A UNIFIED BRAIN EXTRACTION FRAMEWORK POWERED BY STRUCTURE-INTENSITY DISENTANGLEMENT-BASED DATA AUGMENTATION

Xiaoye Li[1] *Shijie Huang*[1] *Yulin Wang*[1] *Zifeng Lian*[1] *Jiameng Liu*[1]
Kaicong Sun[1] *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
[3] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
**Dinggang.Shen@gmail.com*

ABSTRACT

Brain extraction is a fundamental step in neuroimaging analysis, yet existing methods are often constrained to specific modalities or age groups, limiting their applicability in large-scale, lifespan studies. In this work, we present BrainExt, a unified foundation model for brain extraction across multiple imaging modalities and the whole lifespan. To address inter-modality and across-age variations, we first collect a large-scale multi-modal, lifespan brain dataset of 25,487 scans spanning from fetal to elderly subjects and introduce a Structure-Intensity Disentanglement Synthesis (SIDSyn) module to generate real-world distribution-aligned data for robust pre-training. The pre-trained model is subsequently fine-tuned on real clinical scans to better adapt to real-world data distributions. BrainExt demonstrates superior generalization and stability compared to existing methods, achieving high Dice and low surface distance across all modalities. Empowered by disentanglement-based augmentation and a two-stage training strategy, BrainExt provides a scalable, modality-agnostic foundation for unified brain extraction, establishing a strong basis for advancing neuroimaging research and clinical applications.

Index Terms— Brain Extraction, Data Augmentation

1. INTRODUCTION

Brain extraction is a fundamental prerequisite for numerous neuroimaging analyses. By removing irrelevant structures like the skull and dura mater, brain extraction standardizes the input space and improves the accuracy and reliability of subsequent quantitative analyses. However, manual labeling of datasets is exceptionally labor-intensive and prone to inter- and intra-rater inconsistencies, particularly for large datasets spanning different ages and imaging modalities. This underscores the critical need for a unified framework capable of robustly handling such variations across populations, protocols, and the lifespan.

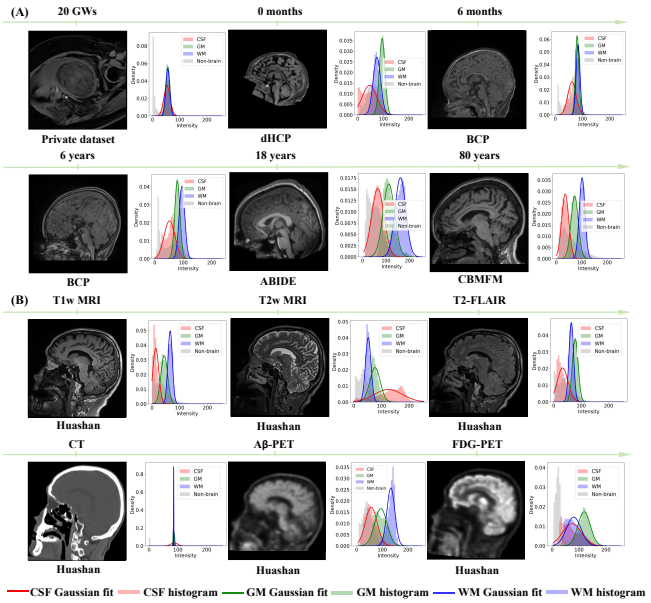

Fig. 1. Illustration of the intensity distributions of different tissue types and non-brain regions across lifespan and diverse imaging modalities. (A) Age-related intensity distribution variations; (B) Modality-related intensity distribution variations (a 69-year-old subject from the OASIS dataset).

Existing brain extraction methods can be broadly classified into optimization-based [1, 2] and deep learning-based [3, 4, 5] approaches. Among pioneering optimization-based methods, the Brain Extraction Tool (BET) [1] initializes a spherical mesh at the barycenter of the brain and projects the vertices outward to align with the brain boundary. FreeSurfer [2] utilizes a hybrid method that leverages a deformable surface combined with a watershed algorithm and statistical atlases to improve the identification of the contour of the

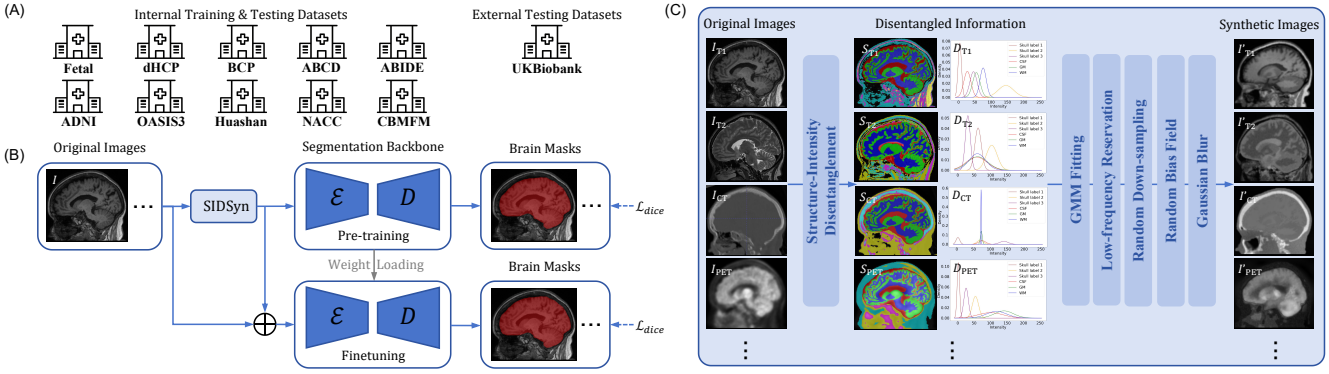

Fig. 2. Overview of the BrainExt framework. (A) Our brain database includes 25,487 scans from 11 medical centers. (B) Two-stage training scheme: pre-training on synthetic data generated by the SIDSyn module, followed by fine-tuning on real scans. (C) Details of the SIDSyn module.

brain. However, these optimization-based methods are inherently constrained by the preset optimization parameters, leading to strong modality-specific dependency and limited adaptability across different modality images.

More recently, deep learning-based methods have been developed for brain extraction. For example, Salehi *et al.* [4] proposed Auto-Net, which learns complementary information from sagittal, coronal, and axial views by three parallel 2D networks, followed by a 2D U-Net to extract the brain. However, Auto-Net is only effective for T1-weighted (T1w) MR images of adult and neonatal groups. Similarly, Kleesiek *et al.* [3] conducted an initial attempt for multi-modal segmentation by training a 3D network on combinations of T1w, T2-weighted (T2w), and T2-FLAIR MR images. Recently, SynthStrip [5] addressed variations in resolution and imaging conditions by training a 3D U-Net on randomly generated synthetic data. However, the substantial domain gap between synthetic and real-world imaging distributions increases the risk of model overfitting [6]. In summary, existing methods are constrained by modality- or age-specific designs, making them unreliable for consistent brain extraction across diverse modalities and developmental stages (Fig. 1).

To overcome these limitations, we present BrainExt, a unified brain extraction framework, for multi-modal neuroimages spanning the whole lifespan (20 gestational weeks to 100 years). BrainExt leverages a structure-intensity disentanglement data augmentation strategy to generate large-scale, realistic, and synthetic data for pre-training, followed by fine-tuning on a modality-balanced hybrid dataset, enabling robust and generalizable performance for neuroscience research and clinical applications.

2. METHOD

We propose a unified model for multi-modality and lifespan brain extraction (BrainExt) as shown in Fig. 2. It is powered by a Structure-Intensity Disentanglement Synthesis (SIDSyn) module for generating augmented training data covering real-world distributions and artifacts, and is trained using a two-stage learning strategy: initial pre-training on synthetic data, followed by fine-tuning on real scans. Additionally, a tailored strategy is designed for fetal brain extraction to address its specific challenges.

2.1. Structure-Intensity Disentanglement Synthesis

The synthetic data is generated using our **SIDSyn** module, designed specifically to generate realistic and diverse multi-modal brain images across the lifespan. The core assumption of **SIDSyn** is that, while brain tissues maintain stable physicochemical properties, variations across scanners, modalities, and ages mainly appear as systematic shifts in tissue-specific intensity distributions (Fig. 1). To implement this, we model each tissue's intensity as a Gaussian distribution, enabling the brain image to be represented as a Gaussian Mixture Model (GMM). Accordingly, a brain image I is disentangled into a structural representation S, describing K anatomical categories including three cerebral tissues (CSF, GM, WM), and three background regions, with an intensity component D reflecting the intensity distribution of each category. Each category k is modeled by a Gaussian distribution $\mathcal{N}(\mu_k, \sigma_k^2)$; therefore, the overall image distribution can be expressed as a GMM parameterized by $\theta = \{\mu_k, \sigma_k\}_{k=1}^{K}$.

The mapping $f(S, D)$ formalizes the process by which these two complementary components are combined to generate the final image. Specifically, for each voxel, the structural label specified by S determines the tissue category, and the corresponding intensity value is drawn from the associated Gaussian distribution in D. In practice, given a source image I_s and a target image I_t, new images are synthesized by combining the anatomical structure of the source (S_s) with

the intensity characteristics of the target (D_t), i.e.,

$$I_{\text{syn}} = f(S_s, D_t).$$

The intensity of each voxel labeled as category k in S_s is sampled from the corresponding distribution $\mathcal{N}(\mu_k^t, (\sigma_k^t)^2)$ in D_t, resulting in a synthesized image that preserves anatomical integrity while transferring realistic contrast characteristics between modalities or age groups.

To further enhance realism, we transfer low-frequency components via Fast Fourier Transform (FFT) [7, 8] and apply random bias field, down-sampling, and gamma transformations to emulate scanner- and site-specific effects.

2.2. Fetal Brain Extraction

Fetal brain extraction presents unique challenges due to (1) large variability in fetal positions and (2) small brain-to-image ratio. Directly combining fetal and adult data can often lead to model confusion and poor segmentation performance. To address this, we design a dedicated strategy for consistent performance across the lifespan.

We employ two preprocessing schemes: (1) center-aligned cropping (Fig. 3, (2–4, 9–11)) and (2) stochastic cropping with random offsets (Fig. 3, (5–7, 12–14)), each applied with 50% probability during training. Both schemes ensure full brain coverage with varying scale and position, enabling the network to learn scale- and position-invariant fetal features.

At inference, a two-stage pipeline is used. The model first generates a coarse mask from the full image, followed by center cropping based on this mask and a refined segmentation using the same model. This fetal-specific preprocessing enables the model to capture brain-relevant features without altering its architecture, ensuring accurate and consistent segmentation across developmental stages.

3. EXPERIMENTS AND RESULTS

3.1. Datasets and Implementation Details

3.1.1. Datasets

We construct a large-scale, diverse cohort to establish the foundation for unified model development and evaluation. This dataset comprises 25,487 scans across 6 modalities from 11 global medical centers [9, 10, 11, 12, 13, 14, 15, 16], spanning ages from 21 gestational weeks to 100 years. The ground truth (GT) for fetal images is generated using MONAIfbs [17] and manually refined by two clinical experts, while GT labels for other age groups are obtained from public datasets and verified and manually corrected by two additional professionals. Each internal dataset is randomly divided into training, validation, and testing datasets in an 8:1:1 ratio. Segmentation performance is evaluated using Dice similarity coefficient (DSC) and average surface distance (ASD) in the format of mean ± standard deviation.

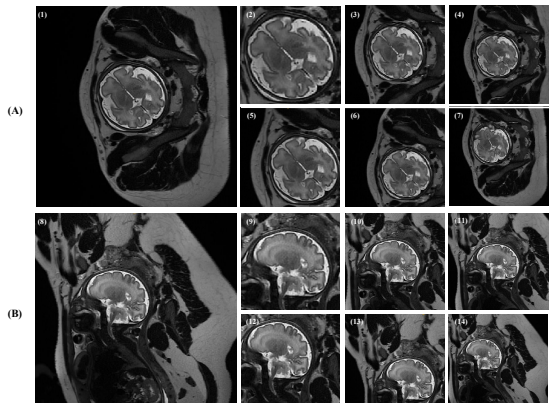

Fig. 3. Visualization of the multi-scale cropping strategy for fetal MRI on two examples. (A) and (B) are from the same subject but shown in different views. (1) and (8): original images; (2)–(4) and (9)–(11): center-cropped sub-images from small to large views; (5)–(7) and (12)–(14): randomly-cropped sub-images from small to large views.

3.1.2. Implementation Details

The proposed method is implemented using PyTorch in Python and is trained on an NVIDIA A100 GPU with 40GB memory. We use Adam as the optimizer with a learning rate of 1.0×10^{-4}, adopt SwinUNETR [18] as the backbone, and optimize the network by Dice loss [19].

3.2. Comparison with State-of-the-art Methods

We compare our method with three state-of-the-art brain extraction methods: SynthStrip [3], BET [1], and HD-BET [20].

3.2.1. Quantitative Comparison

From Table 1, our model achieves the best performance across all modalities, demonstrating superior robustness. Although SynthStrip is also trained on synthesized data, its performance remains moderate across different modalities, likely due to its random synthesis being misaligned with real-world intensity distributions. In addition, other methods are shown to perform well in DSC but have relatively poor performance in ASD. This indicates that, while they achieve satisfactory volumetric overlapping, they struggle to maintain precise boundary alignment and accurate surface positioning. Interestingly, we can observe that BET achieves relatively high DSC on T1w and T2w images, but exhibits a substantial performance drop on CT and PET. In contrast, our model achieves the best performance in terms of all the evaluation metrics across all the imaging modalities, not only in average value but also in

Table 1. Quantitative evaluation of the proposed method and competing methods.

Method	T1w		T2w		CT		PET	
	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD
BET	93.08±5.56	3.20±2.50	95.37±2.31	1.95±1.06	60.58±3.17	24.99±4.94	66.97±5.28	22.52±3.86
HD-BET	96.83±0.44	1.28±0.16	97.28±1.97	1.26±1.58	93.99±2.09	3.12±1.76	94.88±0.84	2.49±0.41
SynthStrip	95.20±2.6	2.00±0.64	96.12±0.90	1.66±0.34	96.12±0.79	1.82±0.44	95.71±0.71	2.20±0.34
Ours	**99.20±0.62**	**0.25±0.16**	**99.23±0.25**	**0.33±0.18**	**98.43±0.39**	**0.36±0.09**	**98.18±0.57**	**0.42±0.11**

Table 2. Validation of the proposed method on whole-lifespan MRI data.

Modality	Fetus		Neonate		Infant		Adolescent		Adult	
	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD	DSC	ASD
T1w MRI	/	/	98.72±0.67	0.20±0.08	98.97±0.22	0.28±0.06	99.24±0.20	0.25±0.04	99.24±0.63	0.24±0.09
T2w MRI	95.98±3.06	0.66±0.52	98.57±0.33	0.22±0.07	99.10±0.25	0.29±0.07	99.19±0.13	0.34±0.07	99.27±0.22	0.34±0.03

Table 3. Quantitative evaluation of the proposed and competing methods on zero-shot external sets.

Method	T2-FLAIR		SWI	
	DSC	ASD	DSC	ASD
BET	96.28±1.04	1.52±0.22	97.52±0.37	1.03±0.14
HD-BET	93.96±0.91	2.49±0.41	95.40±0.98	1.84±0.44
SynthStrip	96.45±0.60	1.49±0.29	96.77±0.66	1.33±0.26
Ours	**97.21±0.45**	**0.65±0.08**	**97.75±0.54**	**0.52±0.07**

standard deviation, highlighting its accuracy and robustness. The performance gain mainly comes from both the proposed real distribution-aligned data synthesis scheme and the two-stage training strategy.

3.2.2. Performance on Lifespan Data

We further evaluate performance across five developmental stages (fetus, neonate, infant, teenage, adult; Table 2). Our model maintains high accuracy (about 99% DSC) across all groups, with a slight decrease in the fetal stage (95.98% ± 3.06%). These results confirm the robustness of our model throughout the lifespan.

3.2.3. Performance on Zero-shot Modalities

On both external modalities (T2-FLAIR and SWI), our model maintains high performance without noticeable degradation (Table 3). Compared with SynthStrip, our method attains higher DSC (97.21% vs. 96.45% on T2-FLAIR; 97.75% vs. 96.77% on SWI) and notably lower ASD, indicating more accurate boundary delineation. These results demonstrate our model's strong generalization and robustness to unseen modalities.

3.2.4. Qualitative Comparison

Fig. 4 demonstrates the superior performance of our method in producing sharper and more accurate brain boundaries across diverse modalities and age groups. Our approach

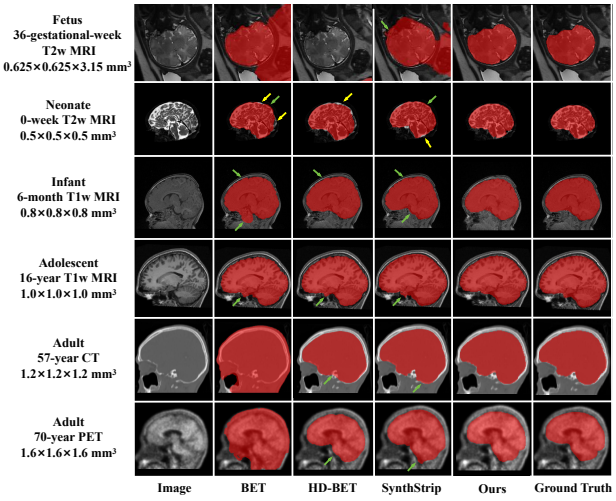

Fig. 4. Visual comparison of the segmentation results. The yellow and green arrows represent the under-segmentation and over-segmentation errors, respectively.

yields stable and anatomically coherent segmentations, ensuring reliability for applications such as longitudinal analyses. In contrast, competing methods exhibit under-segmentation in neonates (yellow arrows) and over-segmentation in older subjects (green arrows), leading to inconsistent results across the lifespan.

4. CONCLUSION

We have presented BrainExt, a unified model that achieves robust and consistent multi-modal brain extraction across the entire lifespan. Empowered by disentanglement-based data augmentation and two-stage training, it mitigates modality and age variations which often affect existing methods. This scalable and foundational tool establishes a strong basis for advancing neuroscience research and clinical neuroimaging applications.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. It involved both publicly available and institutionally approved private datasets. Ethical approval was not required as confirmed by the license attached with the open access data. For the private dataset, our research was approved by the Ethics Committees.

6. ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China (grant numbers 62131015, 82441023, U23A20295, 82394432), the China Ministry of Science and Technology (S20240085, STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001), Key-Area Research and Development Program of Guangdong Province (2023B0303040001), and HPC Platform of ShanghaiTech University.

7. REFERENCES

[1] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.

[2] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.

[3] A. Hoopes et al., "Synthstrip: Skull-stripping for any brain image," *NeuroImage*, vol. 260, pp. 119474, 2022.

[4] J. Kleesiek et al., "Deep mri brain extraction: A 3d convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460–469, 2016.

[5] S. S. M. Salehi et al., "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2319–2330, 2017.

[6] Y. S. Abu-Mostafa, "The vapnik–chervonenkis dimension: Information versus complexity in learning," *Neural Computation*, vol. 1, no. 3, pp. 312–317, 1989.

[7] Y. Yang et al., "Fda: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4085–4095.

[8] S. Huang et al., "Tissue segmentation of thick-slice fetal brain mr scans with guidance from high-quality isotropic volumes," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 4, pp. 1067–1076, 2023.

[9] R. Petersen et al., "Alzheimer's disease neuroimaging initiative (adni): clinical characterization," *Neurology*, vol. 74, pp. 201–209, 2010.

[10] B. Howell et al., "The unc/umn baby connectome project (bcp): An overview of the study design and protocol development," *NeuroImage*, vol. 185, pp. 891–905, 2019.

[11] P. J. LaMontagne et al., "Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease," *medRxiv*, 2019.

[12] C. Sudlow et al., "Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *PLoS Medicine*, vol. 12, pp. e1001779, 2015.

[13] B. J. Casey et al., "The adolescent brain cognitive development (abcd) study: Imaging acquisition across 21 sites," *Developmental Cognitive Neuroscience*, vol. 32, pp. 43–54, 2018.

[14] E. J. Hughes et al., "A dedicated neonatal brain imaging system," *Magnetic Resonance in Medicine*, vol. 78, pp. 794–804, 2017.

[15] L. Cordero-Grande et al., "Sensitivity encoding for aligned multishot magnetic resonance reconstruction," *IEEE Transactions on Computational Imaging*, vol. 2, pp. 266–280, 2016.

[16] A. Di Martino et al., "Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii," *Scientific Data*, vol. 4, pp. 1–15, 2017.

[17] M. Ranzini et al., "Monaifbs: Monai-based fetal brain mri deep learning segmentation," *arXiv preprint arXiv:2103.13314*, 2021.

[18] A. Hatamizadeh et al., "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *Proceedings of the International MICCAI BrainLesion Workshop*. 2021, pp. 272–284, Springer.

[19] A. A. Taha et al., "Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, pp. 1–28, 2015.

[20] F. Isensee et al., "Automated brain extraction of multisequence mri using artificial neural networks," *Human Brain Mapping*, vol. 40, no. 17, pp. 4952–4964, 2019.