Knowledge-Guided Prompt Learning for Lifespan Brain MR Image Segmentation

Lin Teng[1], Zihao Zhao[1], Jiawei Huang[1], Zehong Cao[2], Runqi Meng[1], Feng Shi[2(✉)], and Dinggang Shen[1,2,3(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Shanghai United Imaging Intelligence Co., Ltd. Shanghai 200230, China
feng.shi@uii-ai.com
[3] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
dgshen@shanghaitech.edu.cn

Abstract. Automatic and accurate segmentation of brain MR images throughout the human lifespan into tissue and structure is crucial for understanding brain development and diagnosing diseases. However, challenges arise from the intricate variations in brain appearance due to rapid early brain development, aging, and disorders, compounded by the limited availability of manually-labeled datasets. In response, we present a two-step segmentation framework employing **K**nowledge-**G**uided **P**rompt **L**earning (KGPL) for brain MRI. Specifically, we first pre-train segmentation models on large-scale datasets with sub-optimal labels, followed by the incorporation of knowledge-driven embeddings learned from image-text alignment into the models. The introduction of knowledge-wise prompts captures semantic relationships between anatomical variability and biological processes, enabling models to learn structural feature embeddings across diverse age groups. Experimental findings demonstrate the superiority and robustness of our proposed method, particularly noticeable when employing Swin UNETR as the backbone. Our approach achieves average DSC values of 95.17% and 94.19% for brain tissue and structure segmentation, respectively. Our code is available at https://github.com/TL9792/KGPL.

Keywords: Brain MRI segmentation · Across the lifespan · Knowledge-guided prompt learning · Fine tuning · Transfer learning.

1 Introduction

Dense segmentation of brain MR images involves voxel-wise labeling from different brain tissues, *i.e.*, white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), into fine-grained cortical and subcortical subregions, such as 106 structures according to Desikan-Killiany (DK) atlas [4] and Schaefer 400-parcels as described in [20]. Lifespan brain segmentation is essential for quantitative brain

---

Lifespan refers to the period between birth and death, emphasizing the wide age range covered in our study.

development analysis and neurodegenerative disorders diagnosis [9]. Manual annotation is time-consuming and labor-intensive. Hence, an automated brain image segmentation method is urgently needed.

Several software packages, such as FMRIB Software Library (FSL) [21] and FreeSurfer [8], have been developed to automatically segment brain images. However, processing each subject using these tools costs several hours, and the obtained results often require further manual corrections by experienced experts. Recent advancements in deep learning, such as convolutional neural network (CNN) based U-Net [16], Vision Transformer (ViT) based methods [7] like UN-ETR [11], and Swin Transformer based [15] Swin UNETR [10], have shown promise in medical image segmentation.

Brain morphology is dynamic in the human lifespan, leading to considerable appearance variations in brain images [22]. For instance, CSF exhibits gradual expansion, while GM experiences progressive reduction with aging, particularly among elderly individuals with Alzheimer's disease (AD). However, most previous efforts are established on a single time point, and primarily focus on image-based information while overlooking the complementary textual information that provides essential contextual details and insights into the pathology or anatomy, which often results in lacking generalizability over a range of ages [6,25]. Additionally, the performance of learning-based methods is often restricted by the scarcity of annotated datasets, especially when labeled training samples for specific time points are limited in practice. To solve this problem, transfer learning [17] and fine-tuning [19,23] strategies have been widely adopted. Jia *et al.* [13] proposed an efficient fine-tuning method called Visual Prompt Tuning (VPT), which incorporates a small set of random learnable parameters into the input space. Directly applying random parameters representing unknown information may hinder model training [14].

To tackle the challenges mentioned above, we present a two-step framework with knowledge-guided prompt learning (KGPL) tailored for the dense segmentation of brain MR images across diverse age groups, which are **the first** to incorporate knowledge-wise prompts into brain segmentation models. Specifically, drawing inspiration from VPT, we replace random embeddings with knowledge-driven embeddings, learned from the pre-trained text encoder in BiomedCLIP, into the input space of the encoder. The knowledge-wise learnable embeddings exploit the relationship between images and brain structural and developmental-specific biomedical characteristics, aiming to empower models to learn distinguishing representations of brain data acquired from different age phases. Experimental results demonstrate that our proposed method outperforms state-of-the-art (SOTA) baselines, and performs robustly in different backbone networks.

BiomedCLIP indicates Biomedical Contrastive Language-Image Pre-training (CLIP) [18, 26], which is pre-trained on large-scale figure-caption pairs by jointly training an image encoder and a text encoder, being able to well establish the semantic connection between images and language [27].

2   Methodology

We aim to develop an automatic and robust approach to segment brain structure, based on brain MRI and specific biomedical attribute information. Fig. 1 illustrates the overall framework. Specifically, we first pre-train segmentation models on large-scale datasets with sub-optimal annotation (Sec. 2.1). Subsequently, we refine the pre-trained models on relatively small-scale manually labeled datasets by introducing a few knowledge-driven learnable parameters into the input space, aiming to guide models to extract more crucial features of brain data spanning a large range of age. (Sec. 2.2).
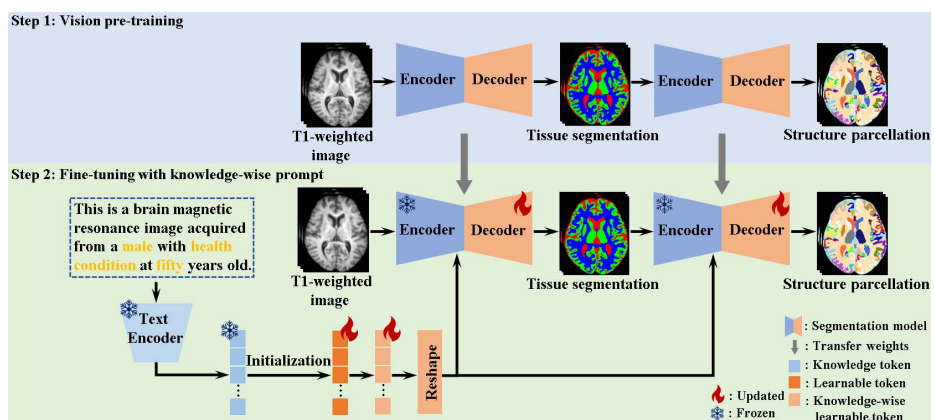

Fig. 1. Overview of our proposed two-step segmentation framework, incorporating Knowledge-Guided Prompt Learning (KGPL) for brain MRI across the lifespan. The top part shows the vision pre-training on the source domain, and the bottom part shows the refinement of the pre-trained models with knowledge-wise prompts on the target domain. Here, we adopt the weights learned from the source domain to initialize the model of the target domain, by only updating the learnable parameters and the decoder while freezing the encoder.

2.1   Vision Pre-training

In this section, we utilize datasets with sub-optimal labels to establish models for segmenting brain MR images into tissue and structure, respectively. Specifically, we select one cutting-edge segmentation model as the backbone. Using the T1-weighted MR image as input, we first train the model to predict the brain tissue segmentation mask, which is supervised by the Dice loss function formulated as Eq. (1). Subsequently, the predicted tissue segmentation results as input to train another model for brain structure parcellation, which is supervised by both Dice and Focal loss functions as expressed in Eq. (2).

$$\text{Dice Loss} = 1 - \frac{2 \times |P \cap G|}{|P| + |G|}, \tag{1}$$

where P is the prediction result, and G denotes the ground truth.

$$\text{Loss} = \text{Dice Loss} - \alpha \times (1 - p_t)^{\gamma} \times \log(p_t), \tag{2}$$

where α and γ are set to 100 and 0.2 based on experience. p_t is the probability of the predicted result.

2.2   Fine-tuning with Knowledge-wise Prompt

To utilize specific textual information reflecting brain properties to enhance the discriminative learning of fine-grained brain anatomical structures, we introduce a novel KGPL approach. The procedure is illustrated at the bottom of Fig. 1.

Knowledge Prompt Generation. Brain morphology, such as cortical thickness, shows a strong correlation with biological attributes like brain age, sex, and disease status [3]. This textual information comes from the corresponding data acquisition centers and official websites. To leverage this prior knowledge for guiding feature extraction, we initially categorize subjects into age groups every decade, considering that brain morphology remains relatively similar at close ages. Then individual attribute details are embedded into a template to generate a descriptive sentence (*e.g.*, "This is a brain magnetic resonance image acquired from a **male** with **mild cognitive impairment** at **fifty** years old"). Finally, this sentence is encoded by a pre-trained text encoder from Biomed-CLIP to obtain knowledge embeddings with dimensions (B, N, D), where B is the batch size, N is identical to the number of word tokens, and D is a fixed hidden dimension of 768.

Learnable Prompt Pre-initialized by Knowledge Prompt. To effectively integrate prior knowledge into models, we first pre-define multiple learnable tokens (*i.e.*, learnable embeddings) with zero values, matching with the dimension of knowledge embeddings. Then, the learnable embeddings are pre-initialized by knowledge embeddings through addition operation, referring to knowledge-wise learnable embeddings.

Model Fine-tuned Guided by Knowledge-wise Prompt. Inspired by VPT, we serve knowledge-wise learnable embeddings as input to deep-level encoder layers, considering knowledge features are considered high-level. Specifically, we denote the collection of p knowledge-wise learnable tokens as $P = \{P_k \in \mathbb{R}^M || k \in \mathbb{N}, 1 \le k \le p\}$, where p is the number of word tokens. To ensure that they can interact with image embeddings, we reshape them on the hidden dimension. For U-Net or Swin UNETR as the backbone, knowledge-wise learnable embeddings are reshaped via the Adaptive Average Pooling (AAP) operator,

transitioning from (B, N, D) to $(B, N, 1)$. Following this, they pass through a linear layer with shape (B, C, N), where C is the number of channels; For UN-ETR, knowledge-wise learnable embeddings are first transposed from (B, N, D) to (D, N, B), then processed by a linear layer to match image embeddings' shape, except for the sequence length dimension. Next, they are inverse-transposed to restore the original dimension order. Finally, they are concatenated with image embeddings to calculate information interaction. The shape is $(B, C, (N + S))$, in which S represents the dot product of the feature map's length, width, and height. The procedure can be formulated as:

$$[X_i, \_] = L_{i-1}([P_{i-1}, X_{i-1}]), \tag{3}$$

$$[X_{i+1}, \_] = L_i([P_i, X_i]), \tag{4}$$

where $X_i \in \mathbb{R}^{B \times C \times S}$ is image embeddings extracted by the *i-1*-th encoder layer L_{i-1}. P_i denotes knowledge-wise learnable embeddings at L_i's input space. During the fine-tuning step, only the image embeddings are directed to the decoder. Notably, we initialize the model of the target domain using weights learned from the source domain, by updating only the learnable parameters and the decoder while freezing the encoder.

3 Experiments and Results

3.1 Datasets and Metrics

Our datasets covering various age stages are collected from multiple repositories, including CBMFM (643 subjects, aged 18-81), Alzheimer's Disease Neuroimaging Initiative (ADNI) (1285 subjects, aged 20-97) [12], Autism Brain Imaging Data Exchange (ABIDE) (1062 subjects, aged 9-64) [5], and Adolescent Brain Cognitive Development (ABCD) dataset (458 subjects, aged 9-11) [2]. Detailed analysis of demographic information can be found in the supplementary material. Most datasets processed by FreeSurfer have sub-optimal labels. To obtain accurate labels for fine-tuning and model inference, FreeSurfer is initially applied, followed by manual annotations. Each dataset is randomly partitioned into training, validation, and testing sets in an 8:1:1 ratio. Each subject has a T1-weighted MR image, brain tissue segmentation mask, and structure segmentation mask (106 regions). T1-weighted MR images are pre-processed by skull stripping, bias field correction [24], and intensity normalization. Performance evaluation utilizes two quantitative metrics: Dice Similarity Coefficient (DSC) and Average Surface Distance (ASD).

3.2 Implementation Details

Experiments are conducted using MONAI [1] on a single NVIDIA TITAN RTX GPU with 24GB memory. Training spans 1000 epochs with early stopping, utilizing an AdamW optimizer with a learning rate of 1e-4 and regularization weight

of 1e-5. We adopt a linear warmup cosine annealing learning rate scheduler. Data augmentation involves randomly flipping along three axes with a probability of 0.5. Images are cropped to (128,128,128) based on foreground regions and a resolution of $1\times1\times1$ mm^3 using linear interpolation.

3.3   Quantitative Comparison Analysis

Effectiveness analysis of backbone architecture. We validate the expansibility of our method based on three representative networks: (1) U-Net, a pure CNN; (2) UNETR, combining CNN and ViT; and (3) Swin UNETR, blending CNN and Swin Transformer. Their original architectures are used in our study, *i.e.*, encoders have four CNN-based or twelve transformer-based blocks, decoders have four or five CNN-based blocks, and skip connections between the encoder and the decoder are included.

Table 1. The quantitative comparison results of brain tissue segmentation using different backbones, in terms of DSC and ASD (*: $p < 0.05$; ✓ denotes to use, and ✗ denotes not to use; The best results under each backbone are in **bold**).

Backbone	Random prompts	Knowledge prompts	Param.	CSF DSC (%)	CSF ASD (%)	GM DSC (%)	GM ASD (%)	WM DSC (%)	WM ASD
U-Net	✗	✗	90.31M	87.24±0.07*	28.14±0.71*	91.18±0.04*	27.55±0.13*	93.84±0.02*	25.26±0.05*
	✓	✗	14.79M	88.68±0.05*	26.28±0.12*	92.30±0.02*	25.70±0.07*	94.27±0.01*	23.78±0.03
	✗	✓	14.79M	**89.03±0.04**	**25.59±0.11**	**93.42±0.02**	**25.02±0.07**	**95.45±0.01**	**23.27±0.03**
UNETR	✗	✗	92.78M	88.24±0.07*	28.50±0.15*	90.09±0.05*	28.07±0.12*	92.20±0.03*	30.02±0.17*
	✓	✗	4.46M	89.43±0.05*	27.44±0.13*	91.19±0.03*	27.78±0.09*	93.99±0.04*	29.62±0.14*
	✗	✓	4.46M	**91.09±0.04**	**24.37±0.09**	**93.54±0.03**	**26.30±0.07**	**95.59±0.03**	**24.93±0.11**
Swin UNETR	✗	✗	62.53M	92.94±0.05*	17.22±0.03*	93.51±0.06*	29.96±0.10*	94.98±0.04*	30.87±0.19*
	✓	✗	54.47M	93.64±0.05*	16.18±0.03*	94.05±0.06*	27.79±0.13*	95.43±0.05*	27.40±0.21*
	✗	✓	54.47M	**94.22±0.04**	**13.01±0.02**	**94.98±0.03**	**22.71±0.05**	**96.32±0.01**	**21.36±0.07**

Tables 1 and 2 present quantitative comparison results of brain tissue and structure segmentation using three backbones, respectively. Due to page constraints, we show partial brain structure segmentation results. The complete comparison results are shown in the supplementary material. It is evident that our approach can significantly improve the performance using different backbones, especially with Swin UNETR obtaining a high prediction accuracy compared with other backbones. This indicates that our method is robust in applying to various network architecture types.

Comparison with full fine-tuning methods. Quantitative analysis of Tables 1 and 2 shows that KGPL using Swin UNETR as the backbone improves average DSC and ASD values by 1.36% and 6.99%, 1.15% and 6.78% for brain tissue and structure segmentation, respectively. Statistical significance of this improvement is confirmed by a paired t-test (p-value < 0.05). Notably, our method achieves optimal performance with minimal trainable parameters, particularly

Table 2. The quantitative comparison results of brain structure parcellation based on different backbones, in terms of DSC and ASD (*: $p < 0.05$; ✓denotes to use, and ✗denotes not to use; The best results under each backbone are in **bold**).

Backbone	Random prompts	Knowledge prompts	Param.	Left hippocampus		Right hippocampus		Average	
				DSC (%)	ASD (%)	DSC (%)	ASD (%)	DSC (%)	ASD (%)
U-Net	✗	✗	92.31M	93.96±0.10*	23.51±0.47*	92.47±0.11*	29.69±0.40*	88.77±0.94*	40.41±1.52*
	✓	✗	14.79M	94.04±0.03*	21.60±0.58*	93.42±0.17*	20.40±0.63*	89.27±0.78*	38.36±0.92*
	✗	✓	14.79M	**94.93±0.06**	**21.24±0.33**	**94.75±0.18**	**22.45±0.35**	**90.42±0.38**	**37.94±0.66**
UNETR	✗	✗	92.78M	94.94±0.05*	21.24±0.23*	94.69±0.02*	22.85±0.25*	91.75±0.33*	30.22±0.37*
	✓	✗	4.46M	93.27±0.08*	25.48±0.31*	93.58±0.06*	23.15±0.36*	91.43±0.47*	24.75±0.63*
	✗	✓	4.46M	**95.10±0.05**	**19.47±0.29**	**95.28±0.03**	**18.63±0.21**	**92.07±0.26**	**20.49±0.25**
Swin UNETR	✗	✗	62.53M	96.17±0.04*	15.10±0.33*	96.62±0.02*	12.71±0.22*	93.04±0.28*	21.28±1.86*
	✓	✗	54.47M	95.43±0.02*	18.14±0.28*	95.86±0.03*	16.75±0.41*	92.28±0.42*	19.79±0.24*
	✗	✓	54.47M	**96.65±0.01**	**11.63±0.21**	**97.57±0.01**	**9.28±0.06**	**94.19±0.21**	**14.50±0.20**

in UNETR, where the parameter count is reduced from 92.78M to 4.46M without sacrificing performance. This suggests that KGPL aids models to prioritize essential information for brain image segmentation, thereby improving efficiency and adaptability.

Comparison with random-prompts-only methods. Table 1 of brain tissue segmentation shows that our proposed KGPL using UNETR as the backbone can offer 1.87% and 3.08% improvements of average DSC and ASD, respectively. In Table 2, KGPL using Swin UNETR shows improvements of 1.71% and 7.47% in average DSC and ASD for right hippocampus segmentation, respectively. Importantly, our proposed method is computationally faster ($3\times$), *i.e.*, extending the time from 3 days to 1 day. This emphasizes that knowledge-wise prompts can direct models to better comprehend and process brain image data with different characteristics and variations, thus promoting efficient convergence.

3.4   Qualitative Comparison Analysis

To more intuitively illustrate the superiority of our method, we display qualitative comparison results of brain tissue and structure segmentation.

In Figs. 2 and 3, our KGPL method yields more accurate segmentation across different backbones, with predictions closer to ground truths compared to full or random prompt-only fine-tuning. For example, Fig. 2 demonstrates our brain tissue segmentation closely matching the ground truth, particularly in complex brain regions like sulci and gyri. In Fig. 3, full fine-tuning backbones exhibit significant errors in regions like the hippocampus and cerebral cortex, while our predictions display greater precision, suggesting potential for aiding in brain disease diagnosis, such as left/right hippocampal sclerosis detection.

4   Conclusion

This work presents a segmentation framework employing knowledge-guided prompt learning for brain MRI analysis throughout the lifespan. It stands as a pioneer-
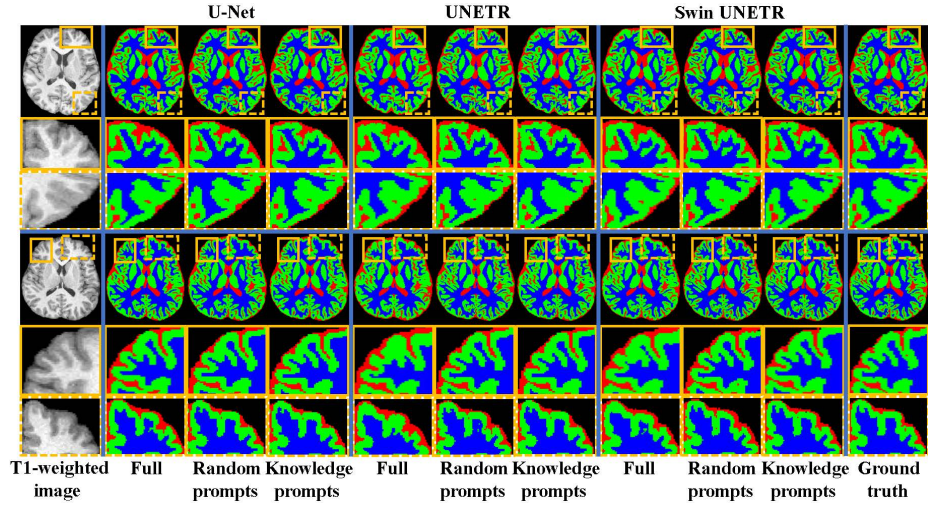
Fig. 2. Brain tissue segmentation results from different backbones are showcased in the transverse view. The first and second rows correspond to subjects aged 20 and 21, respectively. Each panel displays three segmentation results obtained by refining backbones using full, random prompts, and knowledge-wise prompts. Regions enhanced by our method are highlighted and enlarged within yellow boxes.
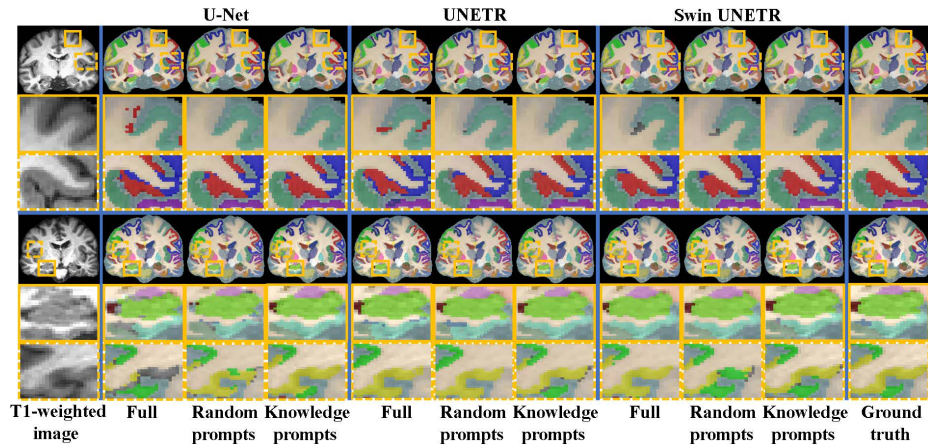

Fig. 3. Parcellation results of brain structure from three different backbones in the coronal view. The first and second rows correspond to subjects aged 17 and 24, respectively. Each panel displays three segmentation results achieved by refining backbones using full, random prompts, and knowledge-wise prompts. Regions improved by our proposed method are enclosed and magnified within yellow boxes.

ing effort in utilizing biomedical attribute knowledge as guidance to achieve precise segmentation for brain data. Experimental results indicate that knowledge benefits models in better learning structural variations of brain data across different age stages. The introduction of knowledge-wise learnable embeddings facilitates the extraction of fine-grained anatomical features and enhances training efficiency, even with small-scale datasets. Hence, the proposed method holds promise for clinical applications such as brain disease diagnosis. Moving forward, we plan to extend our methodology to infant phase datasets, enabling segmentation across the entire lifespan continuum.

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

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. Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D.V., Garavan, H., et al.: The adolescent brain cognitive development (abcd) study: imaging acquisition across 21 sites. Developmental cognitive neuroscience **32**, 43–54 (2018)
3. Cole, J.H., Marioni, R.E., Harris, S.E., Deary, I.J.: Brain age and other bodily 'ages': implications for neuropsychiatry. Molecular psychiatry **24**(2), 266–281 (2019)
4. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage **31**(3), 968–980 (2006)
5. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry **19**(6), 659–667 (2014)
6. Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., Ayed, I.B.: Deep cnn ensembles and suggestive annotations for infant brain mri segmentation. Computerized Medical Imaging and Graphics **79**, 101660 (2020)
7. 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)

8. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
9. González-Villà, S., Oliver, A., Valverde, S., Wang, L., Zwiggelaar, R., Lladó, X.: A review on brain structures segmentation in magnetic resonance imaging. Artificial intelligence in medicine **73**, 45–69 (2016)
10. 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)
11. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
12. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
13. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision. pp. 709–727. Springer (2022)
14. Kang, L., Gong, H., Wan, X., Li, H.: Visual-attribute prompt learning for progressive mild cognitive impairment prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 547–557. Springer (2023)
15. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
17. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on knowledge and data engineering **22**(10), 1345–1359 (2009)
18. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PMLR (2021)
19. Roy, A.G., Conjeti, S., Navab, N., Wachinger, C., Initiative, A.D.N., et al.: Quicknat: A fully convolutional network for quick and accurate segmentation of neuroanatomy. NeuroImage **186**, 713–727 (2019)
20. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebral cortex **28**(9), 3095–3114 (2018)
21. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al.: Advances in functional and structural mr image analysis and implementation as fsl. Neuroimage **23**, S208–S219 (2004)
22. Stiles, J., Jernigan, T.L.: The basics of brain development. Neuropsychology review **20**(4), 327–348 (2010)
23. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J.: Convolutional neural networks for medical image analysis: Full

training or fine tuning? IEEE transactions on medical imaging **35**(5), 1299–1312 (2016)

24. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)

25. Zhang, S., Ren, B., Yu, Z., Yang, H., Han, X., Chen, X., Zhou, Y., Shen, D., Zhang, X.Y.: Tw-net: Transformer weighted network for neonatal brain mri segmentation. IEEE Journal of Biomedical and Health Informatics **27**(2), 1072–1083 (2022)

26. Zhao, Z., Liu, Y., Wu, H., Li, Y., Wang, S., Teng, L., Liu, D., Li, X., Cui, Z., Wang, Q., et al.: Clip in medical imaging: A comprehensive survey. arXiv preprint arXiv:2312.07353 (2023)

27. Zhao, Z., Wang, S., Gu, J., Zhu, Y., Mei, L., Zhuang, Z., Cui, Z., Wang, Q., Shen, D.: Chatcad+: Towards a universal and reliable interactive cad using llms. IEEE Transactions on Medical Imaging (2024)