AmygdalaGo-BOLT3D: A fast and valid transformer tracing boundaries of amygdala from brain images

Bo Dong [1*], Quan Zhou [2,3*], Peng Gao [2,3*], Jintao Wei [1], Jiale Xiao [3], Wei Wang [3], Peipeng Liang [4], Danhua Lin [3], for the Chinese Color Nest Consortium (CCNC) [**], Hongjian He [5,6*], Xi-Nian Zuo [2,7,8,9***]**

[1]College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, 310027, China.
[2]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, China.
[3]Faculty of Psychology, Beijing Normal University, Beijing, 100875, China.
[4]School of Psychology, Capital Normal University, Beijing, 100048, China.
[5]School of Physics, Zhejiang University, Hangzhou, 310058, China.
[6]State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, 310027, China.
[7]Developmental Population Neuroscience Research Center, McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China.
[8]National Basic Science Data Center, Beijing, 100190, China.
[9]Developmental Population Neuroscience Research Base for Education, Nanning Normal University, Nanning, 530001, China.
[*]Equal contribution as the first author.
[**]A list of CCNC authors and their affiliations appears at the end of the paper. More information about CCNC are available at https://ccnp.scidb.cn.
[***]Corresponding author: Xi-Nian Zuo & Hongjian He; Emails: xinian.zuo@bnu.edu.cn, hhezju@zju.edu.cn

ABSTRACT

Tracing boundaries of amygdala from brain images is one of the most fundamental steps in human neuroscience research. However, due to its small volume and complex anatomy, especially in developing brains, the precision and consistency of the tracing outcomes are highly affected by individual differences and methodological inconsistencies between different tools. To address these challenges, we propose an algorithm for learning boundary contrast of 427 manually traced amygdalae in children and adolescents to generate a transformer, AmygdalaGo-BOLT3D, for automatically tracing human amygdala. This method focuses on the boundary to effectively address issues with false positive recognition and inaccurate edges due to small but complex amygdala. AmygdalaGo-BOLT3D first develops the basic architecture of an adaptive cooperation network with multiple granularities. A self-attention-based consistency module is then constructed to improve generalizability by adapting the three-module sample-mask model for the amygdala scene: a lightweight volumetric feature encoder, a 3D cue encoder, and a volume mask decoder. Finally, AmygdalaGo-BOLT3D implements a boundary contrastive learning framework that utilizes the interaction mechanisms between a prior cue and the embedded magnetic resonance images to achieve their effective integration. Our results confirmed that predictions of the overall structure and boundaries of the human amygdala exhibited highly improved precision as the golden-standard - manual tracing, while further achieve high stability across multiple age groups and imaging centers and more efficient for large-scale studies than manual tracing. Open source code for AmygdalaGo-BOLT3D is available at Science Data Bank (SciDB_LINK).

INTRODUCTION

The amygdala, which is vital for mood regulation, cognitive learning, and memory consolidation, has been closely associated with behavioral pathologies, particularly during critical stages of socio-emotional maturation such as childhood and adolescence [1–3]. Magnetic Resonance Imaging (MRI) of brain tissue structures has enabled detailed examination of the amygdala, facilitating the construction of developmental trajectories and the investigation of neurodevelopmental processes underlying various neuropsychiatric conditions [4,5]. Volumetric analyses, based on a precise segmentation of the amygdala by tracing its boundary, are an essential component of these studies. Conventional amygdala segmentation has primarily relied on the use of histological atlases and manual tracing techniques involving in vivo neuroimaging data. However, while serving as the gold standard, these methods predominantly require manual labeling, making them labor intensive and prone to errors. Furthermore, the necessity of processing large-scale datasets makes this task particularly cumbersome, thus requiring the development of automated segmentation techniques of tracing amygdala. One common automated method involves probabilistic estimates derived from Markov random fields [6], in which voxel labels are assigned in anatomical MRI images. This approach has been widely implemented in popular software such as FreeSurfer [6], achieving notable volume-segmentation or boundary-tracing performance [7].

Previous studies also implemented Bayesian methods to delineate brain boundaries [8], based on the MRI signal intensity and the structural configurations defined by established atlases (e.g., FSL-FIRST and similar segmentation algorithms [9]). Multi-atlas label fusion segmentation techniques [10] integrate information from multiple atlases to reduce the mislabeling errors that result from inac-

curate registration, mitigating the inherent instability of MRI data. This methodology has proven to be particularly effective in scenarios involving the high variability of Markov random fields, covering individual disparities and pathological lesions. The technique has also been incorporated into volBrain [11] and more recent algorithms [10,12]. Despite the variety of available automated brain image segmentation tools, including FreeSurfer and volBrain, it is noteworthy that these toolkits were not developed exclusively for amygdala segmentation. Recent studies have identified significant discrepancies in the quantitative estimates of amygdala volumes when using existing software algorithms, which can affect the validity of amygdala-centric research findings [13]. In other words, inaccurate segmentation or tracing performance can introduce systematic biases, compromising the precision and reliability of the amygdala's volumetric measurements.

In summary, the accurate segmentation of the human amygdala poses significant challenges. First, the amygdala constitutes merely 0.05% of the entire brain volume, making it susceptible to false positive segmentation results. Second, the differences between the amygdala and adjacent tissue structures are minimal (∼12.6%) compared to surrounding tissues. The complexity of neighboring structures further complicates the segmentation tasks due to issues with low contrast [6,11,14]. In this work, we propose an innovative algorithm, AmygdalaGo-BOLD, that takes advantage of deep learning from manually traced large-scale amygdala images [13] in the development of a transformer model [15], which emphasizes accuracy on boundary delineation to address the unique challenges of automated boundary-tracing of amygdala from human brain images.

METHOD

The implementation of AmygdalaGo-BOLD was divided into three critical phases: 1) robust global localization, 2) precise delineation, and 3) generalizable segmentation. When examining the structure of the amygdala, the observer must first capture the entire cerebral region and subsequently estimate the approximate position of the amygdala within the image. This estimate relies on prior knowledge of limbic system anatomy and considers adjacent structures such as the hippocampus. Subsequent to this initial phase, the observer leverages a detailed understanding of internal amygdala morphology, gray matter intensity, and other pertinent information used to execute a full regional delineation. The segmentation process is then enhanced using a transformer-based algorithm employed to discern boundary contrast. It aims to improve both the robustness of global amygdala localization and the precision of boundary delineation, thereby improving inter-individual consistency. This process includes the following key elements:

1. Comprehensive detection of amygdala positioning in T1-weighted (T1w) MR images was achieved by employing a transformer-based network. This model (BOLT3D) constructs a multi-granularity adaptive collaboration (MGAC) architecture, which integrates a global coarse-grained perceptual module. It determines the global semantics of the amygdala structure, thereby enhancing positional sensitivity. Since intricate details are often neglected in global representations, a local fine-grained enhancement module was developed to increase the perception and retention of detailed information. Coarse-grained global and fine-grained local data were further harmonized by introducing a multi-granular adaptive fusion module. This allowed BOLT3D to concurrently consider differences between the entire amygdala and adjacent structures, as well as similarities within the amygdala itself. It facilitates the recognition of global positioning characteristics and local disparities, thereby improving the accuracy and robustness of amygdala structure location estimates. The overarching design objective is then to perform a more precise and exhaustive detection of amygdala tracing contours.

2. A novel strategy is introduced for learning boundary contrast, to address dependency challenges for manual segmentation boundaries and enhance the congruence between predicted and actual delineations. It improves the delineation of limbic structures from adjacent tissues by amplifying characteristic differences between the amygdala and surrounding tissues, while minimizing intra-amygdala variations. Consequently, this step refines the alignment between automated boundary segmentation and manual segmentation outcomes, enabling a more precise demarcation of amygdala dimensions.

3. A new Segment Anything Model (SAM) is developed to refine coarse predictions and ensure the generalizability of BOLT3D. Specifically, coarse outputs from the MAGC module are sampled as point cues for the SAM model, utilizing coarse segmentation as a pre-segmentation step for further refinement. The original SAM model is concurrently adapted to the amygdala

context, which contains three essential components: a lightweight volumetric feature encoder, a 3D cue encoder, and a volumetric mask decoder, thereby enhancing generalized segmentation performance.

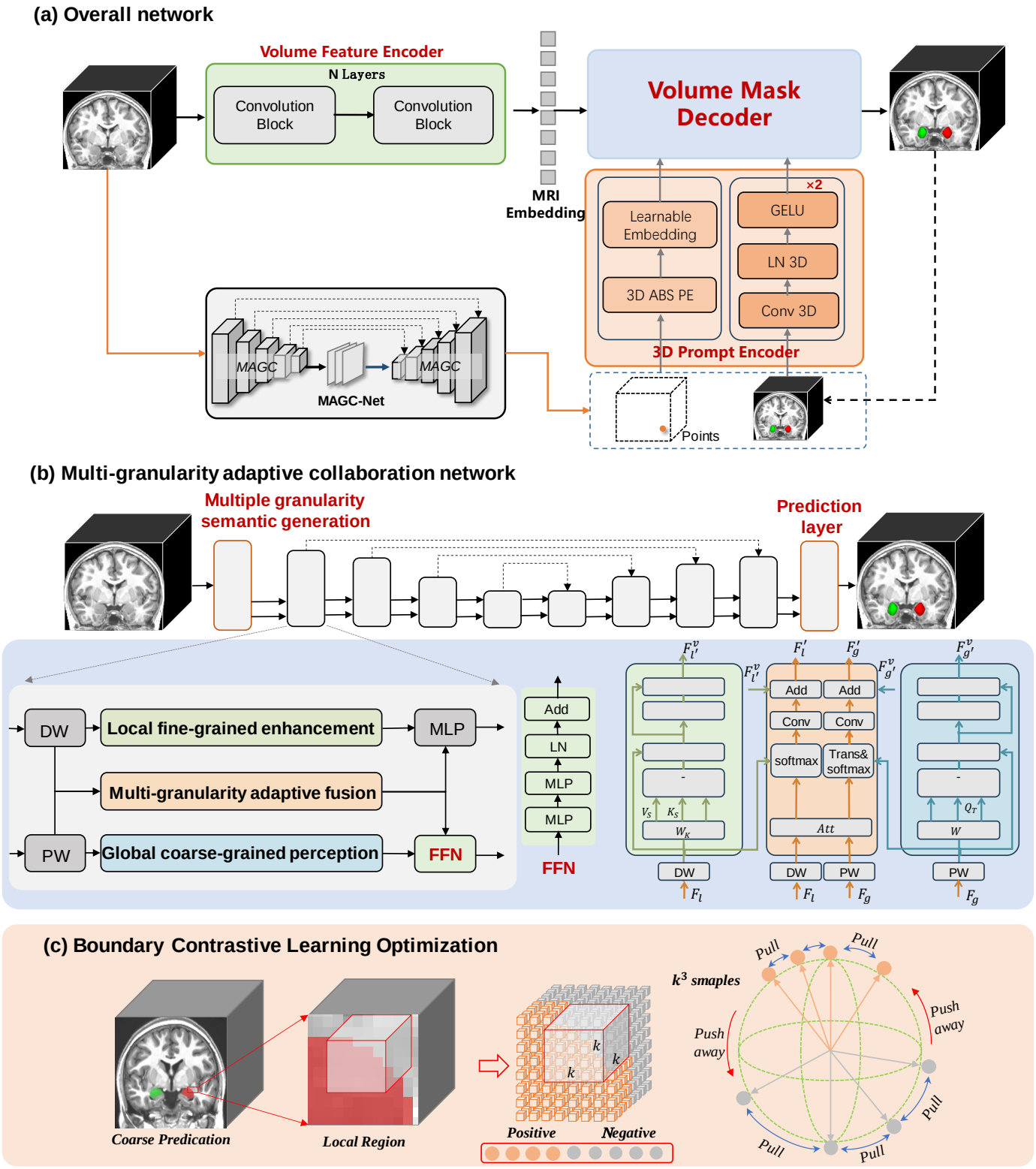

Figure 1: A schematic diagram of the AmygdalaGo-BOLT3D architecture. (a) The overall model structure consists of two principal components: a MAGC-Net (responsible for initial segmentation) and an SAM. The latter is tasked with the refinement of segmentation results, including three essential components: a lightweight Volumetric Feature Encoder (VFE), a 3D Prompt Encoder (PE), and a Volumetric Mask Decoder (VMD). The MAGC network produces preliminary segmentation results and strategically samples voxels, which are then fed into the SAM prompt encoder to guide the refinement process. (b) The MAGC block, a key network module designed to address challenges posed by the size of the amygdala, ensures the extraction of accurate semantic features. (c) The included edge contrast learning optimization strategy, incorporated to further augment edge delineation and enhance overall segmentation precision.

Multi-Granularity Adaptive Collaboration

Specifically, global and local semantic features were initially derived from the original MR image data using a multi-granularity semantic generation module, as depicted in Figure 1(b). Specifically, for T1w images I, the module exploited local induction capabilities by employing two 3D convolutions to extract primary local features F_l. This approach facilitated the computation of relationships among feature points, thereby enabling the extraction of global semantic information from regions exhibiting high responses to initial position features. Subsequently, preliminary global features F_g were constructed as follows. A feature map $F^{'}$ of dimensions $d \times H \times W$ was derived from the input image I using a single convolutional layer and subsequently projected onto a weighting map via two convolutional layers. The channel count in the weight map was given by hw ($h \leq H, w \leq W$), where h and w represent the dimensions of a predefined global semantic map. The weight map was then flattened for the Softmax computation and the resulting weights were used to aggregate semantic information within the $F^{'}$ mapping through a matrix multiplication, producing the final F_g result. In the following illustrations, dimensions are presented in an abbreviated form for brevity.

The MGAC module includes three components: a global module for coarse-grained perception, a local module for fine-grained enhancement, and an adaptive module for multi-granularity fusion, as depicted in Figure 1(c). These components independently optimize global and local features, which are interactively integrated to enhance model performance. Figure 1(b) illustrates the positioning of the adaptive MGAC module within the overall framework, while Figure 1(c) details the corresponding internal structure. Each MGAC module receives two inputs: global coarse-grained (F_g) and local fine-grained (F_l) semantic characteristics. Significant discrepancies in semantic distributions between these two sets of characteristics were mitigated by using BatchNorm3D to normalize the inputs. A linear position projection from previous work was then mapped based on the positional relationships of the elements, ensuring the arrangement of tokens within the transformer remained unchanged while preserving their positional correlations [16]. However, this approach neglects local structural information, which is vital for image data. Although integrating positional encoding enables the transformer to learn positional relationships, such encoding from randomly initialized parameters require extensive training data, which poses a challenge for medical image analysis. Traditional positional encoding data were then substituted with convolutional layers, thereby incorporating the inductive bias inherent in convolutions. This substitution facilitated the preservation of local structural information intrinsic in the images while avoiding the limitations associated with positional encoding.

A three-dimensional (3D) depth-wise convolution strategy was employed to convolve relationships among local neighborhoods within fine-grained semantic features, thereby introducing visual bias. In contrast, 3D pointwise convolutions were used to calculate global coarse-grained semantic features. The results, denoted as F_g and F_l, were subsequently flattened into a one-dimensional (1D) sequence to facilitate Transformer processing. This transformation involved converting F_g into a global query vector F_g^q and a global content vector F_g^v. Similarly, F_l, which encapsulates detailed information, was transformed into a local index vector F_l^k and a local content vector F_l^v through analogous operations. In contrast, the internal architectures of the global coarse-grained perception module and the local fine-grained enhancement module remained consistent, as illustrated in Figure 1b. These modules are primarily comprised of two components: a multi-head self-attention mechanism and a feedforward neural network. The Z_l term initially replaced the input in the multi-head self-attention mechanism and was then expanded into a 1D sequence. A fully connected layer served as the transformation layer, converting the input to a query vector, an index vector, and a content vector. The multi-head attention matrix could subsequently be determined as:

$$Z_l' = MHSA\left(LN(Z_l)\right) + Z_l, \tag{1}$$

where the feedforward layer further refined the relationships among voxel features, while integration with residual mechanisms augmented the data flow. This approach preserved original features and mitigated vanishing gradients, as described by:

$$Z_{l'}^v = FFN\left(LN\left(Z_l'\right)\right) + Z_l'. \tag{2}$$

In this context, $MHSA(\cdot)$ denotes a multi-head self-attention mechanism, $LN(\cdot)$ signifies layer normalization, and $FFN(\cdot)$ represents a feedforward neural network. The global coarse-grained perception module and the local fine-grained enhancement module employed Eqs. 1 and 2 to transform the input global content vector F_g^v and the local content vector F_l^v, respectively. The resulting updated global content vector $F_{g'}^v$ and the local content vector $F_{l'}^v$ were obtained upon completion of the processing step. The correlation matrix Att was then derived using F_g^q as a query vector and F_l^k as an index vector, within the context of the adaptive multi-granularity fusion module, as illustrated in Figure 1. The resulting correlation was calculated as follows:

$$Att_{i,j} = \frac{1}{\sqrt{d_k}}\left(e^{F_{g\,i}^q {F_l^k}_j^\top} / \sum_{j=1}^{n} e^{F_{g\,i}^q}\right), \tag{3}$$

where $Att_{i,j}$ represents the i-th voxel characteristic within the global query vector F_g^q and the corresponding i-th voxel characteristic in the local index vector F_l^k. Correlation scores between characteristics of the j-th voxel were then calculated accordingly. A scaling factor of $1/\sqrt{d_k}$ was then implemented to mitigate overly concentrated correlations in specific local regions and facilitate the model optimization process, where d_k represents the dimensions of an individual attention head in terms of the number of channels within F_g^q. After this derivation, the similarity matrix Att must be remapped into the original value space, comprising both the global content vector space F_g^v and the local content vector space F_l^v. Specifically, the given similarity matrix Att undergoes direct propagation into a Softmax function, to produce a sparse feature distribution for the local fine-grained branch F_l^v. This distribution is then subjected to matrix multiplication with F_l^v, followed by convolution operations to further refine the feature distribution. This sequence can be represented by:

$$F_l^{'} = DW\left(softmax(Att)F_l^v\right) + F_{l'}^v. \tag{4}$$

In the case of a global coarse-grained branch F_g^v, the matrix Att requires transposition to determine the similarity score between individual feature points and the query feature. Subsequently, matrix multiplication with F_g^v is performed and the resultant product is processed using deep separable convolutions to enhance the smoothness of feature distributions. In addition, a residual technique is employed to preserve the integrity of information and mitigate issues with vanishing gradients. Ultimately, this module yields the following output:

$$F_g^{'} = DW\left(softmax(Att^\top)F_g^v\right) + F_{g'}^v. \tag{5}$$

Tracing Boundary Contrast Learning

The vast majority of existing 3D segmentation algorithms [17–19] neglect the explicit segmentation of structural boundaries. However, some algorithms strive to improve boundary delineation through the use of attention mechanisms [20,21]. Despite these efforts, there remains a lack of comprehensive and targeted research focused on evaluating the influence of boundary regions on segmentation outcomes. To address this issue, we propose meticulously aligning boundary information from manual annotations with model segmentation outputs, thereby introducing an innovative boundary contrast learning framework.

Previous contrastive learning methodologies have predominantly been applied at the image level. For example, positional techniques [22] extend 3D images to 2D sequences to determine spatial relationships between positive and negative samples as part of contrastive learning. The Mask Transfiner model [23] adopts contrastive learning within image patches, rather than at the image level, utilizing pseudo-labels to generate patches with distinct semantic categories and forming positive and negative sample pairs. However, these strategies are inadequate for high-resolution segmentation of the amygdala, which exhibits proximity with same-type pixels throughout a 3D image. It requires the separation of pixel types and the clustering of similar pixels based on individual labels.

Here, boundary pixels were used to construct positive and negative sample pairs to maximize distinctions between boundary pixels and adjacent non-relevant pixels, as illustrated in Figure 1(d). Specifically, the local neighborhood around the boundary was designated as a subset, establishing a contrast learning strategy for overlapping 3D patch areas distributed throughout 3D space. Kernels adjacent to the boundary were then used to select the dimensions of 3D patch spaces in each subset. Manually segmented amygdala data thus served as a basis for constructing both positive and negative samples. The InfoNCE information maximization algorithm [24] was then used to calculate the disparity between positive and negative samples in the latitude space on a hyper-

sphere. The kernel size was set to $5 \times 5 \times 5$ (based on the anatomical dimensions of the amygdala) and each subset formed positive and negative samples within this specified area. As illustrated in Figure 1, the model was specifically designed to focus on and exclusively compute boundary voxels. Initially, all boundary regions B_l were considered in manual annotations. For each voxel x_i in B_l, the sampling scope for positive and negative voxels was then confined to the local neighborhood N_i. Under this constraint, positive samples $x_j \in N_i \wedge l_i = l_j$ were identified for boundary voxels x_i, while any remaining neighborhood voxels were treated as negative samples $x_j \in N_i \wedge l_i \neq l_j$. In this paper, j denotes sample voxels within the neighborhood N_i corresponding to the voxel i, with l_i and l_j representing the voxel categories for x_i and x_j, respectively. Consequently, the boundary voxel contrast learning algorithm enhanced feature discernment for structurally organized boundary voxels, thereby establishing a target for contrast optimization within boundary voxels, which is crucial for enhancing the segmentation of boundary regions. Specifically, for the voxels x_i within boundary regions B_l, learned representations were incentivized to be more congruent with voxel features from the same category, while being less congruent with features of disparate categories. This process can be described as follows:

$$L_{CBL} = \frac{-1}{B_l} \sum_{x_i \in B_l} log\Big(\frac{\sum_{x_j \in N_i \wedge l_j = l_i} exp(-d(f_i, f_j)/\tau)}{\sum_{x_k \in N_i} exp(-d(f_i, f_k)/\tau)}\Big), \tag{6}$$

where f_i represents the characteristics of x_i, $d(\cdot, \cdot)$ denotes the distance function, and τ is a temperature coefficient used in contrastive learning. The numerator captures the similarity between positive instances, while the denominator reflects the similarity between positive and negative instances. Consequently, a higher intra-class similarity coupled with a lower inter-class similarity reduces the loss. The temperature coefficient is a predefined hyperparameter that governs model sensitivity to negative samples. Larger values of τ flatten the distribution of $d(f_i, f_j)$, leading to a uniform treatment of all negative samples by contrastive loss, which reduces learning efficacy. Conversely, smaller values of τ focus the model toward particularly challenging negative samples, which could potentially be positive samples. This approach limits model convergence and restricts generalization capabilities. Hence, the determination of an optimal temperature coefficient is critical.

SAM Consistency Module

The SAM framework was used to develop a consistency strategy to mitigate discrepancies arising from the scanning of different subjects. This strategy is comprised of three integral components: a volume feature encoder, a 3D cue encoder, and a volume mask decoder. The principal contributions of each can be specifically delineated as follows. First, an SAM is introduced, which is specifically designed for amygdala segmentation tasks. Unlike existing approaches to improving SAMs, the novelty of our model lies in its comprehensive consideration of factors such as the volume, shape, and relative position of the amygdala. This process eschews conventional slice-by-slice feature extraction in favor of a holistic total volume approach. The model also establishes inter-slice associations, which are exceptionally well-suited to structures with intricate core regions. In contrast to traditional algorithms, this technique provides a more comprehensive representation of anatomical structures, particularly for smaller entities.

Second, the incorporation of a 3D prompt encoder ensures the preservation of prior spatial position information and 3D spatial features provided by the user. In contrast to 2D prompt encoders, this strategy provides more precise relative position cues in multidimensional space. The effective merging of 3D cue information with embedding features derived from a volume feature encoder is then facilitated by a novel volume mask decoder. This decoder harnesses interaction mechanisms between prior cue and MRI embeddings, facilitating seamless integration. The fusion mechanism then adeptly captures nuanced details within the feature space, offering a robust tool for the precise depiction of anatomical structures. SAMs constitute an interactive segmentation methodology predicated on iterative operations. This framework enhances the representational capacity of the model through the incorporation of positional cues and pre-segmentation data, thereby facilitating continuous optimization of segmentation accuracy via multiple interactive iterations. Notably, the accuracy of existing SAMs is suboptimal when applied directly to medical imaging tasks (especially amygdala segmentation). To address this issue, the structure of the proposed 2D SAM has been meticulously customized for amygdala segmentation. This includes the introduction of an advanced SAM, which is comprised of three integral encoders used for volume features, 3D cues, and volume masks.

The function of the Volume Feature Encoder (VFE) is to transform input T1w structural images into an MRI embedding vector. Un-

like the 2D encoder used in traditional SAMs, this process allows simultaneous consideration of volume, shape, and relative position, thereby offering a more comprehensive representation of anatomical structures to enhance the processing of complex regions surrounding the core area. A 3D cue encoder was also used to encode prior information, incorporating interactive cues and existing amygdala segmentation results. The 3D cue encoder consists of two components: a click-cue branch and a pre-segmentation cue branch. The click-cue branch converts user-provided cue points into learnable location embedding features to more precisely delineate amygdala location. The pre-segmentation prompt encoder uses the last click made during interactive segmentation as input. Alternatively, it provides pre-segmentation results to the model, thereby prompting adjustments and optimization of the current segmentation result. As such, a volume mask detector (VMD) was developed to effectively integrate information from the cue encoder with embeddings produced by the feature encoder. This VMD also facilitated the retrieval and refinement of information pertaining to the amygdala, ensuring a seamless alignment of cue feature information and initial feature encoder output.

The VMD is a pivotal component to exhaustively explore key characteristics within the amygdala, thereby enhancing model efficacy in intricate clinical scenarios. It aims to combine MRI embedding vectors (produced by the volume feature encoder) with user-provided prompt inputs (used to determine the correlation between them), culminating in precise predictive results. The 3D mask decoder generates MRI and cue embeddings using a multi-layered fusion 3D feature encoder, facilitating multi-level and multi-perspective information extraction. This comprehensive fusion mechanism enables a more complete assimilation of MR images and improves the differentiation of anatomical structures, thereby increasing predictive precision. In conclusion, the decoder generates precise predictions of specific structures or lesions within images, offering crucial support for clinicians in making accurate diagnoses and devising treatment strategies.

Implementation of Loss Function

We implemented an advanced deep-supervised strategy to formulate the loss function. Specifically, this involved the integration of supervisory signals in each intermediate and terminal module, thereby facilitating enhanced gradient propagation and optimization of network training efficiency. Two distinct loss functions were incorporated for each supervisory signal, namely, binary cross-entropy (BCE) and Dice loss. The overall loss function was constructed as follows:

$$L = \sum_{k=1}^{4} \left(L_{\text{BCE}}^{k} + L_{\text{Dice}}^{k} \right). \tag{7}$$

An adapted version of BCE loss was employed to address issues with imbalances between positive (amygdala) and negative (other tissue) samples. Hereafter, the superscript k is excluded for simplicity, while L_{extBCE} is defined as follows:

$$L_{\text{BCE}} = -\frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(1+\alpha_{i,j})\sum_{l=0}^{1}(g_{i,j}=l)logPr(p_{i,j}=l|\Psi)}{\sum_{i=1}^{H}\sum_{j=1}^{W}(\alpha_{i,j})}, \tag{8}$$

where $l \in (0,1)$ denotes two label categories. The variables $p_{i,j}$ and $g_{i,j}$ represent predicted and ground truth values, respectively, at the spatial coordinate (i,j) within a CT image characterized by a width of W and a height of H. The symbol Ψ encompasses all model parameters and $Pr(p_{i,j}=l)$ signifies the predicted probability. The term $\alpha_{i,j} \in [0,1]$ describes the pixel-specific weight and can be expressed as:

$$\alpha_{i,j} = \left| \frac{\sum_{(m,n)\in A_{i,j}} g_{m,n}}{\sum_{(m,n)\in A_{i,j}} 1} - g_{i,j} \right|, \tag{9}$$

where $A_{i,j}$ denotes the region adjacent to pixel (i,j) and $|\cdot|$ signifies the calculation of absolute value. Furthermore, Dice loss was defined as:

$$L_{\text{Dice}} = 1 - \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}(2g_{i,j}p_{i,j})(1+\alpha_{i,j})}{\sum_{i=1}^{H}\sum_{j=1}^{W}(g_{i,j}+p_{i,j}+g_{i,j}p_{i,j})(1+\alpha_{i,j})}. \tag{10}$$

Testbed Data and Evaluation Metrics

The proposed algorithm was implemented using the PyTorch framework, utilizing the full capabilities of the Tesla A10 GPU to optimize computational efficiency. The AdamW optimizer[25], a common choice within the transformer architecture domain, was used

for routine updates of network parameters. The learning rate was set to $1e^{-4}$ and the weight decay was equally adjusted to $1e^{-4}$ to achieve an equilibrium between training stability and model convergence speed. Memory consumption was mitigated by designating a batch size of 1, with a training process encompassing 100 iterations. This configuration ensured the model underwent sufficient training cycles to achieve optimal performance and effectively discern feature representations. During the testing phase, input image dimensions were down-sampled to $48 \times 64 \times 80$ to minimize computational demands. No post-processing optimization techniques were employed, thereby ensuring an objective assessment of model performance.

In this study, human brain images for the testbed of AmygdalaGo-BOLT3D were from MRI brain scans of three cohorts: the Chongqing cohort of the Chinese Color Nest Cohort (CKG-CCNC) [26–28], the Beijing Normal University Childhood Adversity Cohort (BNU-CAC) [29], and the Zhejiang University Multi-center Traveler Cohort (ZJU-MTC) [30]. CKG-CCNC is part of the CCNP development component (devCCNP), which was collected at Southwest University in Chongqing, China. The devCCNP aims to map normative brain development trajectories within the Chinese population during school years. The devCCNP-CKG dataset includes samples from 201 participants aged 6-17 years, who were invited to participate in three consecutive rounds of data collection at intervals of approximately 1.25 years. At these time points, T1-weighted magnetic resonance imaging (T1w) was performed and images were visually inspected to exclude those with significant head movement artifacts or other structural abnormalities. After this initial quality control phase, the final samples included 427 scans from 198 participants (105 girls and 93 boys). The distribution of participants with one, two, and three scans was 48, 71, and 79, respectively, with an average of 2.16 scans per participant. BNU-CAC, which aims to investigate the impact of childhood adversity on brain development, consists of 116 subjects (60 girls and 56 boys) aged 8-13 years who were scanned once per individual and passed quality control protocols. ZJU-MTC includes three adult travelers [30] to evaluate image consistency under identical parameters, with data collection conducted at 10 scanning facilities to ensure comparability between centers. Model performance was rigorously assessed by incorporating standard evaluation metrics used for medical image segmentation, primarily encompassing similarity metrics and distance-based metrics. These included the Dice coefficient [31], Jaccard coefficient (JS), average surface distance (ASD), average symmetric surface distance (ASSD) [32], and the relative volume difference (RVD) [33]. Precision and recall metrics derived from the confusion matrix were also utilized [34], as described below.

The Dice similarity coefficient (Dice) is used to measure the similarity between a predicted binary region and a ground truth (GT). It can be represented as:

$$Dice = \frac{2|V_{\text{Seg}} \cap V_{\text{GT}}|}{|V_{\text{Seg}}| + |V_{\text{GT}}|}, \tag{11}$$

where V_{Seg} and V_{GT} denote voxel sets for the segmented area and GT, respectively. The term $|\cdot|$ denotes a cardinality computation operation, which provides the number of elements in a set.

Intersection over Union (IoU) is the area of intersection between predicted segments and labels, divided by the area of union between the two, expressed as:

$$IoU = \frac{|V_{\text{Seg}} \cap V_{\text{GT}}|}{|V_{\text{Seg}} \cup V_{\text{GT}}|}. \tag{12}$$

Average Surface Distance (ASD) is the average distance between surface voxels $S(V_{\text{Seg}})$ in the segmentation result V_{Seg} and surface voxels $S(V_{\text{GT}})$ in V_{GT}. This metric can be defined as:

$$ASSD = \frac{1}{|S(V_{\text{Seg}})| + |S(V_{\text{GT}})|} \quad \left(\sum d\left(s_{V_{\text{Seg}}}, S(V_{\text{GT}})\right) + \sum d\left(s_{V_{\text{GT}}}, S(V_{\text{Seg}})\right) \right), \tag{13}$$

where $d(s, S(X))$ is the minimum Euclidean distance from the voxel s to surface voxels in the volume X.

Relative Volume Difference (RVD) is originally used to measure differences between predicted lung infections and a GT. Here, we calculated the absolute value of RVD and report the corresponding results as follows:

$$RVD = \left|\frac{|V_{\text{Seg}}| - |V_{\text{GT}}|}{|V_{\text{GT}}|}\right|. \tag{14}$$

The Asymmetry Index (AI) is used to assess the degree of lateralization for the left and right amygdala and is defined as:

$$AI = \frac{V_{left} - V_{right}}{V_{left} + V_{right}} \times 100\%. \tag{15}$$

Inter-class Correlation Coefficient (ICC) measures the consistency of results across multiple imaging centers and is defined as:

$$ICC = \frac{(MSB - MSW)}{MSB + (k-1) * MSW}, \tag{16}$$

where MSB denotes the mean square between groups, MSW is the mean square within groups, and k is the number of raters.

RESULTS

Assessments on Tracing Capability in Developing Children and Adolescents

Method	Dice	JS	HD95	ASD	ASSD	Precision	Recall	RVD
Freesurfer	0.776	0.636	2.530	1.093	0.954	0.732	0.832	0.142
U-Net3D	0.851	0.743	1.565	0.746	0.693	0.774	0.951	0.239
U-Net3D++	0.787	0.651	1.963	1.064	0.982	0.658	0.985	0.512
ResU-Net3D	0.841	0.728	1.669	0.806	0.736	0.754	0.957	0.279
BOLT3D	0.904	0.825	1.097	0.483	0.467	0.883	0.929	0.056

Table 1: Performance evaluation on different methods of tracing amygdala in the CKG-CCNC cohort.

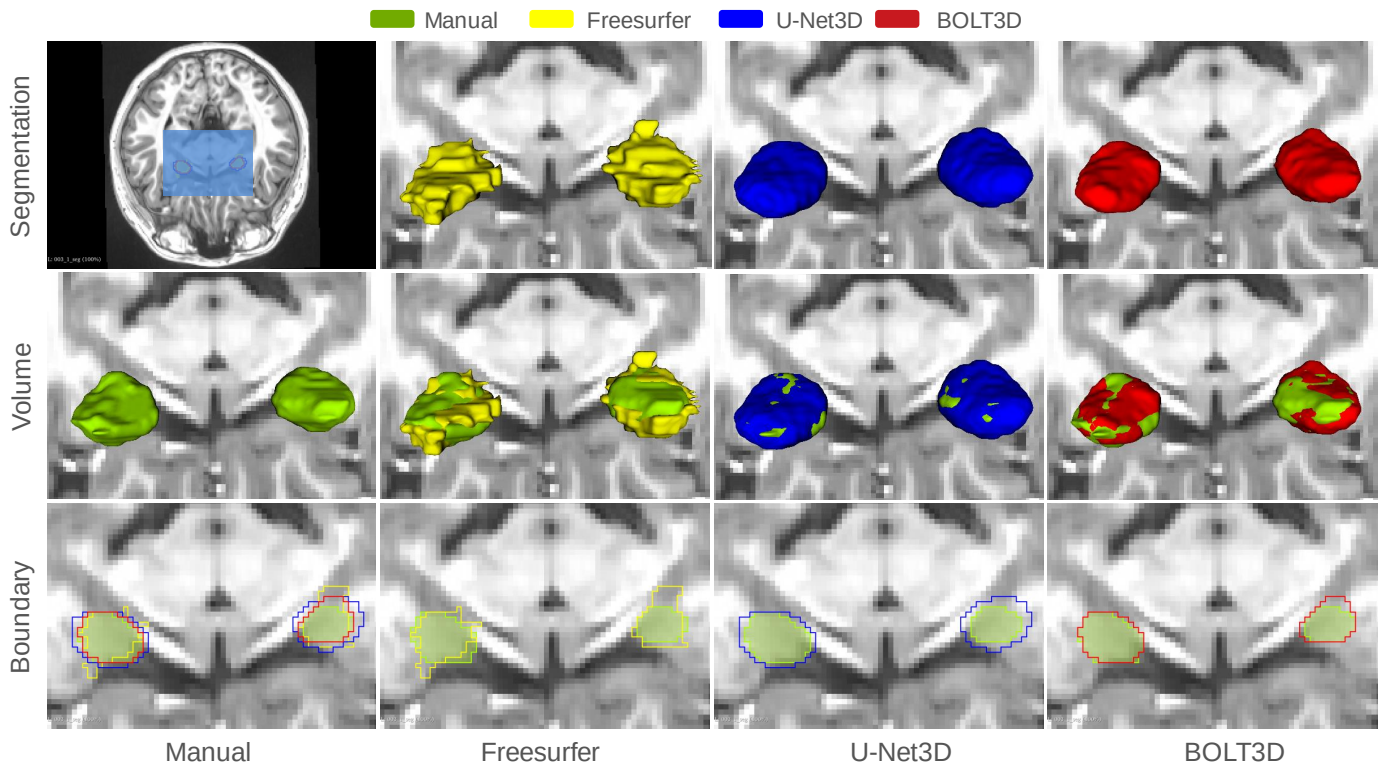

Figure 2: A comparative visual examination of the amygdala in a representative subject, illustrating the preservation of structural integrity (middle row) and the precision of boundary delineation (bottom row). The manual segmentation results (left column) were established as a reference standard for accuracy. As shown, our proposed BOLT3D method (right column) demonstrates superior performance in both volumetry and boundary definition, outperforming Freesurfer (second column) and U-Nest3D (third column).

A comprehensive analysis of individual metrics and the various benefits offered by BOLT3D are presented in Table 1, where it is evident that BOLT3D demonstrates superiority within the test cohort, achieving exemplary performance in critical metrics such as Dice, JS, and HD95, relative to conventional algorithms. Specifically, the Dice and JS coefficients serve as metrics for evaluating the extent of overlap in segmentation outcomes. Notably, BOLT3D achieved a Dice coefficient of 0.904, which surpassed alternative

algorithms such as U-Net3D (0.851), U-Net3D++ (0.787), and ResU-Net3D (0.841). The JS coefficient demonstrated a comparable trend, as BOLT3D (0.825) clearly outperformed other algorithms, underscoring its superior segmentation precision. The HD95 and ASSD metrics quantify the spatial discrepancy between segmentation outcomes and reference annotations, specifically highlighting the interval between segmented amygdala boundaries and true structural boundaries. BOLT3D also achieved a markedly smaller Hausdorff distance of 1.097, relative to other algorithms such as U-Net3D (1.565), U-Net3D++ (1.963), and ResU-Net3D (1.669). This result demonstrates that BOLT3D excelled in preserving the spatial congruence between segmented peripheries and actual structural demarcations. The precision and recall metrics were used to further assess false positives. Precision quantifies the proportion of true positive instances among those designated as positive by the model, while recall evaluates the proportion of true positive instances correctly identified by the model among all positive instances in the dataset. The precision and recall achieved by BOLT3D were 0.883 and 0.929, respectively, surpassing the performance of other algorithms. BOLT3D also exhibited a balance between these metrics, resulting in a superior overall performance. In addition, BOLT3D achieved a relative RVD of 0.056, significantly lower than that of other algorithms, thereby demonstrating exceptional accuracy in volume predictions.

Figure 2 shows comparative segmentation outcomes from multiple algorithms on a representative subject, illustrating the superior performance of the proposed algorithm, as measured by volumetric and boundary correspondence with manual segmentation samples. Notably, the amygdala delineated by Freesurfer deviates markedly from both the morphological and manual segmentation results. The presence of multiple spaces indicates an inability to accurately capture the structural integrity of the amygdala. A closer inspection of the periphery suggests that Freesurfer exhibits numerous irregularities when compared with the gold standard (GS). The amygdala delineated by Unet exhibits morphological characteristics comparable to those of the GS. However, an assessment of the superposition effect and edge segmentation results revealed that volumetric delineations produced by Unet were substantially overestimated. In contrast, results from the proposed BOLT3D model exhibit a morphological resemblance that is closer to the GS. Superposition effects in the BOLT3D samples were similarly congruent with the GS, demonstrating no significant volumetric overestimation or underestimation, as corroborated by edge segmentation results. The superiority of BOLT3D is primarily attributable to the integration of Transformer technology with a boundary contrast learning algorithm, which improved the model's ability to capture intricate structural details in images while preserving the overall structural integrity.

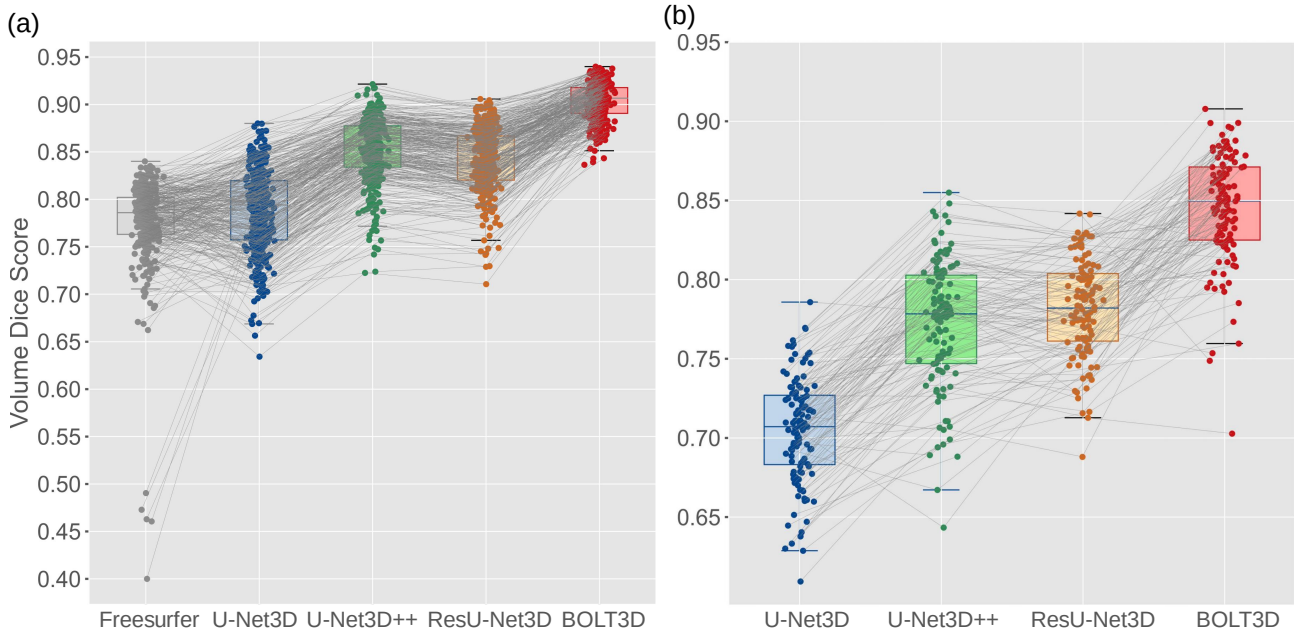

Figure 3: Volume dice analysis of the overall amygdala structure for samples from the CKG-CCNC (a) and BNU-CAC (b) cohorts. Since the CKG-CCNC data were used during training, the overall performance was slightly higher than for the unseen BNU-CAC dataset for all deep-learning network methods. However, the proposed BOLT3D model achieved a dice score above 0.8 for both cases, which was even higher than that of FreeSurfer.

Assessments on Tracing Capability in Traveling Adults

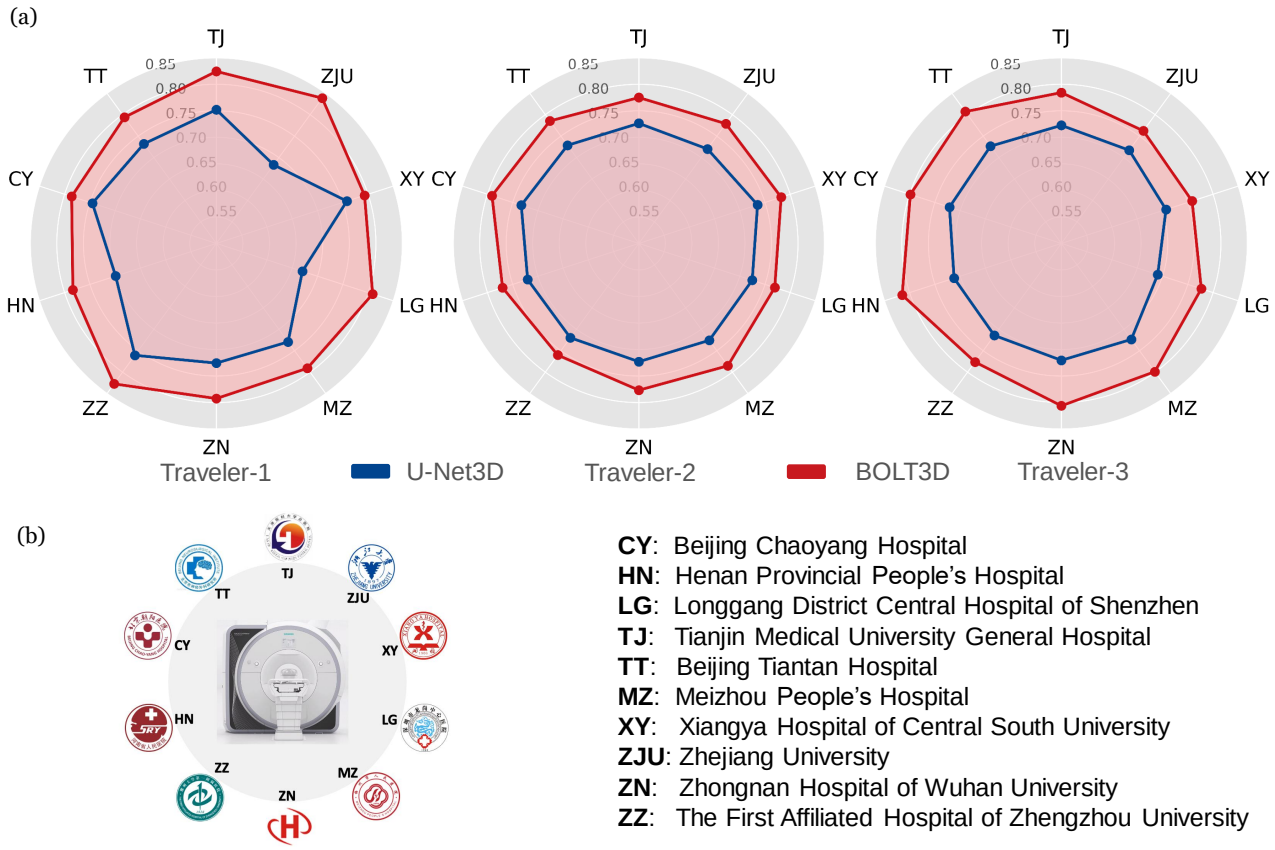

Figure 4: A comparison of amygdala segmentation outcomes for itinerant volunteers from ZJU-MTC. (a) The proposed BOLT3D methodology (depicted in red) consistently demonstrated superior performance relative to U-Net3D (depicted in blue) across all centers for each participant, thereby underscoring its robust cross-center generalization efficacy. (b) All centers were equipped with identical model 3T scanners, geographically dispersed across China, with their respective names and abbreviations provided on the right.

Method	Dice	JS	HD95	ASD	ASSD	Precision	Recall	RVD
U-Net3D	0.705	0.545	2.860	1.445	1.286	0.551	0.980	0.789
U-Net3D++	0.773	0.631	3.551	1.423	1.129	0.665	0.927	0.404
ResU-Net3D	0.782	0.643	2.177	1.107	0.969	0.663	0.955	0.448
BOLT3D	**0.845**	**0.733**	**1.594**	**0.654**	**0.666**	**0.933**	**0.775**	**0.168**

Table 2: Performance evaluation on different methods of tracing amygdala in the ZJU-MTC cohort.

The generalization capabilities of the model were initially validated using data from previously uncounted imaging centers, encompassing a cohort of 116 subjects (see Table 2). The BOLT3D algorithm exhibited robust performance across similarity metrics, achieving a Dice coefficient of 0.845 and a JS of 0.733. This suggests that despite variations in the data distribution, overall algorithm performance remained mostly unchanged. A distance evaluation of BOLT3D produced HD95 and ASD values of 1.594 and 0.654, respectively. Furthermore, the resulting ASSD of 0.666 exhibited less fluctuation compared to the Unet algorithm. BOLT3D also achieved precision and recall scores of 0.933 and 0.775, respectively, underscoring its efficacy in correctly identifying true positives while minimizing false negatives. The equilibrium point of the precision-recall distribution further underscores the ability of BOLT3D to maintain high performance across both precision and recall metrics. The RVD of BOLT3D was recorded at 0.168, reflecting its precision for predicting amygdala volumes with relatively minimal deviations. These indicators suggest the performance of BOLT3D was not compromised significantly, despite the distribution variations encountered at new centers. In addition, this technique significantly outperformed comparable algorithms, thereby affirming the multi-center accuracy and consistency offered by integrating the consistency module.

This section presents results from 116 unseen subjects in BNU-CAC, contrasted with 347 observed scans from CKG-CCNC (see Fig-

ure 3). Given the BNU-CAC subjects represent different data distributions and the model has not previously trained on these data, the overall performance index is lower compared to the CKG-CCNC scans. This discrepancy arises because the model exhibits diminished performance with unfamiliar data relative to familiar data, stemming from variations in feature distributions, inter-data noise, or other unidentified factors that degrade performance. Despite this reduction in the overall index with novel data, the model achieved substantial performance improvements compared to the baseline comparison model, indicating the algorithm offers significant generalization capabilities. These experiments employed three travelers from ten distinct centers to determine the stability and generalizability of the model between multiple facilities (see Figure 4). It is evident from these results that BOLT3D achieved superior segmentation accuracy for each of the three subjects from all ten centers, producing notable improvements (ranging from 0.02 to 0.07) over U-Net 3D. This underscores the robustness of BOLT3D in multi-center image segmentation tasks.

Ablation Study

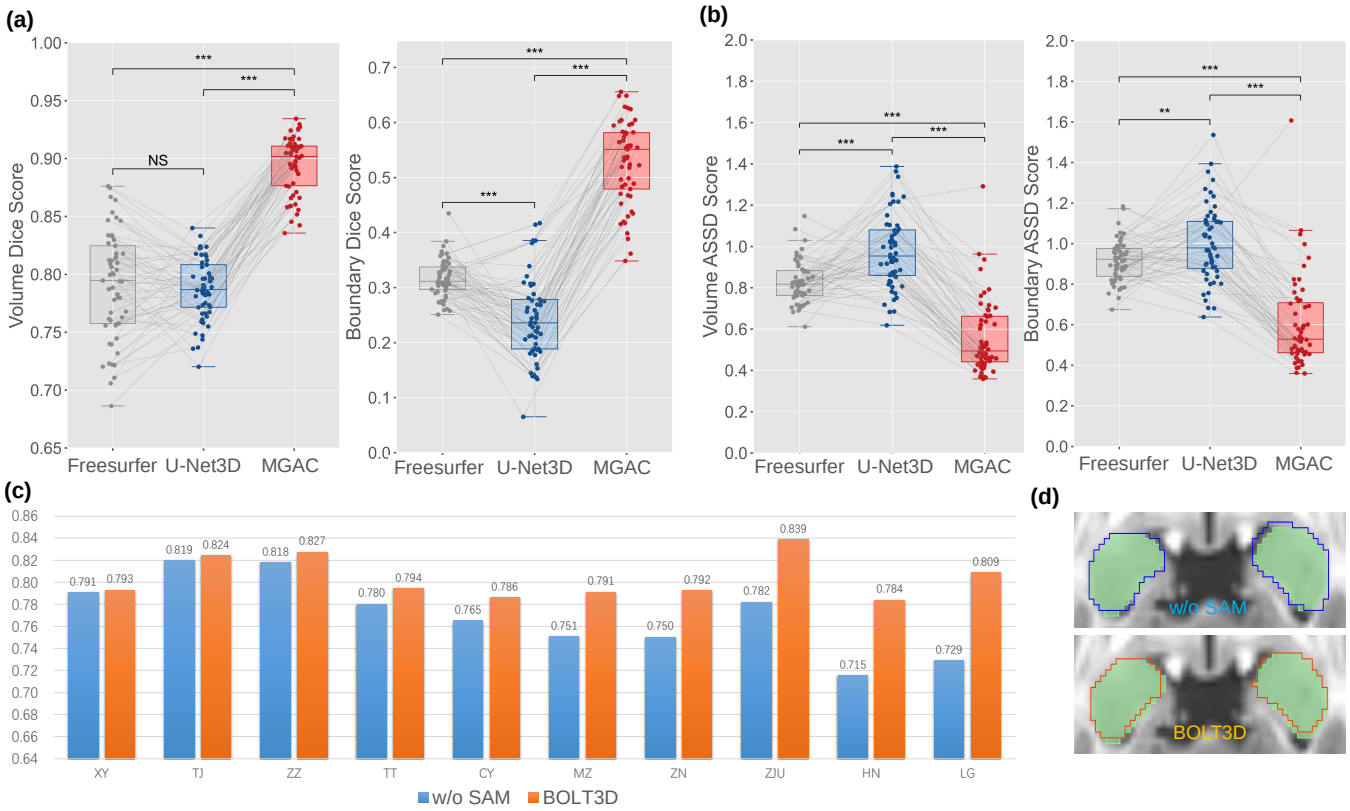

Figure 5: The results of ablation experiments. A comparative analysis of consistency confirmed the transformer-based algorithm (MGAC) within BOLT3D significantly reduced discrepancies when compared with U-Net3D and Freesurfer, in terms of volumetric measurement and boundary delineation, as evidenced by their corresponding (a) Dice and (b) ASSD metrics. This illustrates the superior precision of our algorithm for accurately tracing structural margins and effectively reducing the risk of volumetric overestimation. (c) Integration of the SAM module, in conjunction with the MGAC module, markedly enhanced the robustness of BOLT3D in the ZJU-MTC study, which was quantitatively evaluated using Dice scores. (d) A representative instance both with and without the SAM module, demonstrating the impact of these enhancements on segmentation quality.

The stability of the model was evaluated by analyzing the distribution of Dice and ASSD scores between individuals using both volume and boundary metrics. The consistency analysis depicted in Figure 5(a) demonstrates that our algorithm achieved markedly lower discrepancies than U-Net3D or Freesurfer, thus confirming its stability across a substantial cohort of subjects. The Dice coefficient from the BOLT3D algorithm nearly reached a value of 0.9 (see Figure 5(b)), surpassing the average performance of both Freesurfer and U-Net3D. The proposed MGAC leverages a Transformer mechanism that combines global coarse-grained perceptual processes with local fine-grained enhancement techniques. Such mechanisms enable the acquisition of both global and local amygdala features, while the adaptive multi-granularity fusion mechanism extracted abundant semantic information from varied scales used for optimization. Consequently, the MGAC algorithm achieved remarkable stability and superior efficacy for comprehensive structural processing. The consistency of the segmentation results was assessed using multiple repeated scans of a repre-

sentative subject at various imaging centers, which was imperative for substantiating the effectiveness of the consistency-module. A quantitative comparison between ten MR centers is presented in Figure 5(c), which confirms the SAM module accurately identified and maintained target consistency, producing stable results under diverse scanning conditions. In addition, the optimized segmentation algorithm exhibited exceptional performance in an experiment involving 116 subjects from the BNU-CAC cohort.

Foster Fast and Valid Large-scale Amygdala-centered Neuroscience Research

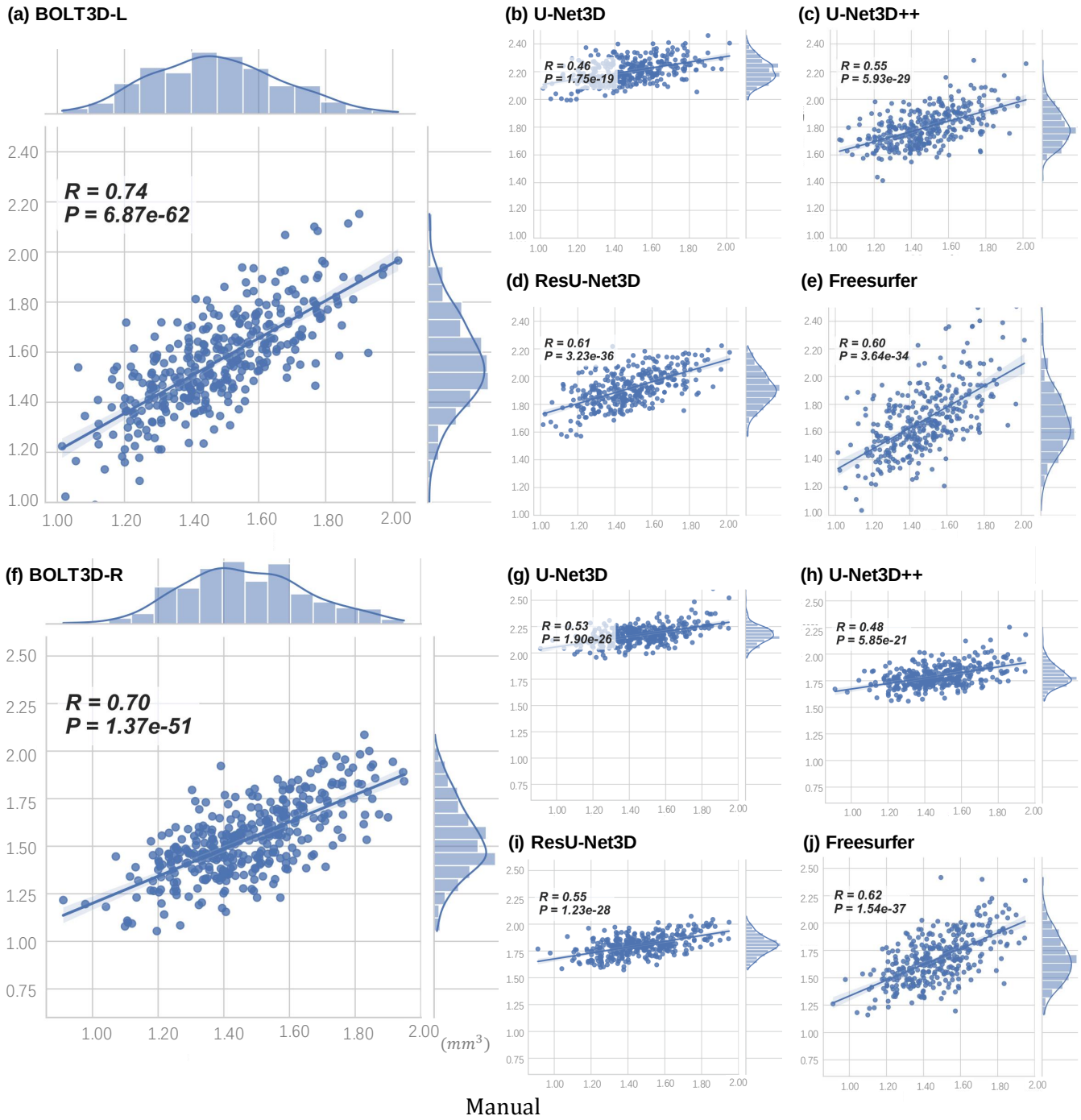

Figure 6: A correlation analysis of amygdala volume as determined by segmentation algorithms compared with manual reference standards. The dataset included 347 scans, with the results of manual segmentation depicted on the horizontal axis. Among the evaluated methods, the proposed BOLT3D algorithm achieved the highest correlation with manual references for the left (a) and right (f) regions of the amygdala, signifying higher consistency in volumetric measurements. In contrast, correlation coefficients for alternative segmentation tools, including U-Net3D (b, g), ResU-Net3D (c, h), U-Net3D++ (d, i), and Freesurfer (e, j), fell within a range of 0.46 to 0.60. The volume distributions derived from each algorithm are represented on the respective vertical axes.

We first calculate the Pearson correlation coefficient of segmented volume between the five automatic methods with the 'gold-standard' manual segmentation to assess their segmentation validity of tracing boundaries of the amygdala in 347 developing brains. Figure 6 shows the results from U-Net3D, U-Net3D++, ResU-Net3D, Freesurfer, and BOLT3D, with correlation coefficients and corresponding P-values provided for each algorithm. These findings indicate varying degrees of correlation with manual segmentation, though some exhibited substantial overestimation. In particular, BOLT3D demonstrated exceptional consistency for volumetric estimation, aligning closely with manual segmentation results and manifesting a substantially positive linear relationship. The correlation coefficient for BOLT3D reached a value of 0.74, exceeding the range of 0.46 to 0.61 observed with other algorithms. Despite the strong correlation results achieved by algorithms such as Freesurfer and U-Net3D, BOLT3D achieved a more definitive overall trend. The elevated correlation coefficient observed in the analysis of the left amygdala underscores the superior performance of BOLT3D for volumetric detection across different structures or tissues, achieving considerable agreement with manual segmentation results. In addition, the narrow dispersion of scatter points accentuates the exceptional consistency of individual measurements obtained with BOLT3D, further attesting to its remarkable stability.

To further demonstrate consistency and comparability between automated segmentation and manual segmentation, we perform and present a detailed analysis of systematic discrepancies in the quantification of amygdala volume based on the 347 pediatric MRI scans. As shown in Figure 7, our method demonstrated superior consistency compared to the other five algorithms used for volume measurements. In this representation, the upper and lower lines denote the 95% limits of agreement, while the central line signifies the mean difference. A qualitative examination of the data distribution near the mean reveals that our algorithm exhibits a moderate and concentrated spread within the range of volume measurements. Specifically, the proposed BOLT3D algorithm registered an average volume discrepancy of -0.152mm^3, while U-Net3D, U-Net3D++, Freesurfer, and ResU-Net3D produced average volume discrepancies of -1.433mm^3, -0.647mm^3, -1.874mm^3, and -0.775mm^3, respectively, indicating significant overestimation. The inflated volumes reported by these algorithms are 9.427, 4.256, and 5.098 times that of BOLT3D, underscoring the substantial advantage of the proposed model in volumetric measurement precision. Notably, minimal systematic deviations were observed between BOLT3D and manually adjusted volumes, highlighting its consistency in the quantification of amygdala volumes.

	Manual	BOLT3D	Freesurfer	U-Net3D	U-Net3D++	ResU-Net3D
Curve edf	1.33	1.59	2.09	2.38	3.02	2.84
Volume (mm^3)	1.47	1.56	1.73	2.20	1.80	1.91

Table 3: Statistics of predictive models on growth curves estimated by different methods of tracing the human amygdala.

Finally, we plot growth curves representing volume changes for the amygdala (derived algorithmically with age) to provide insights into the effectiveness of these algorithms for forecasting growth curves from childhood and adolescence to early adulthood. These results were further evaluated using their alignment with manually annotated results[13]. Specifically, as shown in Figure 8, growth curve patterns were parallel in girls and boys for both manual and automated tracing methods, although boys exhibited larger amygdala volumes than girls across the entire age range. As a reference standard, the manual tracing method revealed the human amygdala (both left and right) exhibited linear growth during childhood and adolescence in both boys and girls. BOLT3D also produced a linear trend consistent with that of manual segmentation (see Figure 8 and estimated degree freedom (edf) in Table **??**). Although the bilateral amygdala volumes predicted by BOLT3D were slightly larger than those of manual annotations, the scatter plot exhibits a more confined aggregation among age cohorts. The FreeSurfer method also exhibits inverted U-shaped curves, which diverge from the linear trend observed in manual segmentation (see Figure 8). Thus, the scatter plots reveal an increase in variability between age cohorts, with bilateral amygdala volumes exceeding those derived from both manual annotations and BOLT3D, which produced growth curves very similar to those established by the manual tracing method. This outcome demonstrates significant advantages for the fitting of growth curves compared to other algorithms, attesting to the robustness and accuracy of BOLT3D in considering gender- and age-related changes when estimating amygdala volume. Compared to conventional and other advanced algorithms, this technique excels at capturing subtle changes in growth curves, not only in terms of overall adaptation efficiency but also in terms of accurately predicting curve details and particular features. These findings improved our understanding

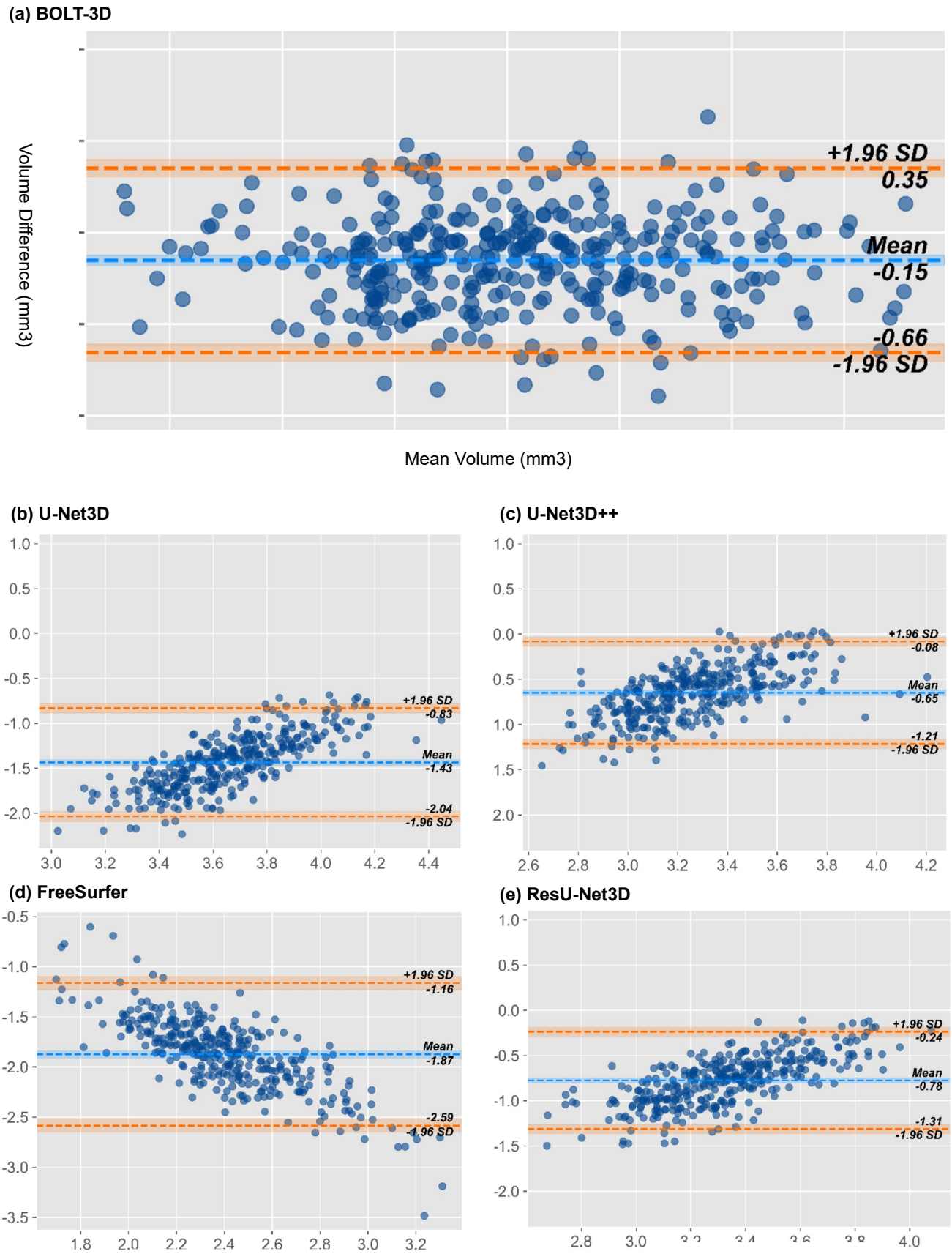

Figure 7: Bland-Altman plots employed to assess the agreement of volume size estimation between various segmentation algorithms and the manual reference standard across 347 scans of the left portion of the amygdala. The central dashed blue lines denote the mean differences, while the dashed yellow lines represent the limits of agreement at ± 1.96 standard deviations from the mean. Each blue dot represents the data for an individual subject.

of amygdala development over time with respect to gender differences and age-related factors.

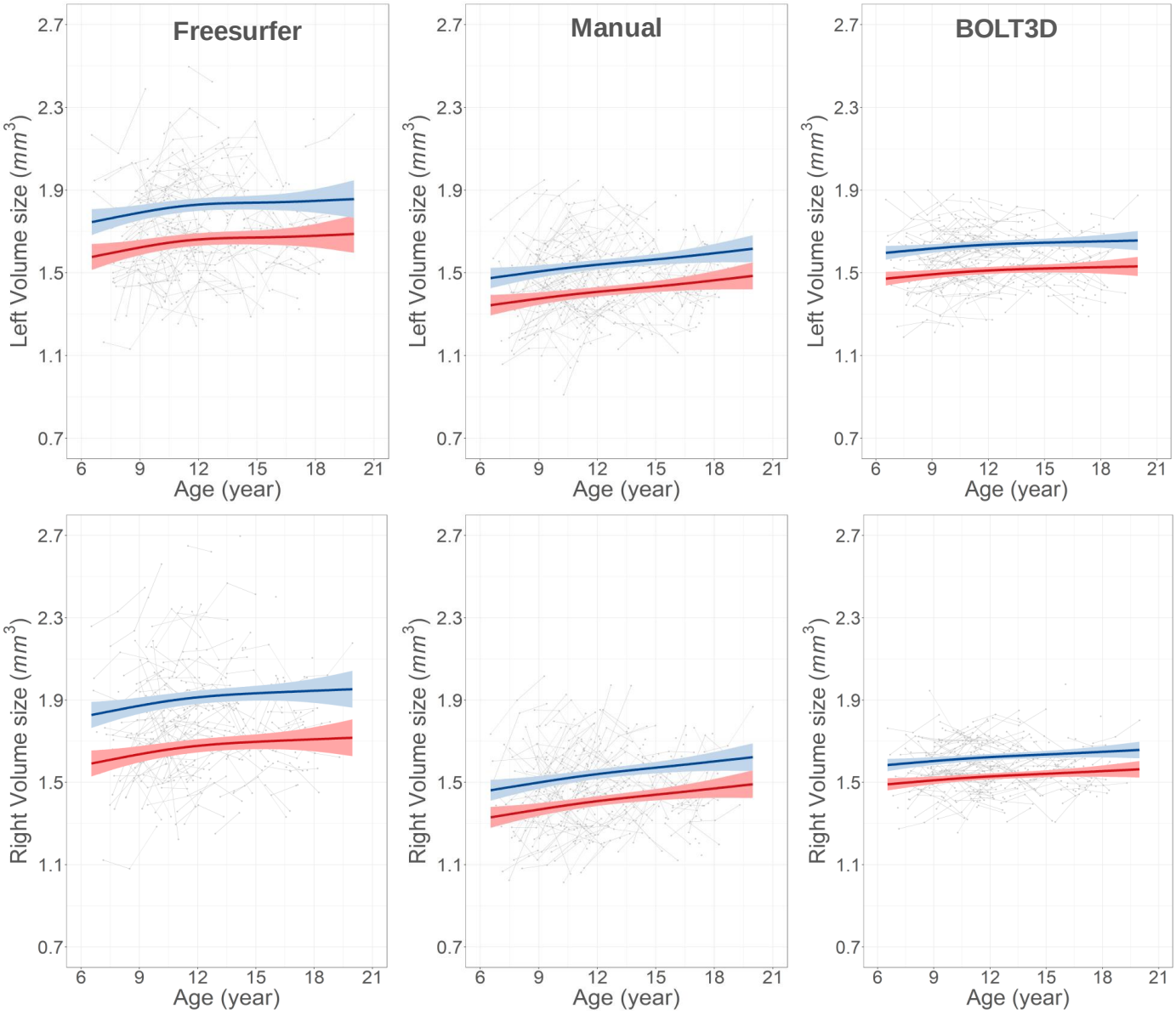

Figure 8: The developmental trajectory of the amygdala, as delineated by our model and comparative methodologies. The FreeSurfer results (left) exhibit an inverted U-shaped curve, indicative of a non-linear growth pattern. Conversely, the growth curves predicted by our BOLT3D model (right) exhibit a striking agreement with the linear trajectory observed in the manual tracing method (middle). Blue denotes trajectories for males, while red represents trajectories for females. These trajectories are enveloped by shaded 95% confidence intervals. Individual scatter points represent segmentation outcomes from distinct cases, while the lines connecting these points indicate results acquired from multiple scans of the same participant.

AmygdalaGo-BOLT3D Model Availability

The AI training platform supporting this work was provided by High-Flyer AI (Hangzhou High-Flyer AI Fundamental Research Co., Ltd.). We deploy the optimal model parameters to be applied to any user-specified dataset in the interest of open science as well as provide tutorials and source codes at the Science Data Bank (SciDB_LINK).

DISCUSSION

Deep-learning methods of tracing boundaries of amygdala are highly valuable for reproducibility by providing standardized, automated pipelines that minimize the variability introduced by manual tracing or traditional atlas-based methods. In open neuroscience, where large-scale datasets are shared across institutions, AI-driven segmentation ensures consistent and reproducible annotations, making results more comparable across studies. In the present work, we develop an open-source AI tool, AmygdalaGo-

BOLT3D, which allows researchers to collaborate, validate, and refine models using publicly available datasets, fostering transparency and innovation. Traditional segmentation techniques, whether manual or atlas-based, are time-consuming and prone to inter-rater variability, especially for small structures like the amygdala. A deep-learning model such as AmygdalaGO-BOLT3D trained on large datasets can provide 1) higher segmentation accuracy, especially for challenging cases with low-contrast boundaries; 2) faster processing, reducing the time required for segmentation from hours to seconds; 3) improved generalization, allowing to adapt to different populations, imaging protocols, and scanner types. another example of AI-based segmentation using multi-atlas priors and 3D convolutional networks[35] has demonstrated superior performance over traditional methods. Our amygdalaGo-BOLT3D addressed the intricacies posed by low contrast and structural complexities that distinguish the amygdala from adjacent tissues. A boundary contrastive learning algorithm was efficient to more accurately align model-predicted boundaries with a ground truth, culminating in more precise segmentation across a wide range of subject ages, ranging from childhood and adolescence to adulthood. It also maintained remarkably consistent performance when deployed in multi-center adult datasets, underscoring exceptional generalization proficiency and robust performance across varied data sources and contexts. This transformer could be a valuable new tool in the field of cognitive neuroscience. For example, previous studies have identified a relationship between the size of social networks and the volume of the amygdala[36–38].

Recent advances in open neuroimaging data have significantly increased our understanding of the intricate structures and networks within the human brain[39]. The aggregation of images from various large-scale data repositories has significantly improved statistical power, allowing for the discernment of subtle variability within the human amygdala[40,41]. However, there is a pressing need for new research methodologies to enhance the computational efficiency and practical applicability of these automated extraction tools[13,42]. Global imaging collaborations have been instrumental in the evolution of population neuroscience[43,44], driven by extensive MRI initiatives[45], which have been crucial for driving innovations in analytical techniques and translational research[46]. As an open-source deep-learning segmentation tool, AmygdalaGO-BOLT3D democratizes access to high-quality neuroimaging analysis, making it possible for researchers without extensive expertise in neuroanatomy to conduct high-precision studies. This is crucial for 1) low-resource institutions that may lack access to expert manual tracers; 2) large-scale, multi-site collaborations where standardized, automated segmentation ensures data harmonization; 3) early-career researchers and students who can leverage open-source AI tools without requiring specialized training in segmentation techniques. By integrating deep learning into open neuroscience platforms, the AmygdalaGO-BOLT3D pipeline can be widely distributed, continually improved, and used globally, accelerating discoveries in neurodevelopment, psychiatric disorders, and neurodegenerative diseases.

We note that the present study lacks a detailed subdivision of the amygdala, specifically finer-grained nuclei labeling. Future work will focus on providing granular annotations and establishing a computational baseline to support a more nuanced analysis of the amygdala. Unlike static atlas-based segmentation tools, AmygdalaGO-BOLT3D improves over time by leveraging crowdsourced data and federated learning approaches. This means that as new datasets become available, its model can be updated, enhancing its robustness and adaptability. Community-driven platforms such as OpenNeuro, NeuroVault, and ScienceDB can serve as repositories for the model training and validation, ensuring that this AI-driven technique remains cutting-edge and inclusive.

CONCLUSION

The deep-learning method of tracing boundaries from human brain images, AmygdalaGo-BOLT3D, offers a scalable, accurate, and open-access solution for human amygdala segmentation, aligning perfectly with the goals of open neuroscience in terms of: 1) Reproducibility - AI ensures consistent results across studies; 2) Efficiency - Faster and more precise segmentation compared to manual tracing; 3) Accessibility - Enables global participation in neuroimaging research; 4) Adaptability - The Model improves over time with diverse datasets. As open neuroscience continues to evolve, AI-driven segmentation will be instrumental in unlocking large-scale brain imaging insights, leading to more inclusive, transparent, and impactful discoveries in human neuroscience of amygdala.

ACKNOWLEDGMENTS

The STI 2030 - the major projects of the Brain Science and Brain-Inspired Intelligence Technology (2021ZD0200500), the Start-up Funds for Leading Talents at Beijing Normal University, the National Natural Science Foundation of China (82102134, 82322035, 62273076), the Key-Area Research and Development Program of Guangdong Province (2019B030335001), the Beijing Municipal Science and Technology Commission (Z161100002616023, Z181100001518003), the Major Project of National Social Science Foundation of China (20&ZD296), the CAS-NWO Programme (153111KYSB20160020), the National Basic Research (973) Program (2015CB351702), the Chinese Academy of Sciences Key Research Program (KSZD-EW-TZ-002) and the Major Fund for International Collaboration of National Natural Science Foundation of China (81220108014) provide funding supports.

AUTHOR CONTRIBUTIONS

Conception and Design: H-J.H., X-N.Z. Data Analysis: B.D., Q.Z., P.G. MRI Data Preprocessing: Q.Z., P.G. Initial Drafting of the Manuscript: B.D., Q.Z. Critical Review and Editing of the Manuscript: All authors contributed to the critical review and editing of the manuscript.

AUTHOR COMPETING INTERESTS

Competing interests

The authors declare that they have no competing interests.

DATA AVAILABILITY

The datasets analyzed in this study are publicly available from the Chinese Color Nest Project - Lifespan Brain-Mind Development Data Community.

CODE AVAILABILITY

All code used in the analyses is available at the Science Data Bank. For detailed information on how to use these codes, please contact the first author by email.

For the Chinese Color Nest Consortium (CCNC)

Xi-Nian Zuo[1,2,3,4,7,18], Ning Yang[1,2,3,4], Zhe Zhang[5], Ye He[6], Hao-Ming Dong[1,4,8], Lei Zhang[9], Xing-Ting Zhu[2], Xiao-Hui Hou[7], Yin-Shan Wang[1,3], Quan Zhou[1,2,3], Zhu-Qing Gong[1,2,3], Li-Zhi Cao[2,4], Ping Wang[1], Yi-Wen Zhang[2], Dan-Yang Sui[20], Ting Xu[21], Gao-Xia Wei[2,4], Zhi Yang[2,4], Lili Jiang[2,4], Hui-Jie Li[2,4], Ting-Yong Feng[11], Antao Chen[11], Jiang Qiu[11], Xu Chen[11], Xun Liu[2,4], Ke Zhao[2,4], Yi Du[2,4], Min Bao[2,4], Yuan Zhou[2,4], Yan Zhuo[14], Zhentao Zuo[14], Li Ke[1], Fei Wang[22], Zhi-Xiong Yan[7], Xue-Quan Su[7], F.Xavier Castellanos[23], Michael Peter Milham[21,24], Yu-Feng Zang[25]

[1]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, China.
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China.
[3]Developmental Population Neuroscience Research Center, International Data Group/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China.
[4]Key Laboratory of Behavioural Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China.
[5]College of Education, Hebei Normal University, Shijiazhuang, 050024, China.
[6]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, 100876, China.
[7]Laboratory of Cognitive Neuroscience and Education, School of Education Science, Nanning Normal University, Nanning, 530299, China.
[8]Changping Laboratory, Beijing, 102206, China.
[9]School of Government, Shanghai University of Political Science and Law, Shanghai, 201701, China.
[10]School of Psychology, Research Center for Exercise and Brain Science, Shanghai University of Sport, Shanghai, 200438, China.
[11]Faculty of Psychology, Southwest University, Chongqing, 400715, China.
[12]NYU-ECNU Institute of Brain and Cognitive Science at New York University Shanghai, Shanghai, 200062, China.
[13]Faculty of Arts and Science, New York University Shanghai, Shanghai, 200122, China.
[14]State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China.
[15]School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China.
[16]Department of Psychology, Renmin University of China, Beijing, 100872, China.
[17]Beijing Key Laboratory of Learning and Cognition, School of Psychology, Capital Normal University, Beijing, 100048, China.
[18]School of Education, Hunan University of Science and Technology, Hunan Xiangtan, 411201, China.
[19]National Basic Science Data Center, Beijing, 100190, China.

REFERENCES

[1] Sara Sammallahti, Fadila Serdarevic, and Henning Tiemeier. Excessive crying, behavior problems, and amygdala volume: A study from infancy to adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, 62(6):675–683, 2023.

[2] Ya Chai, Philip Gehrman, Meichen Yu, Tianxin Mao, Yao Deng, Joy Rao, Hui Shi, Peng Quan, Jing Xu, Xiaocui Zhang, et al. Enhanced amygdala–cingulate connectivity associates with better mood in both healthy and depressive individuals after sleep deprivation. *Proceedings of the National Academy of Sciences*, 120(26):e2214505120, 2023.

[3] Denis Paré and Drew B Headley. The amygdala mediates the facilitating influence of emotions on memory through multiple interacting mechanisms. *Neurobiology of Stress*, page 100529, 2023.

[4] Juhyun Song. Amygdala activity and amygdala-hippocampus connectivity: Metabolic diseases, dementia, and neuropsychiatric issues. *Biomedicine & Pharmacotherapy*, 162:114647, 2023.

[5] Ziv Ben-Zion, Nachshon Korem, Naomi Fine, Sophia Katz, Megha Siddhanta, Melissa C Funaro, Or Duek, Tobias R Spiller, Sarah Danbock, Ifat Levy, et al. Structural neuroimaging of hippocampus and amygdala subregions in post-traumatic stress disorder (ptsd): A scoping review. *Biological Psychiatry Global Open Science*, 2023.

[6] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[7] Marco Trombini, David Solarna, Gabriele Moser, and Silvana Dellepiane. A goal-driven unsupervised image segmentation method combining graph-based processing and markov random fields. *Pattern Recognition*, 134:109082, 2023.

[8] Brian Patenaude, Stephen M Smith, David N Kennedy, and Mark Jenkinson. A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56(3):907–922, 2011.

[9] Chi-Yun Liu and Jui-Sheng Chou. Bayesian-optimized deep learning model to segment deterioration patterns underneath bridge decks photographed by unmanned aerial vehicle. *Automation in Construction*, 146:104666, 2023.

[10] Long Xie, Laura EM Wisse, Jiancong Wang, Sadhana Ravikumar, Pulkit Khandelwal, Trevor Glenn, Anica Luther, Sydney Lim, David A Wolk, and Paul A Yushkevich. Deep label fusion: A generalizable hybrid multi-atlas and deep convolutional neural network for medical image segmentation. *Medical Image Analysis*, 83:102683, 2023.

[11] José V Manjón and Pierrick Coupé. volbrain: an online mri brain volumetry system. *Frontiers in neuroinformatics*, 10:30, 2016.

[12] Huabing Liu, Dong Nie, Jian Yang, Jinda Wang, and Zhenyu Tang. A new multi-atlas based deep learning segmentation framework with differentiable atlas feature warping. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[13] Quan Zhou, Siman Liu, Chao Jiang, Ye He, Xi-Nian Zuo, Chinese Color Nest Consortium, et al. Charting the human amygdala development across childhood and adolescence: Manual and automatic segmentation. *Developmental Cognitive Neuroscience*, 52:101028, 2021.

[14] Jens C Pruessner, Li Min Li, W Serles, Marita Pruessner, D Louis Collins, Noor Kabani, Sonia Lupien, and Alan C Evans. Volumetry of hippocampus and amygdala with high-resolution mri and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cerebral cortex*, 10(4):433–442, 2000.

[15] Youngwan Lee, Jonghee Kim, Jeffrey Willette, and Sung Ju Hwang. Mpvit: Multi-path vision transformer for dense prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

[17] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.

[18] Wei Chen, Boqiang Liu, Suting Peng, Jiawei Sun, and Xu Qiao. S3d-unet: separable 3d u-net for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4*, pages 358–368. Springer, 2019.

[19] Yun Jiang, Yuan Zhang, Xin Lin, Jinkun Dong, Tongtong Cheng, and Jing Liang. Swinbts: A method for 3d multimodal brain tumor segmentation using swin transformer. *Brain sciences*, 12(6):797, 2022.

[20] Hengxin Liu, Guoqiang Huo, Qiang Li, Xin Guan, and Ming-Lang Tseng. Multiscale lightweight 3d segmentation algorithm with attention mechanism: Brain tumor image segmentation. *Expert Systems with Applications*, 214:119166, 2023.

[21] Jianhong Cheng, Jin Liu, Liangliang Liu, Yi Pan, and Jianxin Wang. Multi-level glioma segmentation using 3d u-net combined attention mechanism with atrous convolution. In *2019 ieee international conference on bioinformatics and biomedicine (BIBM)*, pages 1031–1036. IEEE, 2019.

[22] Liyao Tang, Yibing Zhan, Zhe Chen, Baosheng Yu, and Dacheng Tao. Contrastive boundary learning for point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8489–8499, 2022.

[23] Lei Ke, Martin Danelljan, Xia Li, Yu-Wing Tai, Chi-Keung Tang, and Fisher Yu. Mask transfiner for high-quality instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4412–4421, 2022.

[24] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[25] Yifan Xu, Huapeng Wei, Minxuan Lin, Yingying Deng, Kekai Sheng, Mengdan Zhang, Fan Tang, Weiming Dong, Feiyue Huang, and Changsheng Xu. Transformers in computational visual media: A survey. *Computational Visual Media*, 8:33–62, 2022.

[26] Xue-Ru Fan, Yin-Shan Wang, Da Chang, Ning Yang, Meng-Jie Rong, Zhe Zhang, Ye He, Xiaohui Hou, Quan Zhou, Zhu-Qing Gong, et al. A longitudinal resource for population neuroscience of school-age children and adolescents in china. *Scientific data*, 10(1):545, 2023.

[27] Siman Liu, Yin-Shan Wang, Qing Zhang, Quan Zhou, Li-Zhi Cao, Chao Jiang, Zhe Zhang, Ning Yang, Qi Dong, Xi-Nian Zuo, et al. Chinese color nest project: An accelerated longitudinal brain-mind cohort. *Developmental Cognitive Neuroscience*, 52:101020, 2021.

[28] Ning Yang, Yehong He, Zhe Zhang, Hao-Ming Dong, Lei Zhang, Xing-Ting Zhu, Xiao-Hui Hou, Yinshan Wang, Quan Lai Zhou, Zhu-Qing Gong, Lizhi Cao, Ping Wang, Yiwen Zhang, Danyang Sui, Ting Xu, Gao-Xia Wei, Zhi Yang, Lili Jiang, Huijie Li, Tingyong Feng, Antao Chen, Jiang Qiu, Xu Chen, and Xi-Nian Zuo. Chinese color nest project: Growing up in china. *Chinese Science Bulletin*, 62:3008–3022, 2017.

[29] Jianing Sun, Yanping Jiang, Xiaolei Wang, Samuele Zilioli, Peilian Chi, Lihua Chen, Jiale Xiao, and Danhua Lin. Cortisol reactivity as a mediator of peer victimization on child internalizing and externalizing problems: The role of gender differences. *Research on Child and Adolescent Psychopathology*, 50(3):283–294, 2022.

[30] Qiqi Tong, Hongjian He, Ting Gong, Chen Li, Peipeng Liang, Tianyi Qian, Yi Sun, Qiuping Ding, Kuncheng Li, and Jianhui Zhong. Multicenter dataset of multi-shell diffusion mri in healthy traveling adults with identical settings. *Scientific Data*, 7(1):157, 2020.

[31] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision (3DV)*, pages 565–571. Ieee, 2016.

[32] Haben Berhane, Michael Scott, Mohammed Elbaz, Kelly Jarvis, Patrick McCarthy, James Carr, Chris Malaisrie, Ryan Avery, Alex J Barker, Joshua D Robinson, et al. Fully automated 3d aortic segmentation of 4d flow mri for hemodynamic analysis using deep learning. *Magnetic resonance in medicine*, 84(4):2204–2218, 2020.

[33] Ying Wang, Rob J van Klaveren, Hester J van der Zaag-Loonen, Geertruida H de Bock, Hester A Gietema, Dong Ming Xu, Anne LM Leusveld, Harry J de Koning, Ernst T Scholten, Johny Verschakelen, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. *Radiology*, 248(2):625–631, 2008.

[34] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A benchmark. *IEEE transactions on image processing*, 24(12):5706–5722, 2015.

[35] Jiong Wu and Xiaoying Tang. Brain segmentation based on multi-atlas and diffeomorphism guided 3d fully convolutional network ensembles. *Pattern Recognition*, 115:107904, 2021. https://arxiv.org/pdf/1901.01381.

[36] RW McGugin, A Roche, J Ma, and I Gauthier. Challenges in replication: Does amygdala gray matter volume relate to social network size? *Cognitive Affective Behavior Neuroscience*, 24:707–719, 2024.

[37] R Kanai, B Bahrami, R Roylance, and G Rees. Online social network size is reflected in human brain structure. *Proceedings of The Royal Society B: Biological Sciences*, 279:1327–1334, 2011.

[38] KC Bickart, CI Wright, RJ Dautoff, BC Dickerson, and LF Barrett. Amygdala volume and social network size in humans. *Nature Neuroscience*, 14:163–164, 2011.

[39] B Lu, X Chen, F Xavier Castellanos, PM Thompson, XN Zuo, YF Zang, and CG Yan. The power of many brains: Catalyzing neuropsychiatric discovery through open neuroimaging data and large-scale collaboration. *Science Bulletin*, 69:1536–1555, 2024.

[40] DL Cotter, CE Campbell, K Sukumaran, R McConnell, K Berhane, J Schwartz, DA Hackman, H Ahmadi, JC Chen, and MM Herting. Effects of ambient fine particulates, nitrogen dioxide, and ozone on maturation of functional brain networks across early adolescence. *Environment international*, 117:108001, 2023.

[41] ML Hatzenbuehler, DG Weissman, S McKetta, MR Lattanner, JV Ford, DM Barch, and KA McLaughlin. Smaller hippocampal volume among black and latinx youth living in high-stigma contexts. *Journal of the American Academy of Child and Adolescent Psychiatry*, 61:809–819, 2022.

[42] K Ghanem, Saltoun K, Suvrathan A, Draganski B, and Bzdok D. Longitudinal microstructural changes in 18 amygdala nuclei resonate with cortical circuits and phenomics. *Communications Biology*, 7:477, 2024.

[43] Tomáš Paus. Population neuroscience: why and how. *Human brain mapping*, 31(6):891–903, 2010.

[44] Emily B Falk, Luke W Hyde, Colter Mitchell, Jessica Faul, Richard Gonzalez, Mary M Heitzeg, Daniel P Keating, Kenneth M Langa, Meghan E Martz, Julie Maslowsky, et al. What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences*, 110(44):17615–17622, 2013.

[45] Richard AI Bethlehem, Jakob Seidlitz, Simon R White, Jacob W Vogel, Kevin M Anderson, Chris Adamson, Sophie Adler, George S Alexopoulos, Evdokia Anagnostou, Ariosky Areces-Gonzalez, et al. Brain charts for the human lifespan. *Nature*, 604(7906):525–533, 2022.

[46] ZX Zhou, LZ Chen, MP Milham, XN Zuo, and LBCC Lifespan Brain Chart Consortium. Six cornerstones for translational brain charts. *Science Bulletin*, 68(8):795–799, 2023.