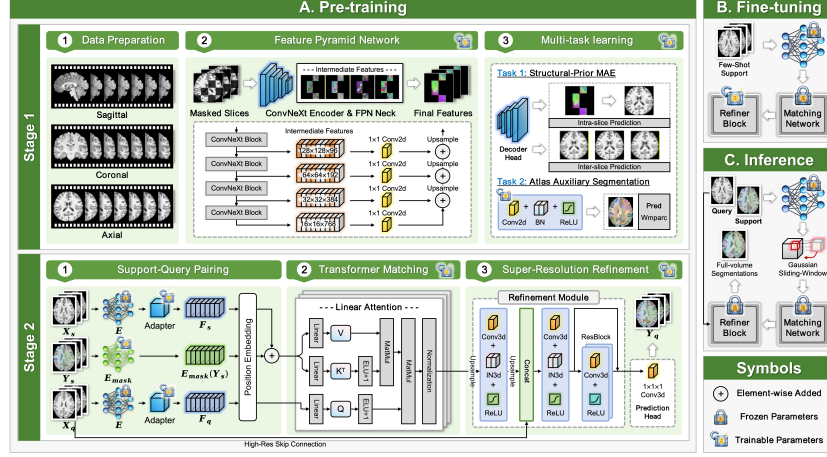


Fig. 1. Overview of the proposed framework. **(A) Training:** Encompasses **Stage I** (2.5D anatomical pre-training using SA-MAE and atlas anchoring) and **Stage II** (3D matching training featuring the ConvNeXt Adapter, Linear Transformer Matcher, and SR Refiner). **(B) Finetuning:** The "Lock-and-Finetune" strategy for rapid few-shot atlas adaptation. **(C) Inference:** Gaussian-weighted volumetric inference for seamless whole-brain segmentation.

3) Seamless Volumetric Inference: Finally, a Gaussian-weighted sliding window scheme aggregates 3D patch predictions, eliminating stitching artifacts and ensuring topologically consistent whole-brain parcellation.

2.2 Stage I: Anatomically-Prioritized Representation Learning

To internalize a generalized 3D brain topology, we pre-train a DINOv3-initialized ConvNeXt encoder E equipped with an FPN neck.

Spatial-aware masked autoencoding (SA-MAE): We propose a 2.5D neighborhood reconstruction task to instill volumetric awareness. Given a central slice (75% masked), the model reconstructs a volumetric context $X \in \mathbb{R}^{7 \times H \times W}$ comprising the target slice and its six immediate neighbors. The encoder receives only the masked central slice as input; the six neighboring slices are used as reconstruction targets rather than additional input channels. To preserve both intensity fidelity and sharp boundaries, we employ a composite L_1 -SSIM loss:

$$L_{recon} = \lambda_{l1} \|X - X_{rec}\|_1 + \lambda_{ssim} (1 - \text{SSIM}(X, X_{rec})) \quad (1)$$

Anatomical anchoring: To supplement low-level textures with high-level semantic partitions, an auxiliary segmentation head is trained on a reference atlas. This anchors the latent space to neuroanatomical boundaries. The joint objective for Stage I is:

$$L_{pre} = L_{recon} + \lambda_{seg} \sum L_{CE}(Y_{atlas}, P_{atlas}) \quad (2)$$

where L_{CE} is the Cross-Entropy loss between logits P_{atlas} and ground truth Y_{atlas} . This synergy ensures the encoder captures fine-grained textures alongside global 3D consistency.

2.3 Stage II: 3D Associative Matching for Label Transfer

In the second stage, we freeze the pre-trained encoder E and introduce an associative matching network designed for cross-subject label transfer. Given a support image X_s with its corresponding label mask Y_s , and a query image X_q , the network learns to propagate anatomical semantics through voxel-wise correlation.

3D feature adaptation: To bridge the dimensional gap, a 3D ConvNeXt Adapter projects multi-scale 2D features into unified volumetric embeddings F_s and F_q . By employing large-kernel 3D depth-wise convolutions ($3 \times 7 \times 7$) and LayerNorm, the module captures long-range spatial dependencies along the depth axis, effectively converting pseudo-3D features from Stage I into true 3D representations.

Linear transformer for dense matching: To overcome the $\mathcal{O}(N^2)$ complexity of standard 3D attention, we employ a Linear Transformer ($\mathcal{O}(N)$ complexity) for dense voxel-wise matching. Support labels Y_s are encoded via a lightweight CNN E_{mask} and fused with support features F_s . With learnable 3D positional embeddings P , the matching memory and target are defined as:

$$\text{Mem} = F_s + E_{mask}(Y_s) + P, \quad \text{Tgt} = F_q + P \quad (3)$$

To formulate the associative matching mechanism, the query (Q) is derived from the target volume, while the keys (K) and values (V) are constructed from the support memory (i.e., $Q = \text{Tgt}$, and $K = V = \text{Mem}$). The matching process utilizes a kernel-based approximation to decompose the standard softmax computation. For the i -th query token, the attention is computed as:

$$\text{Attn}(Q, K, V)_i = \frac{\phi(Q_i) \sum_{j=1}^M (\phi(K_j)^\top V_j)}{\phi(Q_i) \sum_{j=1}^M \phi(K_j)^\top} \quad (4)$$

where i and j index the spatial tokens of the query and memory volumes respectively, M denotes the total number of memory tokens, and $\phi(x) = \text{ELU}(x) + 1$ serves as the positive feature map. This associative mechanism allows the query features to efficiently retrieve anatomical semantics from the support memory, yielding a densely matched feature volume $F_{matched}$.

Image-guided super-resolution refinement: To restore boundaries lost during low-resolution matching (1/4 scale), the Super-Resolution Refinement Module upsamples $F_{matched}$ via trilinear interpolation and 3D residual blocks. We fuse these features with the raw high-resolution MRI X_q through channel concatenation. This leverages high-frequency structural signals from the original image to guide fine-grained boundary delineation for the final high-fidelity segmentation Y_q . Since the segmentation head resides in the SR Refiner, varying target label counts only affect this lightweight head, keeping the backbone and matcher frozen.

2.4 Efficient Atlas Adaptation via "Lock-and-Finetune"

To bridge the gap between experimental models and practical deployment, we design a resource-efficient adaptation strategy and a seamless volumetric reconstruction scheme.

We propose a "Lock-and-Finetune" strategy for rapid adaptation. By freezing the pre-trained backbone and matcher to preserve anatomical priors, we update only the lightweight Super-Resolution Refiner using a few annotated samples to calibrate local boundaries. This adaptation is computationally efficient, requiring about 0.3 h and 4 GB GPU memory on an RTX 3090, compared with approximately 2 h and 22 GB for nnU-Net adaptation.

2.5 Volumetric Inference

For whole-brain parcellation, we adopt a patch-based sliding window approach. To eliminate stitching artifacts, a Gaussian-weighted accumulation scheme is applied, where overlapping predictions $\hat{y}_i$ are aggregated via: $Y_{final} = (\sum w_i \cdot \hat{y}_i) / \sum w_i$, with w_i representing the Gaussian confidence weights. This ensures global geometric consistency and suppresses edge uncertainties.

3 Experiments

3.1 Datasets and Evaluation Metrics

We evaluate our framework on the HCP 3T structural preprocessed dataset [19]. All anatomical labels are standardized via the FreeSurfer pipeline. All methods are evaluated on the same HCP preprocessed data, with identical training subjects and the same validation set (n=25), to ensure consistent comparison across methods. **Stage I** uses the *wmparc* atlas (covering white matter, sub-cortical structures, and Desikan-Killiany regions) [20] as the anatomical anchor. **Stage II** incorporates *ribbon*, *wmparc*, and *aparc* segmentation atlases provided in FreeSurfer to optimize the associative matching network for label transfer. To assess robustness across varying anatomical granularities, we evaluate cross-atlas adaptation in a *coarse-to-fine* manner: (1) the functional Yeo7 atlas (14 labels) [22]; (2) the *aparc.a2009s+aseg* protocol (148 labels) [21]; and (3) the multi-modal HCP-MMP1 atlas (360 labels) [23]. This selection validates our matching and refinement modules across a broad spectrum of label densities and parcellation logic. Performance is quantified by the Dice Similarity Coefficient (DSC) to assess volumetric overlap.

3.2 Implementation and Baselines

We compare our framework against fully supervised baselines (nnU-Net, Medical SAM2 [24], FastSurfer) using 100 subjects and few-shot models (SAM3-Adapter [25], DSPNet [26]) using 5 subjects. Training (Stages I & II) is conducted on four NVIDIA RTX 3090 GPUs using an AdamW optimizer ($lr =$

Table 1. Quantitative comparison of segmentation performance (Dice Similarity Coefficient) across validation atlases. Results are presented as mean ($\pm$ standard deviation). The number of training subjects is indicated in parentheses.

Task	Fully Supervised ($n = 100$)			Few-Shot ($n = 5$)				
	nnU-Net	Medical SAM2	Fast-Surfer	nnU-Net	Medical SAM2	SAM3-Adapter	DSPNet	Ours
aparc2009 (gm)	0.771 ± 0.009	0.744 ± 0.009	0.782 ± 0.010	0.448 ± 0.013	0.512 ± 0.012	0.693 ± 0.009	0.475 ± 0.020	0.758 ± 0.003
yeo7	0.819 ± 0.010	0.803 ± 0.008	0.807 ± 0.003	0.594 ± 0.011	0.549 ± 0.014	0.711 ± 0.004	0.610 ± 0.015	0.797 ± 0.004
hcpmmp1	0.651 ± 0.011	0.601 ± 0.013	0.643 ± 0.009	0.259 ± 0.017	0.390 ± 0.026	0.583 ± 0.009	0.378 ± 0.029	0.632 ± 0.005
aparc2009 (wm)	0.855 ± 0.002	0.849 ± 0.004	0.868 ± 0.003	0.791 ± 0.006	0.807 ± 0.007	0.809 ± 0.005	0.662 ± 0.008	0.864 ± 0.002

1×10^{-4} , weight decay 0.05). Downstream adaptation follows our Lock-and-Finetune strategy on a single RTX 3090, optimizing only the SR Refiner for 5 iterations. For the few-shot setting, nnU-Net and Medical SAM2 are trained from scratch using the same five subjects, SAM3-Adapter follows its pretrain-finetune protocol, and DSPNet is evaluated without pretraining as its framework does not support it.

3.3 Quantitative Results and Analysis

Table 1 presents the quantitative evaluation results across the validation atlases. Here, we separately report the gray-matter and white-matter components of the aparc.a2009s+aseg atlas as aparc2009 (gm) and aparc2009 (wm), respectively. Under the restricted 5-subject few-shot setting, standard methods such as nnU-Net and Medical SAM2 exhibit substantial performance degradation. In contrast, our framework maintains high segmentation accuracy, a stability primarily attributed to the strong structural priors embedded by Stage I Anatomical Anchoring and Stage II Associative Matching. Additionally, pretraining nnU-Net on aparc and fine-tuning its final layer for each target atlas yields Dice scores of 0.704 on Yeo7 and 0.471 on HCP-MMP1, compared with 0.797 and 0.632 from our method.

Furthermore, utilizing data from only 5 subjects, our method achieves performance comparable to fully supervised baselines trained on 100 subjects. On structurally complex atlases like HCP-MMP1 and aparc.a2009s+aseg, the performance gap with the 100-subject nnU-Net is marginal. This indicates that the Super-Resolution Refiner effectively leverages high-frequency MRI signals to mitigate the absence of large-scale annotated data. The consistent accuracy on atlases unseen during pre-training (Yeo7 and HCP-MMP1) further demonstrates effective cross-atlas adaptation under an anatomical-prior and limited-annotation setting.

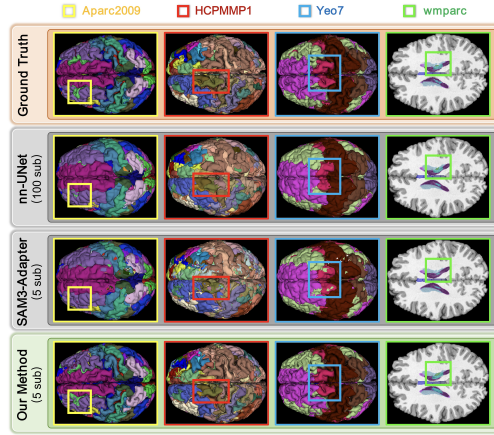


Fig. 2. Qualitative comparison of segmentation results. Compared to fully supervised nnU-Net (100 subjects) and few-shot SAM3-Adapter (5 subjects), our method (5 subjects) demonstrates superior topological consistency and precise boundary delineation across diverse atlas tasks, closely approximating the manual ground truth.

The visual comparisons in Fig. 2 further corroborate the robust performance of our framework. Under the limited data setting of only 5 subjects, the SAM3-Adapter frequently yields fragmented predictions or fails to capture complex cortical folding patterns. In contrast, our method maintains high topological consistency. Even when relying on merely 5 support samples, the masks generated by our framework are visually comparable to the ground truth.

Specifically, in the cortical (aparc.a2009s) and multi-modal (HCP-MMP1) tasks, our model successfully delineates narrow gyri and sulci. This high level of precision is directly attributed to the Super-Resolution Refinement module. Furthermore, the consistent performance across functional (Yeo7) and structural (white matter) tasks suggests that the Linear Transformer Matcher learns stable cross-atlas anatomical correspondences.

3.4 Qualitative Feature Analysis and Interpretability

To evaluate the representation power of our pre-trained encoder, we visualize the latent space using UMAP [27]. High-dimensional features from unseen MRI slices are projected into a 3D RGB space.

As shown in Fig. 3, the pseudo-color maps exhibit semantic consistency: voxels belonging to similar anatomical structures are mapped to coherent color manifolds without target-atlas fine-tuning, while Stage I anchor-atlas supervision shapes the feature space. This suggests that raw signals have been effectively mapped onto a manifold characterized by neuroanatomical semantics. Second, the sharp color transitions at tissue interfaces reflect the model’s exceptional High-Fidelity Boundary Preservation, qualitatively validating the effectiveness of

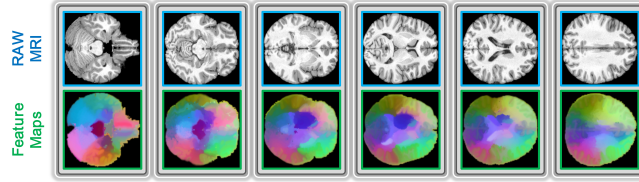


Fig. 3. UMAP feature-space visualization of the pre-trained encoder mapped to RGB channels. The learned latent space exhibits semantic consistency, sharp anatomical boundaries, and 3D topology awareness prior to target-atlas fine-tuning.

the SA-MAE objective and SSIM loss in capturing high-frequency morphological details. Furthermore, the feature maps display a spatial hierarchy that aligns closely with known 3D brain topology. This Anatomical Topology Awareness indicates that the Anatomical Anchoring in Stage I has successfully internalized generalized neuroanatomical priors within the backbone.

3.5 Ablation Study

To investigate the specific contributions of each core component within our framework, we conducted ablation experiments on the *aparc.a2009s+aseg* validation set, with the results detailed in Table 2.

Table 2. Ablation study of the proposed components on the validation set. Metrics are reported as Mean ($\pm$ SD). The best results are highlighted in bold.

Method	Dice $\uparrow$	HD95 (mm) $\downarrow$
Our Full Method	0.758 ± 0.0031	1.9916 ± 0.12
w/o Anatomical Anchoring	0.704 ± 0.0047	2.3870 ± 0.14
w/o 3D ConvNeXt Adapter	0.719 ± 0.0053	2.4190 ± 0.15
w/o Super-Resolution Refinement	0.733 ± 0.0051	2.6310 ± 0.14

Anatomical Anchoring (Stage I): Replacing pure reconstruction with atlas-based auxiliary supervision yielded a significant performance gain. This confirms that explicit structural priors are essential for aligning the latent space with biological boundaries, especially in low-contrast regions.

3D ConvNeXt Adapter: Substituting 2D feature stacking with our 3D Adapter substantially improved the Dice score. The $3 \times 7 \times 7$ depthwise convolutions effectively capture inter-slice dependencies, proving that dedicated cross-dimensional projection is critical for maintaining 3D spatial continuity.

Super-Resolution Refinement: Fusing matched features with raw MRI signals in the SR Refiner dramatically reduces HD95, recovering fine structural boundaries lost during downsampled matching. Furthermore, ablations on Yeo7/HCP-MMP1 show that removing Stage-I anatomical supervision, Stage-I reconstruc-

tion, or the Stage-II matcher decreases Dice from 0.79/0.63 to 0.76/0.55, 0.63/0.47, and 0.71/0.42, respectively. These results support Lock-and-Finetune as a practical trade-off between accuracy and computational cost for few-shot adaptation.

4 Conclusion

In this paper, we present a unified few-shot framework for multi-atlas neuroimaging segmentation that combines anatomical feature anchoring, efficient cross-dimensional matching, and lightweight target-atlas adaptation. Experiments across atlases of varying granularity show that the proposed approach alleviates the dependence on large-scale annotations and reduces retraining costs for new parcellation protocols, establishing a highly efficient paradigm for foundation model adaptation in neuroimaging research.

Acknowledgments. This work is in part supported by the National Key R&D Program of China (No. 2023YFE0118600), the National Natural Science Foundation of China (No. 62371107), and Science and Technology Department of Sichuan Province (No. 2026YFHZ0045).

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

References

1. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. *Journal of Digital Imaging*, **30**(4), 449–459.
2. Despotović, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, **2015**(1), 450341.
3. Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., ... & Buckner, R. L. (2009). The cortical signature of Alzheimer’s disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex*, **19**(3), 497–510.
4. Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., & Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, **3**(3), 165–189.
5. Fischl, B. (2012). FreeSurfer. *Neuroimage*, **62**(2), 774–781.
6. Makropoulos, A., Robinson, E. C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., ... & Rueckert, D. (2018). The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, **173**, 88–112.
7. Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., & Reuter, M. (2020). Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*, **219**, 117012.

8. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, **18**(2), 203–211.
9. Eickhoff, S. B., Yeo, B. T., & Genon, S. (2018). Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, **19**(11), 672–686.
10. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, **63**, 101693.
11. Billot, B., Greve, D. N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., ... & Iglesias, J. E. (2023). SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*, **86**, 102789.
12. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015–4026).
13. Carion, N., Gustafson, L., Hu, Y. T., Debnath, S., Hu, R., Suris, D., ... & Feichtenhofer, C. (2025). Sam 3: Segment anything with concepts. *arXiv preprint arXiv:2511.16719*.
14. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., ... & Bojanowski, P. (2023). Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*.
15. Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., ... & Bojanowski, P. (2025). Dinov3. *arXiv preprint arXiv:2508.10104*.
16. Li, Y., Wu, Y., Lai, Y., Hu, M., & Yang, X. (2025). MedDINOv3: How to adapt vision foundation models for medical image segmentation?. *arXiv preprint arXiv:2509.02379*.
17. Wald, T., Roy, S., Isensee, F., Ulrich, C., Ziegler, S., Trofimova, D., ... & Maier-Hein, K. (2025). Primus: Enforcing attention usage for 3d medical image segmentation. *arXiv preprint arXiv:2503.01835*.
18. Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11976–11986).
19. Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn human connectome project: an overview. *Neuroimage*, **80**, 62–79.
20. Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, **31**(3), 968–980.
21. Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, **53**(1), 1–15.
22. Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*.
23. Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, **536**(7615), 171–178.
24. Zhu, J., Hamdi, A., Qi, Y., Jin, Y., & Wu, J. (2024). Medical sam 2: Segment medical images as video via segment anything model 2. *arXiv preprint arXiv:2408.00874*.

25. Chen, T., Cao, R., Yu, X., Zhu, L., Ding, C., Ji, D., ... & Zang, Y. (2025). SAM3-Adapter: Efficient Adaptation of Segment Anything 3 for Camouflage Object Segmentation, Shadow Detection, and Medical Image Segmentation. *arXiv preprint arXiv:2511.19425*.
26. Tang, S., Yan, S., Qi, X., Gao, J., Ye, M., Zhang, J., & Zhu, X. (2025). Few-shot medical image segmentation with high-fidelity prototypes. *Medical Image Analysis*, **100**, 103412.
27. McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.