NeuroImage: Clinical 43 (2024) 103623

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Geodesic shape regression based deep learning segmentation for assessing longitudinal hippocampal atrophy in dementia progression

Na Gao [a,1], Hantao Chen [a,1], Xutao Guo [a,b], Xingyu Hao [a], Ting Ma [a,b,c,*]

[a] School of Electronic & Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen, China
[b] Peng Cheng Laboratory, Shenzhen, China
[c] Guangdong Provincial Key Laboratory of Aerospace Communication and Networking Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong, China

ARTICLE INFO

Keywords:
Hippocampus
Longitudinal segmentation
Shape analysis
Morphometrics
MRI

ABSTRACT

Longitudinal hippocampal atrophy is commonly used as progressive marker assisting clinical diagnose of dementia. However, precise quantification of the atrophy is limited by longitudinal segmentation errors resulting from MRI artifacts across multiple independent scans. To accurately segment the hippocampal morphology from longitudinal 3T T1-weighted MR images, we propose a diffeomorphic geodesic guided deep learning method called the GeoLongSeg to mitigate the longitudinal variabilities that unrelated to diseases by enhancing intra-individual morphological consistency. Specifically, we integrate geodesic shape regression, an evolutional model that estimates smooth deformation process of anatomical shapes, into a two-stage segmentation network. We adopt a 3D U-Net in the first-stage network with an enhanced attention mechanism for independent segmentation. Then, a hippocampal shape evolutional trajectory is estimated by geodesic shape regression and fed into the second network to refine the independent segmentation. We verify that GeoLongSeg outperforms other four state-of-the-art segmentation pipelines in longitudinal morphological consistency evaluated by test–retest reliability, variance ratio and atrophy trajectories. When assessing hippocampal atrophy in longitudinal data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), results based on GeoLongSeg exhibit spatial and temporal local atrophy in bilateral hippocampi of dementia patients. These features derived from GeoLongSeg segmentation exhibit the greatest discriminatory capability compared to the outcomes of other methods in distinguishing between patients and normal controls. Overall, GeoLongSeg provides an accurate and efficient segmentation network for extracting hippocampal morphology from longitudinal MR images, which assist precise atrophy measurement of the hippocampus in early stage of dementia.

1. Introduction

Persistent hippocampal atrophy has been identified as a crucial indicator of disease progression in dementia (Dubois et al., 2014; Frisoni et al., 2010; Hill et al., 2014). In-depth investigations into localized hippocampal atrophy have further revealed distinct patterns to specific dementia, such as the Alzheimer's disease (Adler et al., 2018; Braak and Braak, 1997a; Braak and Braak, 1997b; Chauveau et al., 2021; Gabere et al., 2020; Kerchner et al., 2010; Martin et al., 2010; Scheff et al., 2007; Tang et al., 2015; Zhang et al., 2020). Currently, advanced automated quantification methods based on segmentation from 3T T1-weighted MR images allow for in-vivo measurement of morphological changes in the hippocampus. Segmentation methods leverage the relatively high contrast provided by these images, enabling delineation of the hippocampal contour. However, the segmentation of the hippocampal boundary can be adversely affected by various signal variations unrelated to disease progression, such as the head motion, changes in slice orientation, susceptibility artifacts, and differences in scanner hardware and software (Dong MJ et al., 2021). These factors can introduce erroneous quantification of longitudinal morphological changes. For instance, one may detect cortical thickening in regions where cortical thinning actually occurs.

The observed morphology from MRI is a combination of disease-related anatomical changes and independent noise-induced variations.

* Corresponding author at: School of Electronic & Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen, China.
E-mail address: tma@hit.edu.cn (T. Ma).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.nicl.2024.103623
Received 17 October 2023; Received in revised form 12 April 2024; Accepted 25 May 2024
Available online 26 May 2024

The hippocampal morphology affected by dementia remains relatively stable for the same subject, exhibiting smooth and continuous anatomical changes. However, independent image noises introduce variations within the time-series of hippocampal intensity and lead to non-smooth boundaries variations from images. To overcome the issue, recent methodologies have been developed to minimize the influence of this error by incorporation of within-subject references in segmentation of longitudinal images (Brown EM et al., 2020; Das SR et al., 2012; Iglesias JE et al., 2016; Li H et al., 2021; Muralidharan P et al., 2014; Platero C et al., 2019; Shaw T et al., 2020). These methodologies can be broadly classified into three categories. First, the Deformation-Based Morphometry (DBM) methods, such as the ALOHA (Das SR et al., 2012), LASHiS (Shaw T et al., 2020) and Freesurfer longitudinal pipeline (Iglesias JE et al., 2016), utilize diffeomorphic deformations of initial segmentation results to calibrate segmentation form independently observed images. The second is to involve a post-processing technique that employ a shape growth model after the initial segmentation stage. (Muralidharan P et al., 2014) have shown that this approach effectively reduces within-subject variability. These methods have demonstrated significant improvements in estimating longitudinal volume changes and are considered state-of-the-art for longitudinal atrophy estimation. However, the DBM methods face large computational burden, which are further exacerbated by an additional process required for longitudinal correction. Moreover, a significant challenge in both approaches is the limited consideration of boundary information conveyed by image intensity in the final segmentation. The third approach is a deep learning-based method that incorporates previous information to enhance learning ability, such as proposed by Li et al. (2021). This method enables efficient and accurate hippocampal segmentation, with a primary focus on dice accuracy, but lacks demonstration of longitudinal morphological consistency. Above all, in addition to efficiency, it is crucial to strike a balance between the boundary information inferred from image intensity and the trajectory of shape evolution to achieve an accurate description of local atrophy in longitudinal images.

In this paper, we introduce GeoLongSeg, a novel approach that incorporates geodesic shape regression into a deep learning network. The geodesic shape regression, a morphological evolution model capable of estimating smooth and continuous deformations of anatomical structures, is integrated with image intensity to enhance the precision of segmentation. The final segmentation of our method is verified to offer higher accuracy than other advanced methods, and local atrophy pattern based on this segmentation demonstrate the highest accuracy in detecting patients in early stage of dementia.

2. Method

The aim of this study is to propose a hippocampal segmentation method for better description of local atrophy in longitudinal 3T T1-weighted MR images. The overall framework, depicted in Fig. 1, consists of two stages of segmentation. In the first stage, the preprocessed longitudinal images undergo independent segmentation using a 3D U-Net network. Subsequently, the hippocampal labels obtained from this segmentation are utilized to estimate the morphological evolution trajectory through geodesic shape regression. This estimated trajectory then serves as input to guide the fine segmentation in the second stage network. By incorporating shape smoothness correction, the longitudinal morphological variability caused by independent segmentation is effectively eliminated, ensuring a more precise and reliable outcome. The detailed methodology steps are described below.

2.1. Data preparation

The data used in this study are collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database (https://adni.loni.usc.edu). All participants have provided informed written consent before recruitment and filled out questionnaires approved by the respective Institutional Review Board (IRB). In order to ensure the generalizability of our network to both dementia and normal aging, we

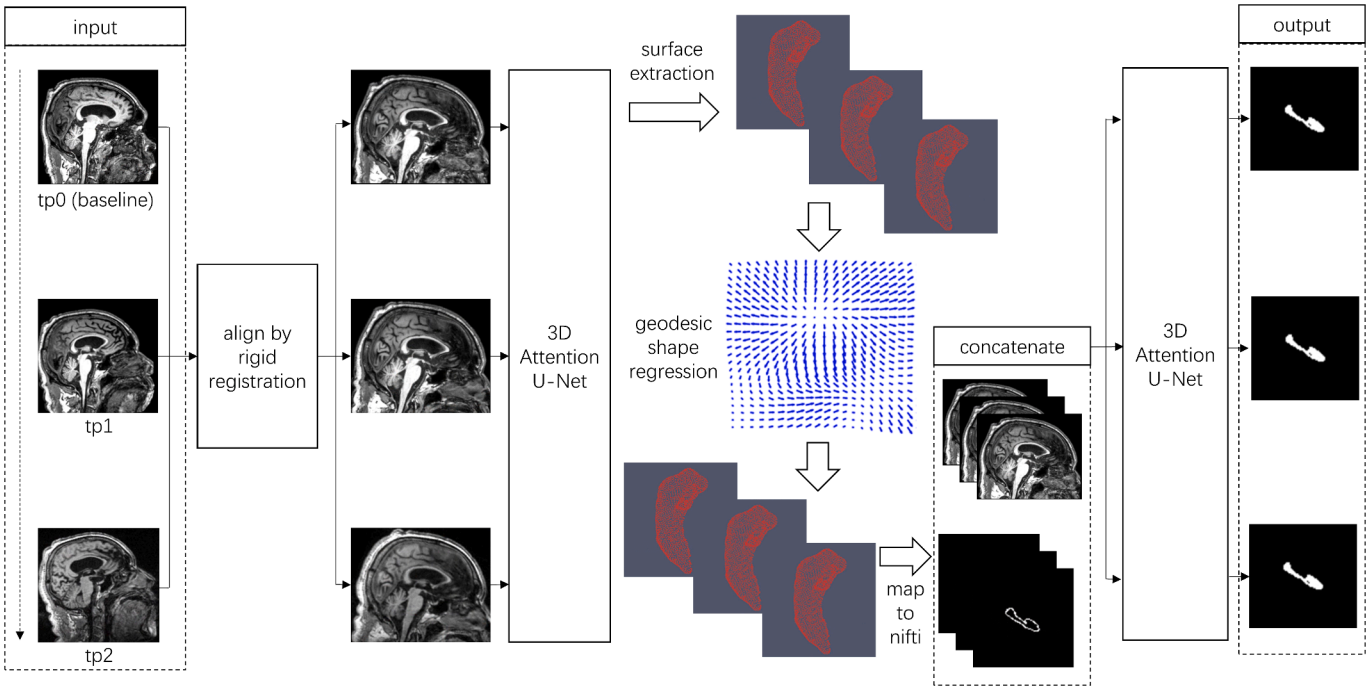

Fig. 1. The framework of GeoLongSeg. The input images are arranged in chronological order and undergo rigid registration for alignment. They are then fed into a first-stage network (3D Attention U-Net) to obtain rough segmentation results. The boundary surfaces are extracted from these results, and a geodesic morphological regression is utilized to estimate the trajectory of morphological changes. The estimated morphological surfaces are mapped back to the original image space and serve as the first channel input for the second-stage network (3D Attention U-Net), while the T1 images serve as the second channel input. The output labels are obtained as the final longitudinal hippocampal segmentations.

enroll individuals in progression of mild cognitive impairment to dementia and age-matched cognitively normal (CN) individuals. Each participant experiences three 3T T1-weighted MRI scans, with an interval of one year. We only select the patients diagnosed as dementia at their third visit. Therefore, the three scans of the patient group constitute a two-year dementia progression.

To construct the annotated longitudinal dataset, we randomly selected 15 participants, including 9 from the dementia progression group and 6 from the CN group. Additionally, 30 participants in dementia progression and 30 CN participants are unlabeled data for evaluating the performance of our method. The demographic information of the dataset is presented in Table 1. All the baseline images are rigidly registered to the MNI152 template using the ANTs toolkit (Avants BB et al., 2014). Subsequently, images of other time points are rigidly registered to their corresponding baseline.

The training set is manually annotated according to the protocol of a high-resolution hippocampal data set from MNI-HiSUB25 (Kulaga-Yoskovitz J et al., 2015), which provides submillimetric T1- and T2-weighted images and stereotaxic probabilistic anatomical maps. The MNI-HiSUB25 segmentation is guided by consistent intensity and morphological characteristics of the densely myelinated molecular layer, along with a few geometry-based boundaries. When manually annotating images of the same subject at different time points, we consider the contour morphology of the baseline hippocampus as a reference to ensure morphological consistency across time. In other words, we consider both image intensity and baseline morphology in the longitudinal annotation. It is also worth noting that this protocol allows for the delineation of the detailed structure of the hippocampal head on 3T images, which is an important region that has been found to be sensitive to dementia.

2.2. Segmentation network

In our study, we introduce a two-stage network approach for longitudinal hippocampal segmentation. The initial stage focuses on independent segmention of individual images, using a U-Net architecture (Fig. 2, stage 1). This choice is motivated by the U-Net's effectiveness on limited training data. We employ the 3D U-Net network to guarantee smooth boundaries of the segmented hippocampi in three-dimensional space, which is crucial for morphological analysis. To enhance the network's sensitivity towards the hippocampal region, we incorporate an attention mechanism into the 3D U-Net architecture, following the approach of Oktay et al. (2018). Attention gate units are introduced within each skip connection's concatenation process, illustrated in the lower panel of Fig. 2.

Then, we introduce a shape regression method to estimate the continuous evolution of shapes based on discrete observations made at different times. The geodesic shape regression utilizes differentiable and invertible deformations to ensure smooth anatomical transformations within the ambient space (Fishbaugh J et al., 2017; Fishbaugh J et al., 2013). Starting with an initial shape S_0 at time t_0, the shape gradually deforms to match target shapes S_i observed at later times. The estimation process is formulated as a variational problem, balancing fidelity to the observed data with regularization, as outlined by a specific regression criterion.

$$E(X_0, \varphi_t) = \sum_i D\left(\varphi_{t_i}(X_0) - O_{t_i}\right) + Reg(\varphi_t) \quad (1)$$

where D represents a distance metric that quantifies the dissimilarity between shapes. $Reg(\varphi_t)$ measures the regularity of the geodesic flow of diffeomorphisms φ_t. By applying the continuous geodesic flow of diffeomorphisms φ_t to the estimated anatomical configuration, we generate a temporally consistent sequence of shapes. To reduce the computational cost, we decrease the number of deformation field control points by assigning them to the vertices of the hippocampal boundary surface. After convergence, the deformed surfaces are mapped back to the original voxel space and serve as input for fine segmentation in the second-stage network.

The objective of the second-stage network is to refine the independently segmented results by leveraging the trajectory estimated through shape regression as a reference for fine segmentation. This strategy ensures that longitudinal segmentation considers both image intensity and temporal consistency. The architecture of the second-stage network also employs a 3D U-Net configuration with two input channels (Fig. 2, stage 2). T1-weighted MRI structural images are fed into the first channel, while estimated labels are provided in the second channel. Finally, we perform a crop around the hippocampal region to mitigate the occurrence of false-positive segmentations outside the hippocampal area. The center of the hippocampal region, determined from the label generated by the previous stage network, is used as the crop center.

2.3. Training strategy

Our method is implemented using the PyTorch 1.6.0 framework and trained on a GPU (NVIDIA TITAN RTX 24G). The training process utilizes the Adam optimizer. The loss function employs cross-entropy loss, and for a single sample:

$$L = -[y log\hat{y} + (1-y) log(1-\hat{y})] \quad (2)$$

where L represents the cross-entropy loss, y denotes the true labels of the samples, and $\hat{y}$ represents the predicted probability of being a positive class for a given input ($\hat{y} = P(y = 1|x)$. The initial learning rate is set to 0.001, and a weight decay of 10^{-8} is applied to prevent overfitting. The training is conducted for 300 epochs, with a batch size of 3 and a patch size of 144 × 144 × 144.

2.4. Evaluation methods

The anatomic structures should undergo smooth temporal changes in the context of disease progression, and it is important for longitudinal segmentation methods to mitigate any unsmooth errors (i.e., longitudinal errors) introduced by imaging noise. Therefore, metrics for evaluating both the individual segmentation accuracy and the within-subject longitudinal consistency are necessary in the evaluation of the longitudinal segmentation method. We comprehensively quantify the segmentation accuracy of our method using various metrics that align with those reported in literatures, including the Dice coefficient, test–retest reliability, and variance ratio. Additionally, we evaluate the efficacy of our method in assessing hippocampal atrophy by detecting significant local atrophy in dementia patients at early stage, and test the classification accuracy of using these atrophic features. The details are outlined as follows.

2.4.1. Metrics for evaluation of segmentation accuracy

We use the Dice coefficient to evaluate the overall segmentation accuracy, calculated by

Table 1
Demographic information for the baseline data enrolled in this study.

	Groups	Number of Subjects	Gender (Male/Female)	Age
Training set	Dementia	9	4/5	75.2 ± 8.3
	CN	6	4/2	72.7 ± 5.4
Testing set	Dementia	30	19/11	75.8 ± 6.8
	CN	30	20/10	75.7 ± 7.1

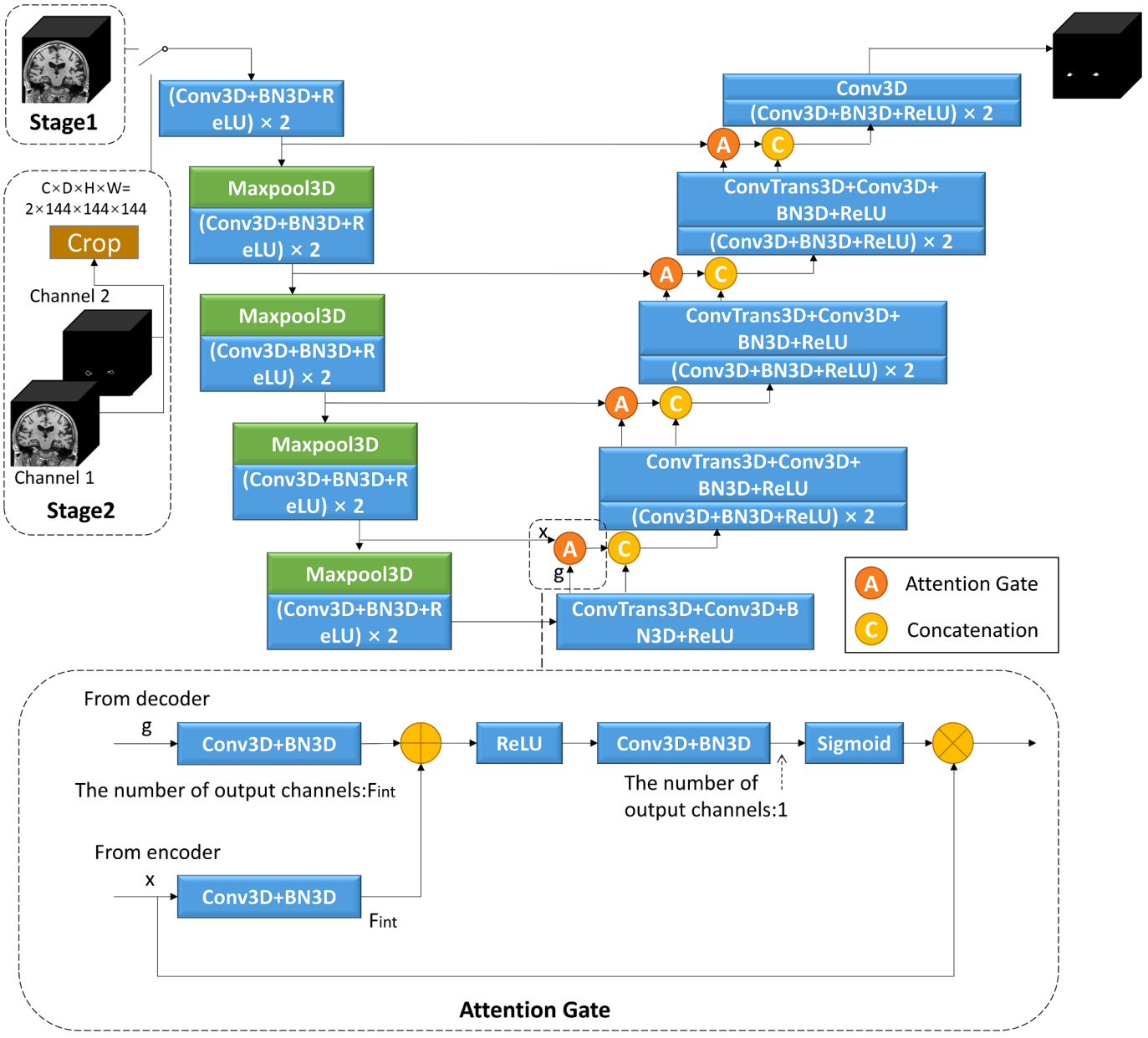

Fig. 2. The network architecture is the 3D Attention U-Net. The left side represents the encoder part that extracts features, while the right side represents the decoder part which restores the size and generates segmentation results. The network employs attention gates, which producing a feature map that highlights important regions, and the output of these gates is connected to the upsampled output of the decoder part through concatenation. In the first stage, the network only inputs T1 images, while in the second stage, the number of input channels in the network doubled, and the contour information of the hippocampus after geodesic regression was concatenated. Cropping operations were applied to preserve only the surrounding area images of the hippocampus.

$$Dice = 2\frac{|X \cap G|}{|X| + |G|} \times 100\% \tag{3}$$

where X represents the segmented label and G represents the ground truth label.

In addition, we use the evaluation strategies in line with previously published longitudinal segmentation methods (Shaw T et al., 2020; Tustison N J et al., 2020), including the test–retest reliability (also known as the overlap coefficient) and variance ratio, which reflect the longitudinal morphological consistency of the segmentation. The test–retest reliability is calculated by

$$test - retestreliability = 2\frac{|A \cap B|}{|A| + |B|} \times 100\% \tag{4}$$

where A and B represent the volume of labels at two different time points.

The variance ratio is computed using a linear mixed-effects (LME) model. Intuitively, a good longitudinal segmentation exhibits larger between-subject variability and smaller within-subject variability. For each morphological measurement, a linear mixed-effects model is fitted by:

$$V_{st} = kt + h_s + \varepsilon \tag{5}$$

where V_{st} represents the hippocampal volume, k represents the slope of the regression line, t represents the age, h_s represents the bias introduced by each subject, and ε represents the residual. The variability between subjects is quantified by the variance of the individual biases τ. The variance of the residuals for all subjects is quantified by the residual

variability σ. Then, the temporal consistency of each morphological feature can be quantified by dividing the between-subject variability by the residual variability:

$$r = \frac{\tau}{\sigma} \tag{6}$$

A larger variance ratio indicates better consistency within individual.

2.4.2. Evaluation of atrophy from the segmentation

To assess longitudinal atrophy, we extract local morphological features from segmentation utilizing the multiscale skeletal representation (m-s-rep) method proposed by Gao, N., et al. (2023). The method defines local thickness measurement on the hippocampus and exhibits superior performance in detecting hippocampal atrophy in the early stage of dementia. In this paper, we measure the atrophy from both spatial and temporal dimensions. The spatial atrophy is quantified by the relative atrophy in the patient group compared to the control group. The temporal atrophy is quantified by the annualized decreasing rate of local thickness in individual hippocampi, calculated by the slope of linear regression of local thickness versus scan date. The statistical significance of intergroup differences is tested using general linear models. We set each measurement as the dependent variable, with group membership as the factor of interest and age as a covariate. Sex is included as an additional covariate for cross-sectional volume measurements. FDR correction for multiple comparisons is performed (Benjamini and Yekutieli, 2001). Cohen's d is used to measure the effect size (d), which can be considered as small (0.2), medium (0.5), or large (0.8).

To validate the effectiveness of features from GeoLongSeg segmentation in assisting dementia detection in early stage, we applied a random forest model to classify patients with pMCI and healthy controls. We used a 5-fold cross-validation to avoid overfitting caused by improper data set partitioning. The classification performance was evaluated by sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve (AUROC) on the testing set.

3. Results

We have conducted experiments to evaluate the accuracy of our proposed method in segmentation of longitudinal hippocampi, while also assessing the effectiveness of the method in evaluating hippocampal local atrophy during the dementia progression. The accuracy of longitudinal segmentation can be evaluated from two dimensions. Firstly, the accuracy of the overall morphology of the segmentations is quantitatively assessed by comparing them with annotated ground truth labels, using the Dice coefficient as a metric. Secondly, the test–retest reliability and variance ratio are employed to quantitatively measure the consistency of longitudinal morphology. In addition, significant hippocampal atrophic features are computed in dementia converters based on the segmentations obtained from different methods. These features are then utilized to classify patients and controls.

In our experiments, we select the four other advanced segmentation pipelines for comparison with our method: Synthseg (Billot B et al., 2023), BrainLabel (Wei C et al., 2022), Freesurfer's independent segmentation pipeline (FSCross) (Fischl B et al., 2002), and Freesurfer's longitudinal segmentation pipeline (FSLong) (Reuter M et al., 2012). The detailed results are presented below. The Synthseg, BrainLabel and FSCross are methods that employ independent segmentation on longitudinal images, while the FSLong is a longitudinal segmentation mehod.

3.1. Global segmentation accuracy

To demonstrate the global segmentation accuracy of the proposed network, we conduct ablation experiments to assess the Dice coefficients of the single-stage and two-stage networks. In the case of the two-stage networks without integrated geodesic shape regression (stage2_ngr), we extract hippocampal surfaces from the initial segmentation (stage1) and utilize them as input to the secondary channel for the second-stage network. Following a five-fold cross-validation, the average Dice similarity coefficients for each network segmentation outcome are computed, shown in Table 2.

Notably, the single-stage network (stage1) exhibits the highest average Dice coefficient. The two-stage network employing geodesic shape regression (stage2_gr) achieves a higher Dice coefficient than the two-stage network without geodesic shape regression (stage2_ngr). The observed decrease in the Dice coefficient for stage2_ngr compared to stage1 is attributed to spatial deviations introduced during the conversion of the hippocampal boundary to voxel contours. The average Dice coefficient of stage1 is slightly higher, within 0.003, than both stage2_ngr and stage2_gr. Despite this superiority in terms of the Dice coefficient, the evaluation of longitudinal morphological consistency is also necessary for a comprehensive evaluation.

We further calculate Dice coefficients of the segmentations across time points, shown in Table 2. Among the results of the stage1, the Dice coefficient is the lowest at the second time point (tp1) and the largest at the third time point (tp2). The results of stage2_ngr exhibited a similar trend, with the lowest Dice coefficient observed at tp1 and the highest at tp2. However, with geodesic shape regression applied prior to the second stage network, the Dice coefficient at the first two time points is enhanced, although the Dice coefficient at tp3 is slightly reduced. This observed pattern appears to align with the trend exhibited by the results of stage1_gr. Across all three time points, result of stage1_gr at tp1 has the highest Dice coefficient, which possibly enhances the final output of the two-stage network at tp1.

3.2. Longitudinal morphological consistency

The longitudinal morphological consistency is assessed from two aspects. The global longitudinal consistency quantifies the smoothness of overall morphologic changes over time, and the local longitudinal consistency characterizes the smoothness of temporal changes at specific spots. Additionally, to ensure the generalizability of our method, we conducted the experiment in both the dementia and CN groups.

3.2.1. Global longitudinal consistency

The global longitudinal consistency is evaluated by the test–retest reliability and global variance ratio. The distributions of the test–retest reliability for the five methods in each group are shown in Fig. 3. We observe that the average test–retest reliability improved after incorporating geodesic regression. In addition, the distribution of test–retest

Table 2
Average Dice Coefficients for comparison of intermediate and final results in our proposed segmentation network.

Pipeline	stage1	stage1_gr	stage2_ngr	stage2_gr (GeoLongSeg)
Dice coefficient (all time points)	0.8758 ± 0.0113	0.7836 ± 0.0580	0.8734 ± 0.0106	0.8747 ± 0.0108
Dice coefficient (tp 0)	0.8745 ± 0.0147	0.7794 ± 0.0567	0.8773 ± 0.0106	0.8751 ± 0.0115
Dice coefficient (tp 1)	0.8690 ± 0.0038	0.7911 ± 0.0633	0.8657 ± 0.0041	0.8712 ± 0.0136
Dice coefficient (tp 2)	0.8839 ± 0.0060	0.7850 ± 0.0524	0.8812 ± 0.0040	0.8778 ± 0.0029

Abbreviations: stage1, segmentation results from the first stage network; stage1_gr, segmentation results of directly performing geodesic shape regression on the segmentation labels of the first stage; stage2_ngr, segmentation results from the second stage network without geodesic shape regression guidance; stage2_gr, segmentation results from the second stage network with geodesic shape regression guidance; tp, time point.

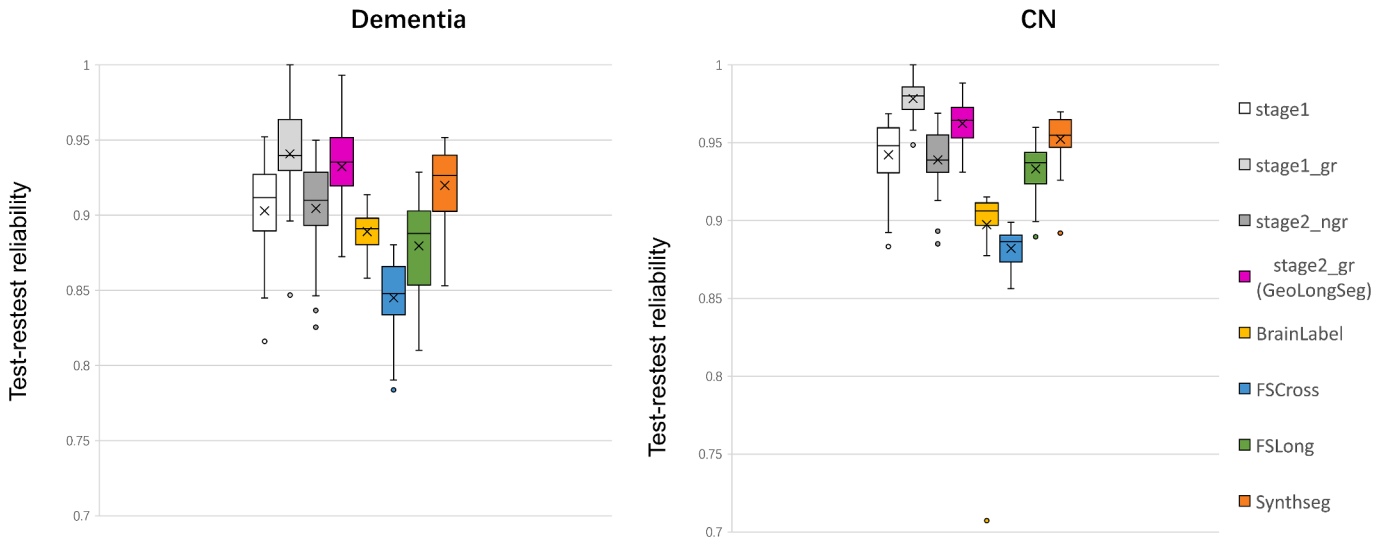

Fig. 3. Global longitudinal morphological consistency measured by the test-Retest reliability of the hippocampal segmentations from different methods in dementia and CN groups respectively. The results obtained from different segmentation methods are represented using different colors. The cross symbol denotes the average value. Abbreviations: stage1, segmentation results from the first stage network; stage1_gr, segmentation results of directly performing geodesic shape regression on the segmentation labels of the first stage; stage2_ngr, segmentation results from the second stage network without geodesic shape regression guidance; stage2_gr, segmentation results from the second stage network with geodesic shape regression guidance; FSCross, the Freesurfer independent segmentation method; FSLong, the segmentation method using the Freesurfer longitudinal pipeline.

reliability for the final output of the second-stage network (stage2_gr) is more concentrated than that of the stage1 and stage2_ngr in the two groups. This observation suggests that the final network output exhibits a higher degree of longitudinal consistency and stability.

It can be observed that our method achieves the highest average test–retest reliability among the other methods evaluated. It is worth noting that the Freesurfer cross-sectional method (FSCross) exhibits the lowest test–retest reliability, while the Freesurfer longitudinal method (FSLong) demonstrates an improvement compared to the FSCross. In the dementia group, our method shows a comparable compactness of distribution to the FSCross, FSLong, and Synthseg. Althoug results of the Brainlabel exhibits the most compact distribution, it has relatively low average test–retest reliability. In the CN group, our method achieves a comparable level of concentration in the distribution of test–retest reliability compared to other methods.

We also observe that the test–retest reliability of the dementia group has lower average and larger variance than that of the CN group. This is probably because that the dementia group exhibits more pronounced and different extent of hippocampal atrophy compared to the CN group, leading to lower degree of overlapping between labels of two adjacent years in the dementia group.

The volumetric variance ratio serves as another metric to assess the overall longitudinal consistency. A higher value indicates better longitudinal consistency, whereas a lower value signifies poorer consistency. The results of our tests for different segmentation methods are presented in Table 3. Notably, the volumetric variance ratio of the two-stage network incorporating geodesic shape regression (stage2_gr) demonstrates an improvement compared to that of stage1 and stage2_ngr, which did not employ geodesic shape regression. Furthermore, the output from stage1_gr, which incorporates geodesic regression after the first stage, exhibits the highest longitudinal consistency. This observation is within expectations, as geodesic regression estimates the segmented shapes by solely considering morphological smoothness. However, the segmentation requires consideration of both the shape variation smoothness and the image intensity. As can be observed from the Table 2 that the Stage1_gr has poor segmentation accuracy, although it improves the longitudinal consistency.

In both the dementia and CN groups, our method (stage2_gr) demonstrates the highest volumetric variance ratio when compared to the other four tested methods. In the CN group, Synthseg exhibits a marginally lower volumetric variance ratio compared to stage2_gr, whereas in the dementia cohort, the difference is large. FSLong shows a greater volumetric variance ratio than FSCross, indicating better longitudinal consistency. BrainLabel exhibits a much larger volumetric variance ratio in the dementia group than in the CN group. In contrast, our method exhibits a more consistent volumetric variance ratio across both groups.

3.2.2. *Local longitudinal consistency*

To evaluate local longitudinal morphological consistency of the segmentations from different methods, we calculate local thickness variance ratios at 1738 distinct locations on the hippocampal surface. The results are presented in Fig. 4, with the horizontal axis representing the 1–1738 locations on the hippocampal surface and the vertical axis representing values of the variance ratios. The different methods are differentiated by distinct colors and shapes.

Table 3
Volumetric variance ratio of different segmentation methods.

Pipelines	stage1	stage1_gr	stage2_ngr	stage2_gr (GeoLongSeg)	BrainLabel	FSCross	FSLong	Synthseg
Dementia group	3.9584	7.4772	3.8459	8.1508	6.7269	3.1376	4.1478	7.7925
CN group	8.8374	13.6687	8.5113	9.4329	1.5611	5.8979	7.8718	9.417

Abbreviations: stage1, segmentation results from the first stage network; stage1_gr, segmentation results of directly performing geodesic shape regression on the segmentation labels of the first stage; stage2_ngr, segmentation results from the second stage network without geodesic shape regression guidance; stage2_gr, segmentation results from the second stage network with geodesic shape regression guidance; FSCross, the Freesurfer independent segmentation method; FSLong, the segmentation method using the Freesurfer longitudinal pipeline.

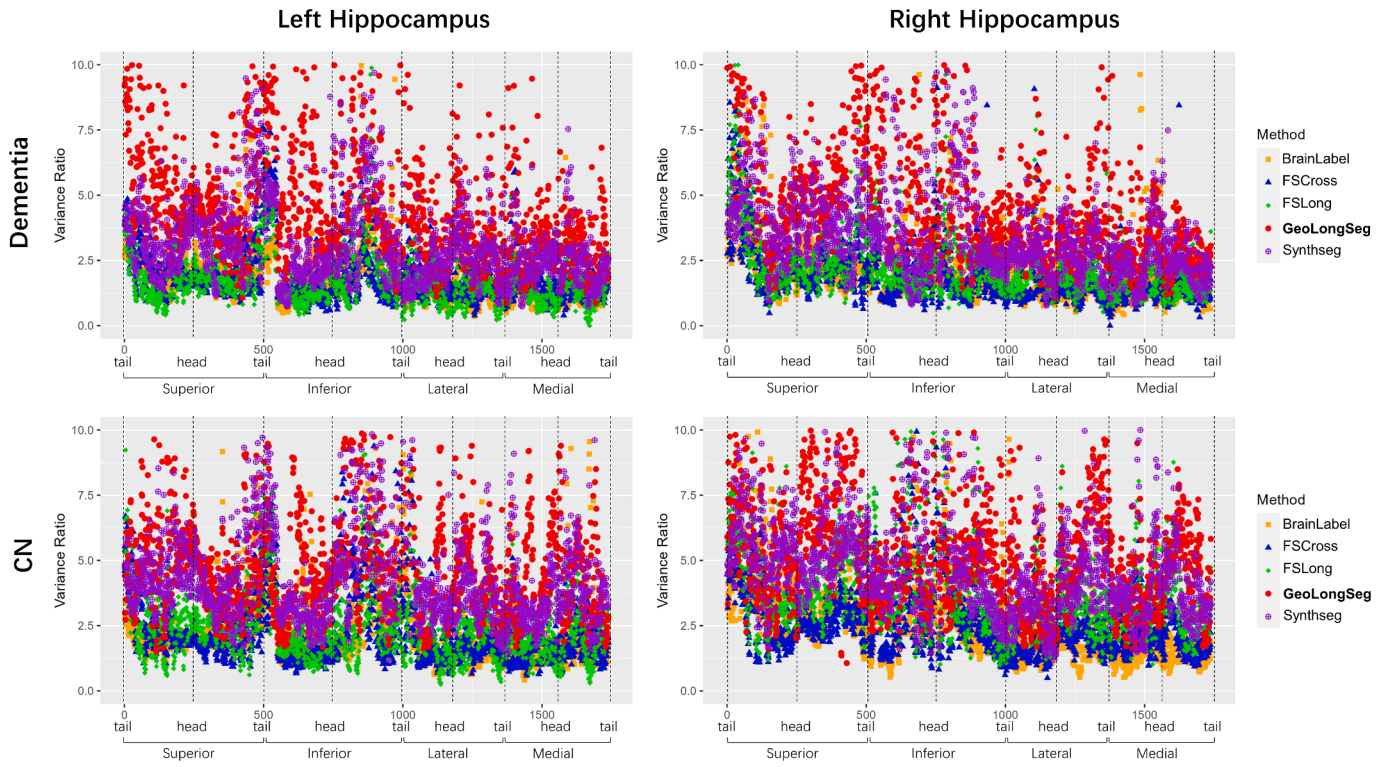

Fig. 4. Local longitudinal morphological consistency measured by the thickness variance ratio of segmentations from different methods in dementia and CN groups respectively. The x-axis represents different locations on the hippocampus. The results obtained from different segmentation methods are depicted using different colors and shapes.

We observed that the variance ratio of the segmentations can vary between the left and right sides, as well as across different groups. For instance, in the tail of the hippocampus, approximately at the horizontal axis scale of 500, the FSCross, FSLong, Synthseg, and BrainLabel methods exhibit higher variance ratios in the left hippocampus of the dementia group, whereas lower variance ratios are observed in the right hippocampus. A similar trend can be observed in the CN group. Also, in this region, the distribution of variance ratios of Synthseg in CN group is higher than those observed in the dementia group. However, overall, our proposed method shows the highest variance ratios across most of local locations compared to the other methods, and the FSLong shows higher variance ratios than the FSCross.

We further calculate the means and standard deviations of the variance ratio, which are summarized in Table 4. We observe that GeoLongSeg exhibits the highest average variance ratio among all methods, but also has a larger variance. Except for FSCross in the dementia group, in all methods, the CN group has a larger average variance ratio than the dementia group, indicating higher longitudinal consistency in overall morphology, consistent with the conclusion in Section 3.2.1.

Table 4
Local thickness variance ratio of different segmentation methods.

		Dementia group		CN group	
		Left	**Right**	**Left**	**Right**
variance ratio	**GeoLongSeg**	**4.18 ± 1.94**	**4.13 ± 1.91**	**4.49 ± 1.74**	**5.00 ± 1.87**
	Synthseg	3.00 ± 1.47	3.31 ± 1.51	4.13 ± 1.58	4.54 ± 1.65
	BrainLabel	1.79 ± 0.89	2.02 ± 1.16	2.23 ± 1.33	2.60 ± 1.45
	FSCross	1.84 ± 1.06	1.82 ± 1.07	2.17 ± 1.28	2.54 ± 1.24
	FSLong	1.68 ± 0.98	2.19 ± 1.20	2.38 ± 1.25	3.59 ± 1.54

We evaluate the local morphological accuracy of different segmentation methods by temporal changes of local thickness. Specifically, we randomly select a subject and test whether the bilateral hippocampi obtained using different segmentation methods exhibits consistent atrophy at specific locations with the ground truth. The results are depicted in Fig. 5, with warm colors indicating regions of decreased thickness and cool colors representing regions of increased thickness over time.

Visual evaluation observes that both the shapes of segmentations from GeoLongSeg and BrainLabel are closer to the ground truth hippocampi. As expected, the segmentation from GeoLongSeg exhibits better similarity in the hippocampal head, whereas the segmentation from BrainLabel is more similar to the ground truth in the tail. The segmentation from Synthseg displays some missing parts at the top of the hippocampal tail and some redundant parts in the head. Conversely, the two Freesurfer segmentation methods introduce large sawtooth-shaped protrusions along the hippocampal lateral boundary, particularly evident in the segmentation using FSCross in the right hippocampus. The line charts in Fig. 5 show the longitudinal local thickness variations. It can be observed that our method exhibits the most consistent segmentation of local atrophic regions that align with the ground truth. In contrast, segmentations from other pipelines exhibit thickness increases in some atrophic regions, which is inconsistent with the ground truth.

3.3. Detecting patients in the early stage of dementia using local atrophic features of the hippocampus

To confirm whether the improvement in segmentation by GeoLongSeg is more advantageous for discriminating dementia in early stage, we conduct a classification using significant spatiotemporal atrophic features between patients and normal controls. We only identify features that exhibited significant atrophy ($p < 0.05$) consistently across all time points, which we considered as significant spatial features

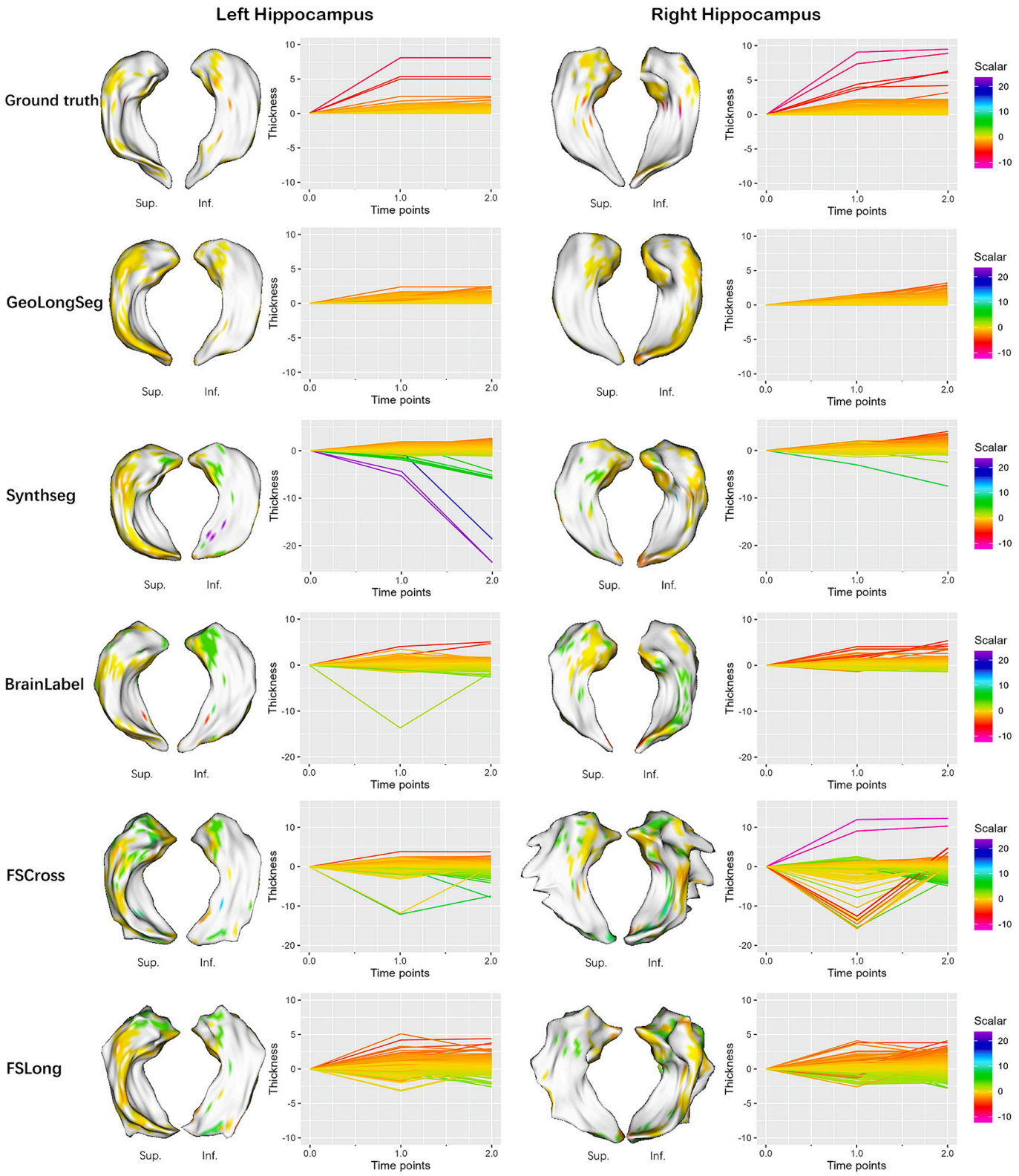

Fig. 5. Local atrophy trajectories of different segmentation methods on a typical case. The stained regions on the hippocampus indicate deformations compared to the baseline morphology, with colors representing the degree of atrophy and expansion. Warm colors represent atrophy (<0), while cool colors represent expansion (>0). The line graph illustrates the deformation trajectories at different locations, with colors indicating the degree of atrophy and expansion.

distinguishing dementia in two years preceding the dementia conversion. Additionally, we calculate average atrophy rates of each morphological features over the two-year period using linear regression as temporal atrophy features. The statistical results are shown in the left columns of Fig. 6, with red indicating the Cohen's d effect size.

It can be observed that the local atrophy patterns vary across different segmentation methods. Specifically, analysis based on Synthseg identifies a greater number of atrophic locations, whereas analysis

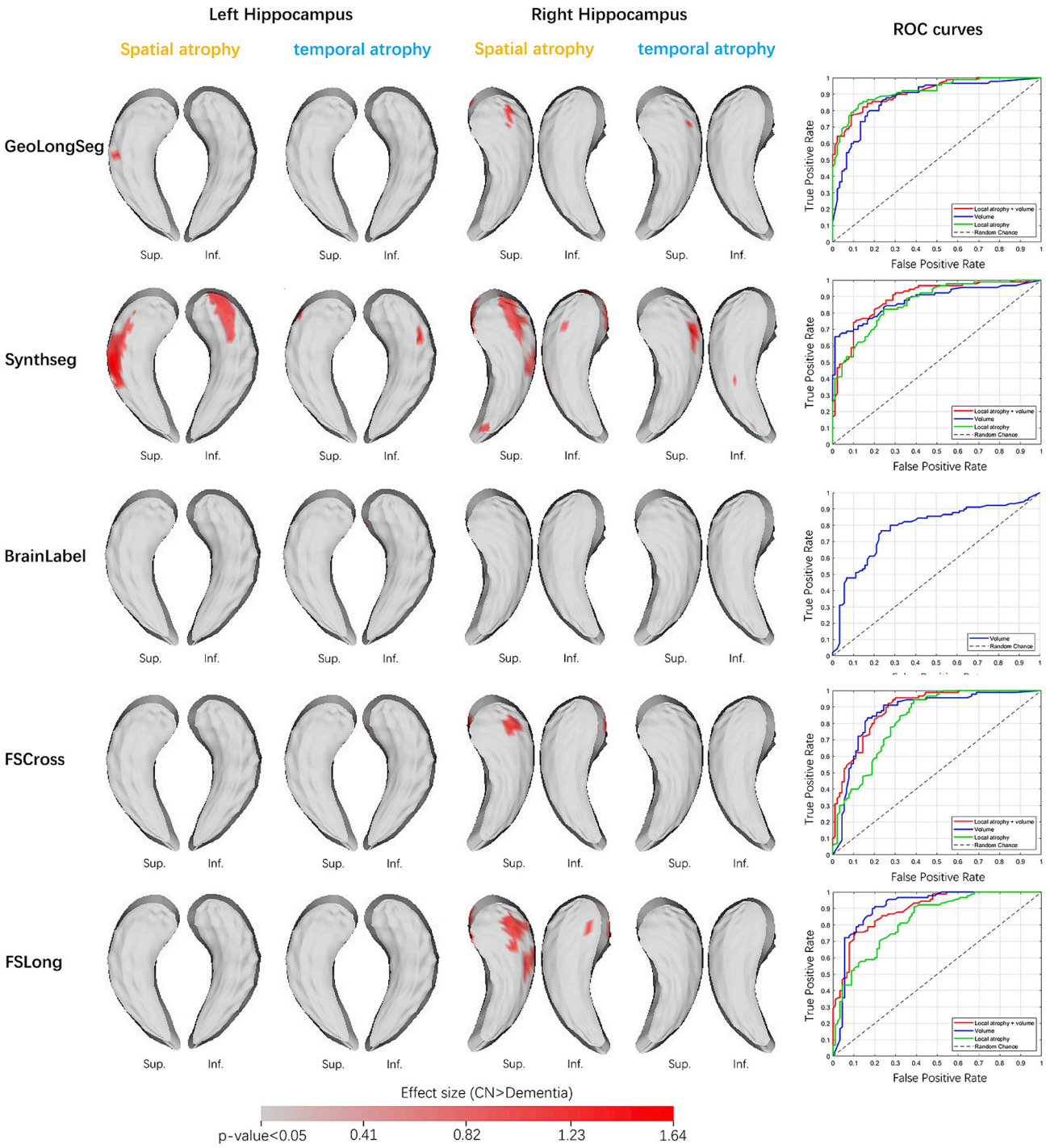

Fig. 6. Using different segmentation methods for local atrophy assessment in two years prior to dementia conversion, and classification based on significant local atrophic and volumetric features. The red regions on the hippocampus represent areas of significant atrophy in dementia compared to CN, observed consistently across three consecutive scans (spatial atrophy differences), as well as regions where dementia exhibits significantly greater hippocampal atrophy rate compared to CN (temporal atrophy differences). The color intensity indicates the degree of significance, measured by the Cohen's d effect size. In the last column, red ROC curve represents the classification performance of both volumetric features and significant local atrophic features for the two groups of subjects. The blue curve represents the classification results using only volumetric features, while the green curve represents the classification results using only significant local atrophic features. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

based on BrainLable does not detect any significant atrophy. The results of FSCross and FSLong are consistent, both finding spatial atrophy in the head of the right hippocampus, as can be seen in the last two rows of Fig. 6. This indicates that the local thickness in this region of the hippocampus is significantly less in dementia patients than in the control group. Our method also found significant temporal atrophy at almost the same position. This is similar to the atrophy pattern indicated by the results from Synthseg. Results from GeoLongSeg and Synthseg also reveal spatial atrophy in the superior lateral part of the left hippocampus.

Note that the volumetric features incorporate both spatial and temporal characteristics, namely, absolute volume and annualized change

rate. In the right column of Fig. 6, we present ROC curves for classifying dementia and CN individuals based on segmentations from different methods and different combinations of features (local atrophy + volume, volume, and local atrophy features). The accuracy, sensitivity, specificity, and AUROC are listed in Table 5. Since no significant atrophy feature is found in the segmentation results based on BrainLabel, only volume features are tested to distinguish between the two groups.

As observed from the figure, the morphological features calculated from different segmentations have different performance in classifying the dementia and CN groups. Specifically, the volumetric features derived from FSLong segmentation demonstrate a superior ability to distinguish dementia compared to other methods. And the volumetric features outperforming both the local features and the combinations of the two. This observation suggests that, in the context of dementia classification, local features may serve as a limiting factor. Nevertheless, our methodology demonstrates that local atrophy features also contribute significantly to the classification. The results based on FSCross indicate that the inclusion of local features has little impact classification accuracy, but using only local features leads to low accuracy.

In terms of the volumetric features, the Freesurfer Longitudinal pipeline (FSLong) exhibits the highest AUROC, while having comparable accuracy to that of the Freesurfer cross-sectional pipeline (FSCross). Our method demonstrates comparable performance to the Synthseg. When using only local features for classification, our approach demonstrates superiority across all metrics, followed by SynthSeg, with FSCross and FSLong performing comparably. The classification that incorporates both local atrophic features and volumetric features demonstrates superior performance of the GeoLongSeg segmentation in terms ofaccuracy, sensitivity and AUROC compared to other methods.

Table 5
Results of inter-group classification using local atrophic features and volume.

Volumetric measurements				
	Accuracy	**Sensitivity**	**Specificity**	**AUROC**
GeoLongSeg	0.789 ± 0.010	0.778 ± 0.019	0.800 ± 0.00	0.887 ± 0.011
Synthseg	0.783 ± 0.000	0.789 ± 0.019	0.778 ± 0.019	0.890 ± 0.038
BrainLabel	0.761 ± 0.019	0.7755 ± 0.019	0.767 ± 0.033	0.804 ± 0.067
FSCross	**0.833 ± 0.017**	0.833 ± 0.067	**0.833 ± 0.058**	0.880 ± 0.020
FSLong	**0.833 ± 0.017**	**0.856 ± 0.051**	0.811 ± 0.039	**0.929 ± 0.017**
Local atrophy measurements				
	Accuracy	**Sensitivity**	**Specificity**	**AUROC**
GeoLongSeg	**0.861 ± 0.010**	**0.878 ± 0.019**	**0.844 ± 0.019**	**0.927 ± 0.029**
Synthseg	0.789 ± 0.026	0.756 ± 0.019	0.822 ± 0.051	0.877 ± 0.038
BrainLabel	–	–	–	–
FSCross	0.744 ± 0.026	0.711 ± 0.039	0.778 ± 0.019	0.840 ± 0.042
FSLong	0.733 ± 0.050	0.733 ± 0.067	0.7733 ± 0.033	0.846 ± 0.022
Local atrophy and volumetric measurements				
	Accuracy	**Sensitivity**	**Specificity**	**AUROC**
GeoLongSeg	**0.833 ± 0.017**	**0.844 ± 0.019**	0.822 ± 0.051	**0.911 ± 0.028**
Synthseg	0.794 ± 0.026	0.800 ± 0.058	0.789 ± 0.019	0.898 ± 0.031
BrainLabel	–	–	–	–
FSCross	0.811 ± 0.051	0.778 ± 0.039	**0.844 ± 0.077**	0.891 ± 0.023
FSLong	0.794 ± 0.035	0.833 ± 0.033	0.756 ± 0.039	0.901 ± 0.029

4. Discussion

This paper introduces a novel hippocampal segmentation method based on 3T T1-weighted MRI for assessing morphological atrophy in the progression of dementia. The proposed method, GeoLongSeg, integrates geodesic shape regression into a 3D U-Net network, enhancing the accuracy of longitudinal image segmentation by improving intra-individual morphological consistency. To validate our method, we evaluate the Dice coefficient and longitudinal morphological consistency of the segmentation results. The results demonstrate the superiority of GeoLongSeg over independent segmentation by a one-stage 3D U-Net network and four existing state-of-the-art segmentation methods. Additionally, we conduct a comprehensive assessment of the hippocampal atrophy in dementia patients relative to normal controls using different segmentation methods. The results show that the features derived from GeoLongSeg segmentation achieve higher segmentation accuracy compared to other methods.

4.1. GeoLongSeg improves atrophy measurement by enhancing longitudinal local segmentation of the hippocampus

There are two dimensions of atrophy in the progression of dementia. At the individual level, hippocampal atrophy manifests as a gradual process over time, known as temporal atrophy, and is typically quantified through annual atrophy rates. At the group level, comparisons reveal significant volumetric or thickness reductions in the hippocampus, known as spatial atrophy. Increasing evidence suggests that subfields of the hippocampus are differentially affected by the progression of dementia, and some regions experience local atrophy at very early stage of the disease (Chauveau et al., 2021; Braak and Braak, 1997a; Dubois et al., 2014; Jack CR et al., 2010; Sperling RA et al., 2011; Weiner MW et al., 2015). Future research endeavors may potentially expand to establish a more comprehensive connection between these evidences and clinical imaging discoveries, to aid the early diagnosis of dementia. Against this background, accurate segmentation of the hippocampal morphology from longitudinal 3T T1-wighted MRI data is essential for capturing the continuous and evolving local atrophy patterns caused by disease. However, current hippocampal morphological studies in longitudinal MRI often use independent segmentations at each scan, by considering each segmentation as a discrete event. This may introduce intra-subject variability induced by MRI signal variations that unrelated to diseases (Dong MJ et al., 2021). This error in segmentation can result in potential over- or underestimation of atrophy and affect estimation of local atrophy trajectories.

One approach to address the problem is to incorporate intra-individual morphological consistency in segmentation, such as the method used in Freesurfer longitudinal pipeline. The fundamental assumption underlying the method is that deforming a template tailored specifically to an individual subject to match each of the subject's time point scans, more accurate results can be achieved compared to independently deforming a generic template to each scan of the same subject. This approach actually enhances intra-individual consistency by reducing overall deformation errors. As expected, in our experiments, we observe that the Freesurfer longitudinal pipeline exhibits better longitudinal morphological consistency compared to the Freesurfer cross-sectional pipeline. However, the estimation of the segmentation solely depends on image intensity. Despite that the methods emphasized the errors inherent in MRI imaging when explaining the genesis of longitudinal errors, they fail to address the issue directly. Different from the above method, we hypothesize that the morphological variations observed in the image arise from a combination of deformations due to disease progression or aging, along with imaging noise. To mitigate the influence of image noise, we aim to achieve a balance between the image intensity and smooth deformation estimations in our method, thereby

obtaining an optimal segmentation.

Consistent with the metrics employed by literatures of other methods, we assess the Dice coefficient, test–retest reliability, and variance ratio of the segmentations from different approaches. We find that our method, using shape regression integrated two-stage network, achieves a better dice coefficient compared to which utilizes the shape regression as a post-processing technique for segmentation. Based on the Dice coefficients obtained from the segmentation outputs at each time point, the geodesic shape regression serves as a crucial intermediate step that balances the segmentation results across different observation points. Furthermore, our method demonstrates the best overall longitudinal morphological consistency when compared to three advanced independent segmentation methods and one longitudinal segmentation approach. It is worth noting that the test–retest reliability and volumetric variance ratios of all the methods, except the BrainLabel, in CN group performs better than those in the dementia group, indicating inferior longitudinal consistency in dementia hippocampi. This suggests more complex deformation patterns occur in the hippocampus during disease progression than in normal ageing.

Different from other literature, we conduct a comprehensive evaluation of our method in terms of local morphology. This includes calculating the variance ratio of the local thickness of the hippocampal tail and assessing the atrophy progression of the hippocampus for an individual subject. Among them, the observation of variance ratios around the hippocampal tail indicates that the accuracy of local morphology segmentation is not uniform bilaterally. Despite this, our proposed method demonstrates the highest variance ratio in most locations compared to other methods. We further demonstrate enhanced accuracy in detecting longitudinal atrophy in a randomly selected longitudinal hippocampus. Our method shows the most consistent segmentation of local atrophy regions with the ground truth. In contrast, other segmentation pipelines reveal thickness increases in some atrophic areas, which contradicts the ground truth. Furthermore, both the Freesurfer cross-sectional and longitudinal approaches demonstrate unsmooth hippocampal boundaries, which are anatomically implausible. This issue requires particular attention in local morphology studies.

In addition, from Fig. 5, we noted that the GeoLongSeg demonstrates a conservative estimation of atrophy in regions that should have exhibited pronounced local atrophy, identified by red in the first row of the figure. Also, it exhibits some overestimation in regions with subtle atrophy primarily in the edge of lateral posterior part of the hippocampus. This is attributed to errors in local segmentation. The principal challenge of our method in local segmentation lies in the accuracy of the geodesic morphological regression in estimating the location and magnitude of deformation. Additionally, the longitudinal hippocampus requires a surface rigid registration process prior to the morphological regression. The precision of the registration process influences the accuracy of the morphological regression, potentially resulting in exaggerated or underestimated local atrophy identification. Future improvements should take into consideration both of these aspects.

4.2. GeoLongSeg facilitates the discovery of local hippocampal atrophy patterns in dementia progression

To demonstrate that our method generates more accurate segmentations and thereby enhancing the measurement of local atrophy, we conduct an assessment of local hippocampal atrophy in patients at early stage of dementia. The identified local features are utilized to classify dementia patients and controls. This verification is based on the hypothesis that improved longitudinal segmentation accuracy would lead to more informative features for disease classification.

As shown in Fig. 6, statistical analysis based on different segmentation methods yields different significant atrophy distributions. The results obtained through Synthseg segmentation reveal more pronounced areas of atrophy compared to other methods. Results based on GeoLongSeg segmentation display similar atrophy patterns, but identify fewer positive locations. Classification based on these local features reveals that the features derived from Synthseg segmentation achieve lower accuracy compared to those of GeoLongSeg, suggesting the presence of some false positives in detecting significant atrophy. Results based on segmentations from FSCross and FSLong do not identify significant local atrophy in the left hippocampus, and no temporal atrophy is found in the right hippocampus. However, FSLong exhibits significant spatial atrophy in the right hippocampus, which is consistent with Synthseg. Both methods yield comparable performance in identification of dementia in early stage, but they have lower performance than that of GeoLongSeg and Synthseg. This suggests that Freesurfer's segmentation fails to adequately reveal atrophy on the left hippocampus.

Interestingly, our results show different local atrophy patterns based on different segmentation methods. The results based on methods other than GeoLongSeg indicate that volumetric features have more advantages than local features. For instance, the cross-sectional and longitudinal pipelines of Freesurfer exhibit higher classification accuracy when using the volumetric features, whereas the accuracy decreases when incorporating local atrophic features. However, the local features derived from GeoLongSeg exhibit superior classification performance to other features, suggesting a negative impact of volumetric features in classification. Therefore, it remains unclear whether local atrophic features or volume contribute more to the dementia detection. To address this question, larger samples and more representative datasets should be used for further validation.

4.3. Limitations

This study has several limitations that should be acknowledged. Firstly, we annotate only a limited amount of data as the training set due to the labor-intensive manual labeling in longitudinal images. This may introduce bias and variability in the training data, potentially impacting the generalizability of the proposed pipeline. However, we believe that utilizing the proposed method for segmenting data from ADNI would still provide more reliable results compared to alternative methods. Secondly, the geodesic regression technique relies on the baseline morphology to estimate other observations, which necessitates high-quality baseline images. In cases where the baseline hippocampal segmentation results exhibit significant morphological errors, it can potentially affect local atrophy assessment. Above limitations emphasize the need for future research to address these challenges and refine the proposed methodology.

5. Conclusion

This paper introduces GeoLongSeg, a novel method for longitudinal hippocampal segmentation in 3T T1-weighted MRI scans. By incorporating diffeomorphic geodesic guidance into a deep learning network, GeoLongSeg achieves improved accuracy and morphological consistency within longitudinal hippocampi. Extensive evaluations on ADNI data are conducted to validate the proposed method's effectiveness in assessing the progression of hippocampal atrophy in dementia. GeoLongSeg provides a deep learning network pipeline designed for longitudinal MRI, offering precise and efficient segmentation for morphological studies of the hippocampal atrophy in dementia progression in the early stage.

CRediT authorship contribution statement

Na Gao: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Hantao Chen:** Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Xutao Guo:** Methodology, Software, Writing – original draft. **Xingyu Hao:** Formal analysis, Software. **Ting Ma:** Conceptualization, Funding acquisition, Methodology, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This study is supported by grants from the National Natural Science Foundation of P.R. China (62276081), Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (NO:ZYYCXTD-C-202004).

Data and Code Availability Statement

Data supporting the findings of this study were enrolled from Alzheimer' s Disease Neuroimaging Initiative (ADNI) database: www.loni.ucla.edu/ADNI/. The inclusion and exclusion criteria were listed in "Data preparation" section of manuscripts. All code for this project, including the manual segmentation, are available at a github repository: https://github.com/calliegao/ARMM.

References

Adler, D.H., Wisse, L.E.M., Ittyerah, R., Pluta, J.B., Ding, S.L., Xie, L., Wang, J., Kadivar, S., et al., 2018. Characterizing the human hippocampus in aging and Alzheimer's disease using a computational atlas derived from ex vivo MRI and histology. Proc Natl Acad Sci U S A 115, 4252–4257.

B.B. Avants N.J. Tustison M. Stauffer G. Song B.H. Wu J.C. Gee The Insight ToolKit image registration framework. Front Neuroinform 8 2014.

Billot, B., Greve, D.N., Puonti, O., et al., 2023. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Med Image Anal. 86, 102789.

Braak, H., Braak, E., 1997b. Staging of Alzheimer-related cortical destruction. Int Psychogeriatr 9 Suppl 1, 257-261; discussion 269-272.

Braak, E., Braak, H., 1997. Alzheimer's disease: transiently developing dendritic changes in pyramidal cells of sector CA1 of the Ammon's horn. Acta Neuropathol 93, 323–325.

Brown, E.M., Pierce, M.E., Clark, D.C., Fischl, B.R., Iglesias, J.E., Milberg, W.P., McGlinchey, R.E., Salat, D.H., 2020. Test-retest reliability of FreeSurfer automated hippocampal subfield segmentation within and across scanners. Neuroimage 210.

Chauveau, L., Kuhn, E., et al., 2021. Medial Temporal Lobe Subregional Atrophy in Aging and Alzheimer's Disease: A Longitudinal Study. Frontiers in Aging Neuroscience 13.

Das, S.R., Avants, B.B., Pluta, J., Wang, H., Suh, J.W., Weiner, M.W., Mueller, S.G., Yushkevich, P.A., 2012. Measuring longitudinal change in the hippocampal formation from in vivo high-resolution T2-weighted MRI. Neuroimage 60, 1266–1279.

Dong, M.J., Xie, L., Das, S.R., Wang, J.C., Wisse, L.E.M., deFlores, R., Wolk, D.A., Yushkevich, P.A., et al., 2021. DeepAtrophy: Teaching a neural network to detect progressive changes in longitudinal MRI of the hippocampal region in Alzheimer's disease. Neuroimage 243.

Dubois, B., Feldman, H.H., et al., 2014. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. Lancet Neurol 13, 614–629.

Fischl, B., Salat, D.H., Busa, E., et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Fishbaugh, J., Prastawa, M., Gerig, G., Durrleman, S., 2013. Geodesic shape regression in the framework of currents. Inf Process Med Imaging. 23, 718–729.

Fishbaugh, J., Durrleman, S., Prastawa, M., Gerig, G., 2017. Geodesic shape regression with multiple geometries and sparse parameters. Med. Image Anal. 39, 1–17.

Frisoni, G.B., Fox, N.C., et al., 2010. The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. 6, 67–77.

Gabere, M., Thu Pham, N.T., et al., 2020. Automated Hippocampal Subfield Volumetric Analyses in Atypical Alzheimer's Disease. Journal of Alzheimers Disease 78, 927–937.

Gao, N., Liu, Z., Deng, Y., Chen, H., Ye, C., Yang, Q., Ma, T., 2023. MR-based spatiotemporal anisotropic atrophy evaluation of hippocampus in Alzheimer's disease progression by multiscale skeletal representation. Hum. Brain Mapp. 1–18.

Hill, D.L.G., Schwarz, A.J., et al., 2014. Coalition Against Major Diseases/European Medicines Agency biomarker qualification of hippocampal volume for enrichment of clinical trials in predementia stages of Alzheimer's disease. Alzheimers Dement. 10 (421–429), e423.

Iglesias, J.E., Van Leemput, K., Augustinack, J., Insausti, R., Fischl, B., Reuter, M., AsDN, I., 2016. Bayesian longitudinal segmentation of hippocampal substructures in brain MRI using subject-specific atlases. Neuroimage 141, 542–555.

Jack, C.R., Bernstein, M.A., Borowski, B.J., Gunter, J.L., Fox, N.C., Thompson, P.M., Schuff, N., Krueger, G., et al., 2010. Update on the Magnetic Resonance Imaging core of the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement. 6, 212–220.

Kerchner, G.A., Hess, C.P., et al., 2010. Hippocampal CA1 apical neuropil atrophy in mild Alzheimer disease visualized with 7-T MRI. Neurology 75, 1381–1387.

J. Kulaga-Yoskovitz, B.C. Bernhardt, S. Hong, T. Mansi, K. Liang, A. van der Kouve, J. Smallwood, A. Bernasconi, N. Bernasconi (2015) Multi-contrast and submillimetric 3T hippocampal subfield segmentation protocol and dataset. Scientific data, in press.

Li, H., et al., 2021b. Longitudinal subcortical segmentation with deep learning. Conference on Medical Imaging: Image Processing.

Li, H., Zhang, H.H., Johnson, H., Long, J.D., Paulsen, J.S., Oguz, I., 2021a. Longitudinal subcortical segmentation with deep learning. Proc Spie 11596.

Martin, S.B., Smith, C.D., et al., 2010. Evidence that volume of anterior medial temporal lobe is reduced in seniors destined for mild cognitive impairment. Neurobiol. Aging 31, 1099–1106.

Muralidharan, P., Fishbaugh, J., Johnson, H.J., Durrleman, S., Paulsen, J.S., Gerig, G., Fletcher, P.T., 2014. Diffeomorphic Shape Trajectories for Improved Longitudinal Segmentation and Statistics. Medical Image Computing and Computer-Assisted Intervention - Miccai 2014. Pt Iii 8675, 49–56.

Oktay O , Schlemper J , Folgoc L L ,et al.Attention U-Net: Learning Where to Look for the Pancreas[J]. 2018.

Platero, C., Lin, L., Tobar, M.C., 2019. Longitudinal Neuroimaging Hippocampal Markers for Diagnosing Alzheimer's Disease. Neuroinformatics 17, 43–61.

Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61 (4), 1402–1418.

Scheff, S.W., Price, D.A., et al., 2007. Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment. Neurology 68, 1501–1508.

Shaw, T., York, A., Ziaei, M., Barth, M., Bollmann, S., Neuroimaging, A.D., 2020. Longitudinal Automatic Segmentation of Hippocampal Subfields (LASHiS) using multi-contrast MRI. Neuroimage 218.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., et al., 2011. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 280–292.

Tang, X.Y., Holland, D., et al., 2015. The Diffeomorphometry of Regional Shape Change Rates and its Relevance to Cognitive Deterioration in Mild Cognitive Impairment and Alzheimer's Disease. Hum. Brain Mapp. 36, 2093–2117.

Wei C, Yang Y, Guo X, et al. MRF-Net: A multi-branch residual fusion network for fast and accurate whole-brain MRI segmentation. Front Neurosci. 2022;16:940381. Published 2022 Sep 12.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Luthman, J., Morris, J.C., Petersen, R. C., Saykin, A.J., Shaw, L., Shen, L., Schwarz, A., Toga, A.W., Trojanowski, J.Q., 2015. Neuroimaging Initiative D, Dement Author manuscript A, 2015. 2014 update of the Alzheimer's disease neuroimaging initiative: a review of papers published since its inception HHS public access author manuscript. Alzheimers Dement. 11, 1–120.

Zhang, L., Mak, E., et al., 2020. Longitudinal trajectory of Amyloid-related hippocampal subfield atrophy in nondemented elderly. Hum. Brain Mapp. 41, 2037–2047.