Hippocampus Analysis by Combination of 3-D DenseNet and Shapes for Alzheimer's Disease Diagnosis

Ruoxuan Cui, Manhua Liu, and the Alzheimer's Disease Neuroimaging Initiative

***Abstract*—Hippocampus is one of the first involved regions in Alzheimer's disease (AD) and mild cognitive impairment (MCI), a prodromal stage of AD. Hippocampal atrophy is a validated, easily accessible, and widely used biomarker for AD diagnosis. Most of existing methods compute the shape and volume features for hippocampus analysis using structural magnetic resonance images (MRI). However, the regions adjacent to hippocampus may be relevant to AD, and the visual features of the hippocampal region are important for disease diagnosis. In this paper, we have proposed a new hippocampus analysis method to combine the global and local features of hippocampus by three-dimensional densely connected convolutional networks and shape analysis for AD diagnosis. The proposed method can make use of the local visual and global shape features to enhance the classification. Tissue segmentation and nonlinear registration are not required in the proposed method. Our method is evaluated with the T1-weighted structural MRIs from 811 subjects including 192 AD, 396 MCI (231 stable MCI and 165 progressive MCI), and 223 normal control in Alzheimer's disease neuroimaging initiative database. Experimental results show the proposed method achieves a classification accuracy of 92.29% and area under the ROC curve of 96.95% for AD diagnosis. Results comparison demonstrates the proposed method performs better than other methods.**

***Index Terms*—Alzheimer's disease, hippocampus, 3D densenet, deep learning, structural magnetic resonance image.**

I. Introduction

ALZHEIMER'S disease (AD) is an irreversible and chronic neurodegenerative disease with progressive impairment of the memory and cognitive functions. In the World Alzheimer Report 2015, AD is the most common form of dementia as it accounts for 60–80% of the cases [1]. It was proven that AD causes cell death and damage of tissue nerves throughout the brain. Mild cognitive impairment (MCI) is a prodromal stage of AD including the progressive MCI (pMCI) and stable MCI (sMCI). pMCI means that the MCI subjects will convert to AD after some time while sMCI subjects are stable. Currently, there are no effective cure to stop or reverse AD progress, but some treatments can be developed to delay its progression, especially if AD is diagnosed at the early stage. Magnetic resonance images (MRI) is widely used as imaging modality for AD diagnosis in clinic. It can non-invasively capture the internal body structures and the regional brain atrophy, helping us understand the anatomical and functional changes [2]. Structural MRI scans provide detailed information about the anatomical structures and the morphology of brain tissues such as white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF). In recent years, extensive efforts have been done to find the biomarkers and develop computer-aided diagnosis system using pattern recognition methods to decode the disease states with structural MRIs [2], [3].

One popular method is based on the regions of interest (ROIs), which partition the MR brain image into multiple anatomical regions by warping of a labeled atlas, and the regional measurements such as volumes and shapes are computed as the features for AD diagnosis [3], [4]. Zhang *et al.* [4] proposed to extract tissue volumes from 93 ROIs to train the support vector machine (SVM) classifier for AD diagnosis. A approach was proposed to combine the marginal fisher analysis with norm based multi-kernel learning to achieve the sparsity of ROIs, which can simultaneously select the relevant brain regions and learn a dimensionality transformation for classification [5]. Recently, deep learning networks have been investigated to learn the image features and classifiers for AD/MCI diagnosis [3], [6]–[11]. A deep learning network on stacked auto-encoder (SAE) was proposed to learn the complicated features inherent in the low-level ROI features, which were combined with the original features to enhance the classification [3]. The whole brain is parcellated into a number of brain regions with a deep belief network trained on each region, and multiple deep belief networks was ensembled for AD diagnosis [10]. The ROI based method can make use of all brain regions to achieve good classification performance. There are still some limitations in these methods. First, the definition of ROIs requires accumulation of long-term experience

Manuscript received May 4, 2018; revised August 28, 2018 and October 10, 2018; accepted November 9, 2018. Date of publication November 20, 2018; date of current version September 4, 2019. This work was supported in part by the National Natural Science Foundation of China (No. 6181101049, 61375112, and 61773263), in part by the National Key Research and Development Program of China (No.2016YFC0100903), in part by the National Key Basic Research Program of China (973 Project, No.2015CB931802), and in part by the SMC Excellent Young Faculty Program of SJTU. *(Corresponding author: Manhua Liu.)*

The authors are with the Department of Instrument Science and Engineering, School of EIEE, Shanghai Engineering Research Center for Intelligent Diagnosis and Treatment Instrument, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: sodacui@sjtu.edu.cn; mhliu@sjtu.edu.cn).

Digital Object Identifier 10.1109/JBHI.2018.2882392

of researchers. Second, ROI segmentation is affected by the individual differences and subjective factors of experts. Third, the morphological abnormalities caused by the brain disorders may involve multiple ROIs or part of the pre-defined ROIs, so the performance may not be stable. To avoid the multiple processing steps in feature extraction, the residual and plain 3D CNN architectures have been investigated for automatic feature generation and more straightforward analysis of whole brain MRI scans [11]. The landmark-based deep learning framework was proposed to automatically extract patch-based representation of MRI for AD diagnosis [6], [9].

Among all parcellated ROIs in brain, hippocampus is proven to be a vitally important region for AD. There are many studies in the literature to extract features from hippocampus region for AD diagnosis using structural MRI [12]–[17]. Most existing methods are based on the shape and volume analysis of bilateral hippocampi. Hippocampal atrophy is one of the most validated, easily accessible and widely used biomarkers of AD. The reduction of hippocampus volume in brain structure can be visible and measured using structural MRI. Thus, there are some studies proposed to compute the hippocampal volumes using MRI for AD diagnosis [13], [15], [17]. In [18], a fully automatic method using probabilistic and anatomical priors was proposed for hippocampus segmentation and the hippocampus volumes were computed for AD classification. In addition, a fast multi-atlas segmentation method was proposed to automatically measure hippocampal volume for discriminating AD or MCI from NC [13], [17]. However, volumetric analysis only assesses global hippocampal changes and suffers from the variety of hippocampus volumes among different individuals. Therefore, shape analysis was used to capture the hippocampus morphology for AD diagnosis [12], [14], [19], [20]. It can unveil the local atrophy of hippocampus and is more sensitive in the early stage of AD. In [12], the hippocampus shape analysis was performed by mapping the segmentation mask into spherical harmonic shape description (SPHARM-PDM) and then the original surface locations were described with sets of coefficients weighting spherical harmonic basis functions.

However, there are still some limitations on the existing hippocampus analysis methods. First, both volumetric and shape analysis of hippocampus depend on the accurate segmentation of brain ROIs. It is not an easy task to achieve the accurate ROI segmentation. Second, hippocampus is not sufficient for discrimination of MCI from NC. Other regions adjacent to hippocampus such as parahippocampus and amygdala are also involved in AD. Third, the visual features of MR images derived from the hippocampal region can be of great help for AD diagnosis [16], [21].

To overcome the above limitations, this paper proposes a new method for hippocampus analysis based on combination of 3D Densely Connected Convolutional Networks (3D DenseNet) and traditional shape features for AD diagnosis, which integrates the multi-level and multi-type features for improving the disease classification. First, hippocampus segmentation is performed to generate the hippocampus masks and two 3D image patches centered on the centroids of bilateral hippocampi are extracted for each MRI. Second, motivated by the success of Convolutional Neural Networks, a 3D DenseNet is trained with the 3D patches of each hippocampus to learn the visual features for image classification. Third, the shape analysis is performed on the hippocampus mask using SPHARM-PDM tool to extract the shape descriptors, followed by a Multi-Layer Perceptron (MLP) for classification task. Finally, the outputs of 3D DenseNets and MLP are combined for disease classification.

Compared to the existing methods, our proposed method has following improvements: 1) The deep DenseNet can jointly learn the features of image patches and disease classifier, while the shape features capture the global shapes of hippocampus mask. These two types of features can provide complementary information and thus are combined to enhance disease classification. 2) Training a DenseNet requires a large image set, which is not applicable for AD and MCI diagnosis. Instead of training a deep DenseNet with the whole brain image, we build a DenseNet on the local patches from hippocampus, which can reduce the number of parameters and alleviate the problem of small image set. 3) No tissue segmentation and nonlinear registration are required in image processing, which can simplify the diagnosis procedure and save computation costs. The rest of this paper is organized as follows. In Section II, we present the image set used in this study and the proposed method in details. In Section III, we provide the experimental results and discussion. A conclusion is given in Section IV.

II. Proposed Method

In this section, we presented the proposed classification algorithm for hippocampus analysis. Our method works on the T1-weighted MR brain images which are widely available, non-invasive and often used as the first biomarker in AD diagnosis. Fig. 1 shows the flowchart of our proposed algorithm, which consists of four main steps: hippocampus segmentation and patch extraction, construction of 3D DenseNet models, MLP based shape analysis and final ensemble classification. More details of these steps were provided in the following subsections.

A. The Image Set and Processing

The structural MR brain images studied in this work were obtained from ADNI database, which can be freely downloaded from the website (www.loni.ucla.edu/ADNI). In 2003, the ADNI was launched as a $60 million, 5-year public–private partnership to test whether serial MRI, Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Detailed information about data acquisition is available at the ADNI website.

In this study, we use the T1-weighted MR images acquired with 1.5T scanners from 811 subjects as shown in Table I. The MR images are first preprocessed by using a nonparametric nonuniform intensity normalization (N3) algorithm [22] to correct the non-uniform intensity. After correction, skull-stripping is performed by the method [23] followed by cerebellum removal with Freesurfer software. We manually check the skull-stripped images to ensure clean and dura removal. The rigid registration is used to align the MR images to a

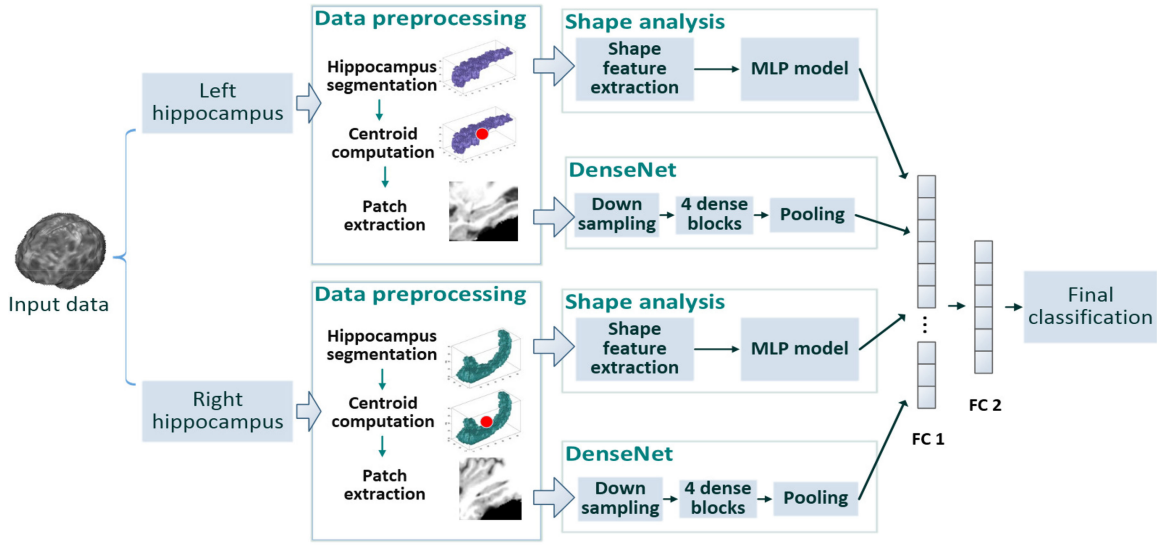

Fig. 1. The flowchart of our proposed algorithm consisting of four main steps: hippocampus segmentation and patch extraction, construction of 3D DenseNet models, MLP based shape analysis, and final ensemble classification for AD and MCI diagnosis.

TABLE I
DEMOGRAPHIC CHARACTERISTICS OF THE STUDIED SUBJECTS FROM ADNI DATABASE (THE VALUES ARE DENOTED AS MEAN ± STANDARD DEVIATION)

Diagnosis	Subject #	Age	Gender (M/F)	MMSE	Education	CDR
AD	192	75.7 ± 7.7	100/92	23.3 ± 2.0	15.0 ± 3.0	0.8 ± 0.3
pMCI	165	74.4 ± 7.2	102/63	26.5 ± 1.7	15.9 ± 2.5	0.5 ± 0
sMCI	231	74.6 ± 7.7	158/73	27.2 ± 1.8	15.4 ± 2.7	0.5 ± 0
NC	223	76.0 ± 5.0	118/105	29.1 ± 1.0	15.6 ± 2.7	0 ± 0

template space with FMRIB Software Library (FSL) 5.0 from https://fsl.fmrib.ox.ac.uk/.

B. Hippocampus Segmentation and Patch Extraction

Hippocampus is often considered as one of the first affected brain regions in AD. There are two hippocampi (left and right) in brain. For hippocampus analysis, we first segment the hippocampus from other regions using FSL 5.0 and a binary mask is generated for each hippocampus. We also manually check the results to ensure the correct segmentation. For the failures, we remove the subjects. The MRIs of 834 subjects were initially downloaded from ADNI to test our algorithm. There are 23 failed subjects of 7 AD, 4 pMCI, 6 sMCI and 6 NC. After removing the failed subjects, the final sample size is 811 including 192 AD, 165 pMCI, 231 sMCI and 223 NC subjects. Then, we calculate the centroid of each hippocampus based on the mask. Finally, a 3D patch of fixed size is extracted from the MR images to be centered on hippocampus centroid as the inputs of DenseNets. To achieve the robustness of location variance, each hippocampus centroid is shifted by ±2 voxels in x, y and z coordinates to extract more patches for training. In the test phase, no shift is for patch extraction. Since some regions surrounding the hippocampus are also useful for AD diagnosis, the patch size is so large to cover the whole hippocampus and its surrounding regions such as amygdala and parahippocampal gyrus. The patches from left and right hippocampus have overlapping of 6–16 voxels.

C. The Deep 3D DenseNet Model

After extraction of patches from hippocampus region, we learn features from image patches for classification. Convolutional neural networks (CNNs), which alternatively stacks convolutional and pooling layers followed by fully connected and softmax layers, were investigated to learn features for image classification. To enhance the representation power, more layers are added to build the deep CNNs, which increases the information loss when the input information passes through many layers to reach the end of network. Thus, the Densely Connected Convolutional Networks (DenseNet) was proposed to connect each layer to every other layer in a feed-forward fashion, which increased direct connections between the low and high layers [24]. In the DenseNet, the layers from different levels are densely connected to improve information flow between layers. Compared to deep CNN, DenseNets can alleviate the vanishing-gradient problem because there is a direct connection from the low to high layers. In addition, feature propagation is strengthened to reuse the low-level features with less information loss and the number of parameters is reduced.

In this work, we propose to construct a 3D DenseNet for learning the features from 3D image patches for AD diagnosis. Fig. 2 shows the structure of 3D DenseNet model consisting of a convolutional layer, 4 dense blocks, 3 transition layers, an average pooling layer and a softmax layer. First, convolutional layer is added after the input layer with a stride of $2 \times 2 \times 2$, followed by 4 dense blocks. The dense block uses a dense connectivity through which the lth layer receives the output feature maps of all preceding layers as [24]:

$$x_l = H_l([x_0, x_1, \ldots, x_l - 1]) \quad (1)$$

where $[x_0, x_1, \ldots, x_{l-1}]$ is the concatenation of the feature maps from all previous layers into a single tensor, and H_l denotes

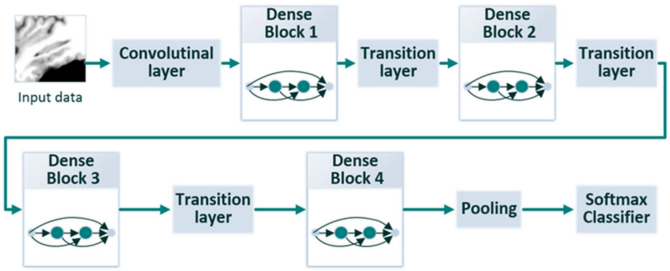

Fig. 2. The structure of 3D DenseNet consisting of a convolutional layer, four dense blocks, three transition layers, an average pooling and a softmax layers.

a composite nonlinear transformation function of four consecutive operations: batch normalization, leaky rectified linear units, $3 \times 3 \times 3$ convolution and dropout. In each dense block, every dense layer receives the feature maps of all previous dense layers by shortcut connection. The dense layer includes one $1 \times 1 \times 1$ and one $3 \times 3 \times 3$ convolutional layers, two batch normalization layers and two activation layers. Dense block 1 and 2 consist of three dense layers, while the other two dense blocks contain two dense layers. Between two dense blocks, a transition layer is set to achieve dimension reduction of the feature maps. It consists of five consecutive operations: batch normalization, leaky rectified linear units, a $1 \times 1 \times 1$ convolution, dropout and a $3 \times 3 \times 3$ convolution with a stride of $2 \times 2 \times 2$. Following the last dense block, an average pooling and a softmax classifier are appended to reduce the feature dimension and make the classification. The subject labels are used through back-propagation for updating the weights of DenseNet and learning the relevant features. All layers receive the direct supervision from the loss function through shortcut connections. The 3D DenseNets for the left and right hippocampi have the same structure but they are trained with different patches. Table II lists the output size and the parameters of each network layer in the 3D DenseNet model.

TABLE II
THE ARCHITECTURE OF THE 3D DENSENET

Layers	Output size	Parameters
Input layer	62×48×58	
Convolution layer	64, 31×24×29	Kernel size: (3,3,3) Stride: (2,2,2)
Dense Block (1)	64, 16×12×15	$\begin{bmatrix} 1\times1\times1 \\ 3\times3\times3 \end{bmatrix} \times 2$
Transition layer (Conv_pooling)	64, 9×7×8	
Dense block (2)	64, 9×7×8	$\begin{bmatrix} 1\times1\times1 \\ 3\times3\times3 \end{bmatrix} \times 2$
Transition layer (Conv_pooling)	64, 5×4×5	
Dense block (3)	64, 5×4×5	$\begin{bmatrix} 1\times1\times1 \\ 3\times3\times3 \end{bmatrix} \times 2$
Transition layer (Conv_pooling)	64, 3×3×3	
Dense block (4)	64, 3×3×3	$\begin{bmatrix} 1\times1\times1 \\ 3\times3\times3 \end{bmatrix} \times 2$
Global average pooling	128, 1×1×1	
Softmax layer	2	

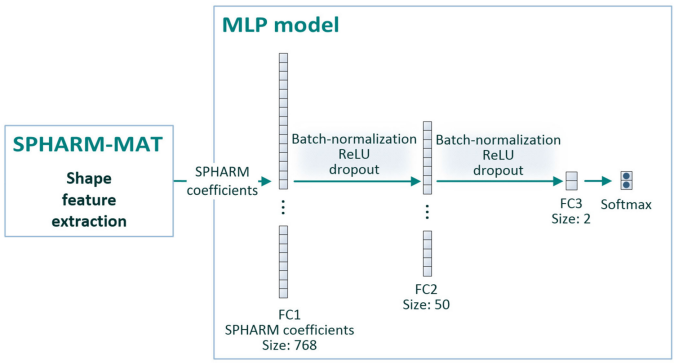

Fig. 3. The flowchart of hippocampal shape analysis, where SPHARM coefficients are computed with SPHARM-MAT, followed by MLP model for classification task.

D. Shape Analysis

It was proven that the AD progression has a strong correlation with the hippocampal atrophy. Although the 3D DenseNet is effective to learn the visual features of MR images, the hippocampal shapes are also important features to differentiate the hippocampal atrophies. Thus, we perform the shape analysis on the hippocampal mask for classification. Fig. 3 illustrates the flow chart of the hippocampal shape analysis which consists of shape feature extraction and classification. For feature extraction, the tool of SPHARM-MAT [22] (SPHARM Modeling and Analysis Toolkit) on MATLAB 2012b is used for shape analysis in this work. It consists of four processing steps: topology fix, spherical parameterization, SPHARM expansion and surface alignment. First, the voxel-wise surfaces of all binary hippocampus masks are fixed to have a spherical topology. Second, we create spherical parameterization for surface meshes and generate a continuous mapping from the hippocampus surface to the surface of a unit. Third, SPHARM expansion is performed on the surface into a complete set of spherical harmonic basis functions, which are essentially Fourier basis functions defined on the sphere. Finally, surface alignment is conducted for group analysis across different surface samples.

After the above operations, we obtain the SPHARM coefficients for each hippocampus mask, which are used as the shape features to train a Multi-Layer Perceptron (MLP) model with three full connected layers and one softmax layer. Similar to the DenseNet models mentioned above, two MLP models sharing the same structure are constructed for the bilateral hippocampi. In total, the MLP model contains ten layers, including the input layer, two batch normalization layers, two *ReLU* activation, two dropout, two full-connected layers and one softmax layer. It performs well with a simple structure and less number of parameters for shape analysis.

E. Final Classification

In the above sections, the DenseNets capture the visual features while the shape analysis by MLP is effective to capture the hippocampal shape features for classification. In this work, we

TABLE III
COMPARISON OF DIFFERENT PATCH SIZES FOR CLASSIFICATIONS OF AD VS. NC, PMCI VS. SMCI, AND MCI VS. NC

Patch size	AD vs. NC (%)				pMCI vs. sMCI (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
57×45×53	88.43	84.38	91.93	91.32	71.72	66.67	75.32	75.26	72.54	75.51	67.26	75.67
62×48×58	90.12	86.98	92.83	92.83	73.23	69.70	75.76	76.74	73.02	75.25	69.06	76.17
68×54×64	89.88	86.46	92.83	92.56	72.98	70.30	74.89	76.39	72.86	75.00	69.06	76.20

investigate two strategies to combine these networks to improve classification. First, we take the weighted sum of prediction scores from 4 separately training networks. Considering that DenseNet can capture more discriminative features than shape analysis, the weights are set to 0.2 and 0.3 for each MLP and DenseNet output scores, respectively (all weights are summed to 1). Second, instead of taking the sum of output scores, we use two full connected layers to combine the 4 networks which are initially trained individually and finely tuned by end-to-end training. We have implemented and compared these two methods and found that the combined network by full connection performs slightly better than the weighted sum. The combined network is an end-to-end training to optimize the parameters of whole network.

In summary, our proposed method combines the deep learning model and MLP based shape analysis to make full use of the shape and local visual features of MR brain images for AD diagnosis. Specially, we learn the visual features using 3D DenseNets and extract the shape features with SPHARM-MAT. These two methods are further combined to make the final classification. Compared to the existing methods for classification of MR brain images, our proposed method has the following advantages. First, both the local and global hippocampal information are used to enhance the classification. Second, instead of using the whole image, the proposed method extracted the patches from hippocampal regions to train 3D DenseNets. This can avoid building very deep DenseNets and the training data can be augmented by shifting patches. Third, no tissue segmentation and nonlinear registration are required in image processing, which can simplify the diagnosis procedure and save the computation costs.

III. EXPERIMENTAL RESULTS

In this section, we first introduce the datasets and implementation of our method. Then, we present the extensive experiments to test our method on the classifications of AD vs. NC, pMCI vs. sMCI and MCI vs. NC. Finally, we compare our method with other methods and give the discussion.

A. Datasets and Implementation

The proposed method is tested on the T1-weighted MR brain images from ADNI database. The image preprocessing was conducted as illustrated in Section II. The proposed method is tested on classifications of AD vs. NC, pMCI vs. sMCI and MCI vs. NC. The MR images were taken from the baseline visits of 811 ADNI participants for evaluation. The proposed method is implemented with python 2.7.9 and Keras library on the Tensorflow backend. For training the DenseNets, the initial weights for whole network is uniform, and Adam optimizer is adopted with a low learning rate of $1 \times 10-4$. The networks are stable after iteration of 200 epochs. The batch sizes are set to 64. *ReLU* activation is used for each neuron of DenseNet. All the experiments were conducted in the environment of Ubuntu14.04-x64/GPU of NVIDIA GeForce GTX 1080 Ti.

To avoid the overfitting problem, we used three techniques in building the DenseNet model. First, we shift the centroid of each hippocampus by ± 2 voxels in x, y and z coordinates to extract more patches from training subjects to greatly augment the training data. Second, we constructed a simplified version of DenseNet with the numbers of dense layers in the dense blocks set to [2, 2, 2, 2], which is much smaller than those of [6], [12], [24], [16] in the original DenseNet proposed by Huang *et al.* [24]. This reduces the depth of network structure and helps to avoid overfitting. Third, a dropout layer was added in each convolutional layer of DenseNet model to avoid overfitting.

To evaluate the performance, we used a 5-fold cross-validation strategy to train and test the proposed method. Each time, 1 fold of the data set was used for testing, while the other 4 folds was further split into training and validation parts. The validation part is used to adjust the training process for the optimal parameters. Four measures, i.e., classification accuracy (ACC), sensitivity (SEN), specificity (SPE), receiver operating characteristic (ROC) curve and the area under ROC curve (AUC), are computed for evaluation. ACC is computed as the proportion of correctly classified subjects among the whole population. SEN is the proportion of correctly classified positive samples (AD/MCI subjects) among the total number of positive samples. SPE is the proportion of correctly classified negative samples (NC subjects) among the total number of negative samples. The ROC curve is generated by plotting the true positive rate (TPR) against the false positive rate (FPR) at various thresholds on the class prediction scores.

B. Test the Effects of Patch Size

The first experiment is to test the effects of different patch sizes on the classification performance of DenseNet. Large patch covers more neighboring information of hippocampus. In the experiment, we extracted the patches by gradually increasing its size from $57 \times 45 \times 53$, $62 \times 48 \times 58$ to $68 \times 54 \times 64$. Table III compares the classification performances of three different patch sizes for AD vs. NC, pMCI vs. sMCI and MCI vs. NC. From these results, we can see the classification performances are improved by increasing the patch size from $57 \times 45 \times 53$ to $62 \times 48 \times 58$. But they cannot be improved by further increasing

TABLE IV
COMPARISON OF DIFFERENT DEEP LEARNING MODELS FOR CLASSIFICATIONS OF AD VS. NC, PMCI VS. SMCI, AND MCI VS. NC

Method	AD vs. NC (%)				pMCI vs. sMCI (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
LeNet	82.17	79.17	84.75	83.14	67.93	61.21	72.73	70.34	67.85	70.71	62.78	70.97
VGGNet	83.86	82.29	85.20	85.12	69.95	64.85	73.59	71.09	69.47	71.97	65.02	73.05
Dense Block 1	81.93	76.56	86.55	83.09	68.94	64.24	72.29	71.38	68.17	71.21	62.78	72.12
Dense Block 2	87.95	84.38	91.03	90.37	70.96	66.67	74.03	72.27	71.24	73.23	67.71	73.25
Dense Block 3	88.67	85.42	91.48	91.04	71.46	67.88	74.03	74.05	71.57	73.48	68.16	74.32
Dense Block 4	90.12	86.98	92.83	92.83	73.23	69.70	75.76	76.74	73.02	75.25	69.06	76.17

TABLE V
COMPARISON OF DIFFERENT METHODS FOR CLASSIFICATIONS OF AD VS. NC, PMCI VS. SMCI, AND MCI VS. NC

Method	AD vs. NC (%)				pMCI vs. sMCI (%)				MCI vs. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
DenseNet	90.12	86.98	92.83	92.83	73.23	69.70	75.76	76.74	73.02	75.25	69.06	76.17
Shape analysis	83.86	80.21	87.00	85.54	71.46	68.48	73.59	76.09	70.27	72.47	66.37	74.55
Combination by scores summation	92.05	90.10	93.72	96.77	74.49	72.73	75.76	79.03	74.15	76.77	69.51	77.48
Combination by full connection	92.29	90.63	93.72	96.95	75.00	73.33	76.19	79.70	74.64	77.27	69.96	77.70

the patch size. Thus, the $62 \times 48 \times 58$ patch is used to train the DenseNets in our following experiments.

C. Test the Effectiveness of 3D DenseNet and Shape Analysis

The second experiment is in depth analysis of all stages in the proposed method. We first compare classification performances using the DenseNet with different number of dense blocks and two well-known CNN models: LeNet [25] and VGGNet [26], as shown in Table IV. The LeNet and VGGNet are implemented with the released codes by replacing the 2D convolution with 3D. We can see VGGNet performs better than LeNet and the performance of DenseNet is gradually improved by increasing the number of dense blocks till 4. The DenseNet model with 4 dense blocks achieves the best performance because of its reuse of multi-level rich features.

We further test the classification performances of 3D DenseNet, shape analysis and two different combination methods, separately. One combination method is by weighted sum of four output prediction scores from MLP based shape analysis and DenseNets. Another method appended the full connected layers to combine the features extracted from shape analysis and DenseNets for classification. The classification performances of these methods for AD and MCI diagnosis are compared in Table V. We can see that the combination method by full connection performs slightly better than the scores summation. Fig. 4(a), (b) and (c) illustrate the ROC curves for classifications of AD vs. NC, pMCI vs. sMCI and MCI vs. NC, respectively. The results show that the DenseNet performs better than the shape analysis, especially for classification of AD vs. NC. The classification performances are further improved by combination of DenseNet and shape analysis with the full connection.

D. Comparison With Other Methods

The third experiment is to compare our proposed method to other methods that are also based on T1-weighted structural MRI data of ADNI in the literature. Specifically, our method is compared to three conventional features such as hippocampal volume, GM volumes of ROIs and voxel-wise features in [4], [18], [27]. In [18], a fully automatic method was proposed for hippocampus segmentation to compute the hippocampal volumes. The GM volumes of 93 ROIs by parcellation of the whole brain were computed for classification [4]. The voxel-wise GM density maps were used for classification [27]. For fair comparison, we implemented these methods with our best efforts on the same training and test data sets.

In our experiments, the same image processing and classifier, i.e., softmax classifier, were used for all methods while feature extractions were different. To extract the conventional features, tissue segmentation and rigid registration were performed using the FAST in the FSL package [28] and HAMMER [29], respectively. The image is mapped into 93 ROIs to compute their GM volumes as the features. To extract the voxel-wise features, the warped tissue volumes by HAMMER [29] reflected the spatial density of tissues in an original brain and were downsampled by 4 and t-test is used to select the 10000 most discriminative features for classification. Table VI shows the comparison of classification performances by our proposed method and other three feature extractions with their ROC curves shown in Fig. 5. It is worthy to note that this evaluation is based on different feature extractions, not the design of classifier, so the results may be different from those reported in literature.

In addition, we compare our results with the reported results by the deep learning methods on ADNI in Table VII. We can see that our method performs better than other methods for

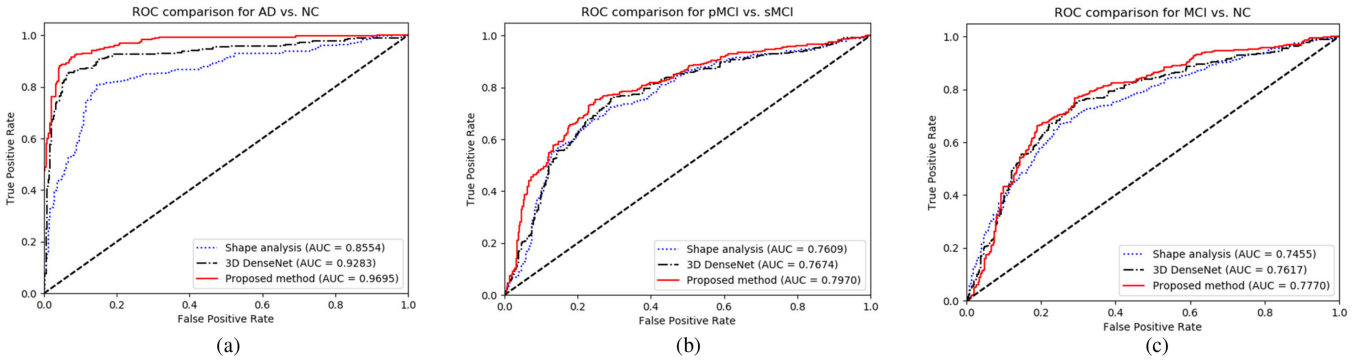

Fig. 4. Comparison of ROCs by shape analysis, DenseNet and proposed combination method for (a) AD vs. NC, (b) pMCI vs. sMCI, and (c) MCI vs. NC.

TABLE VI
COMPARISON OF OUR METHOD WITH OTHER METHODS IN THE LITERATURE FOR CLASSIFICATION OF AD VS. NC, PMCI VS. SMCI, AND MCI VS. NC

Method	AD VS. NC (%)				pMCI VS. sMCI (%)				MCI VS. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Hippocampal volumes	78.55	72.92	83.41	82.81	67.93	60.61	73.16	68.97	68.82	75.76	56.50	69.93
ROI features	86.51	81.77	90.58	90.53	73.23	70.30	75.32	71.89	73.34	78.03	65.02	71.83
Voxel-wise features	86.02	81.77	89.69	88.66	72.98	69.09	75.76	73.83	73.99	79.55	64.13	74.33
Proposed method	92.29	90.63	93.72	96.95	75.00	73.33	76.19	79.70	74.64	77.27	69.96	77.70

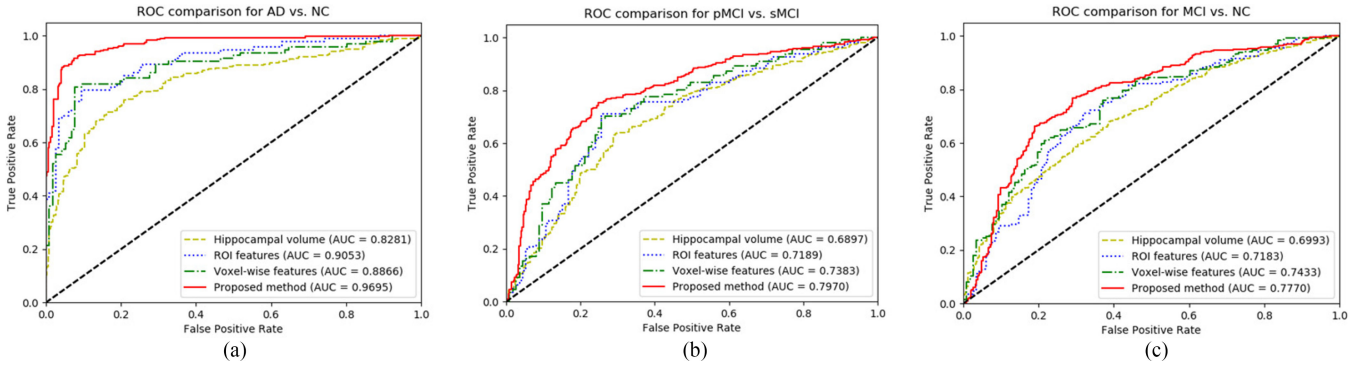

Fig. 5. Comparison of ROCs of our proposed method with other four methods for (a) AD vs. NC, (b) pMCI vs. sMCI, and (c) MCI vs. NC, respectively.

TABLE VII
COMPARISON OF OUR RESULTS WITH THE RESULTS OF OTHER DEEP LEARNING METHODS FOR CLASSIFICATION OF AD VS. NC, PMCI VS. SMCI, AND MCI VS. NC

Method	AD VS. NC (%)				pMCI VS. sMCI (%)				MCI VS. NC (%)			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Liu M. et al. [6]	91.09	88.05	93.50	95.86	76.90	42.11	82.43	77.64	-	-	-	-
Ortiz A. et al. [10]	90.00	86.00	94.00	95.00	78.00	61.00	88.00	82.00	80.00	60.00	90.00	84.00
Sergey K. et al. [11]	80.00	-	-	87.00	52.00	-	-	52.00	61.00	-	-	65.00
Proposed method	92.29	90.63	93.72	96.95	75.00	73.33	76.19	79.70	74.64	77.27	69.96	77.70

classification of AD vs. NC. Specifically, the whole brain images were used to train the residual and plain 3D CNNs for AD and MCI diagnosis [11]. Our method performs much better than this method for all classifications. The whole brain is parcellated into 116 regions with each trained on a deep belief network for ensemble classification using MRI and PET images [10]. Our method using only MRI achieves better results than this ensemble methodw for AD classification, but has lower performance for classification of other groups. A landmark-based deep learning framework was proposed for AD diagnosis [6]. This framework learns an end-to-end classifier to combine both the local and global features from landmarks. The result comparison further validates the efficacy of our proposed method for AD and MCI diagnosis. It is worth noting that the differences of these results may be caused not only by the deep learning methods and but also by different ADNI subjects.

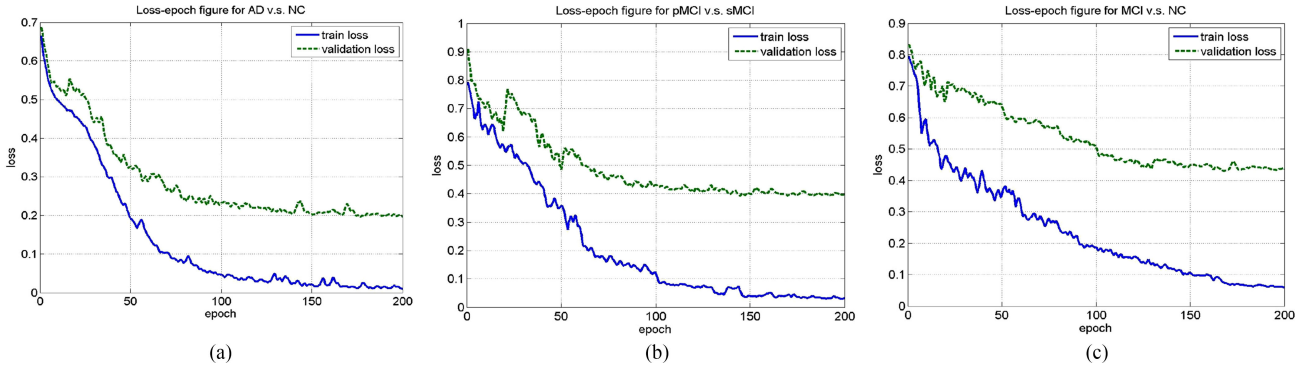

Fig. 6. The convergence of loss functions for training the classification models of (a) AD vs. NC, (b) pMCI vs. sMCI, and (c) MCI vs. NC.

E. Discussion

From the above results, we can see that the classification performance is low by using the hippocampal volumes because they have limited information for AD diagnosis. Our proposed method achieves a very competitive classification accuracy of 92.29% for classification of AD vs. NC, which is higher than the ROI-based method (86.51%) and voxel-based method (86.02%). Our method also achieves the accuracy of 75.00% for classification of pMCI vs. sMCI, higher than the ROI-based method (73.23%) and voxel-based method (72.98%).

Compared to the ROI methods that use multiple brain regions [10], [30], the proposed method focuses on the hippocampus with the following advantages. First, the proposed method can make full use of the visual and shape features of hippocampus, while the ROI methods just use the hippocampal volume which is not enough to represent the hippocampus. Second, different from the ROI methods, no tissue segmentation and nonlinear registration are required in the proposed method, which can reduce the computation costs and fast the process. The disadvantage is that the regions far from the hippocampus will not be considered in the proposed method.

As for the computational complexity, the proposed method includes the offline training and online test stages. Fig. 6(a), (b) and (c) show the convergence of the loss functions for training the classification models of AD vs. NC, pMCI vs. sMCI and MCI vs. NC, respectively. The results show that the AD vs. NC model converges after 150 epochs which is faster than others. Other models converge after 200 epochs. In the test stage, it takes about 0.058 s on average for a subject, which demonstrate the usefulness of the proposed method in real application.

IV. Conclusion

This paper has proposed a hippocampus analysis method based on combination of 3D DenseNets and shape analysis for AD and MCI diagnosis using MR images. The DenseNets are constructed to learn the visual features of hippocampus region. The final image classification is made by combination of the 3D DenseNets and shape analysis. Experimental results and comparison on T1-weighted structural MR images from ADNI database demonstrate that the proposed method have promising performances for AD and MCI diagnosis.

References

[1] M. Prince, A. Wimo, M. Guerchet, G. C. Ali, Y. T. Wu, and M. Prina, "World Alzheimer report 2015. The global impact of dementia. An analysis of prevalence, incidence, cost and trends," Alzheimer's Dis. Int., London, U.K., Tech. Rep., 2015.

[2] J. CR, Jr. *et al.*, "Introduction to the recommendations from the National Institute on Aging—Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimers Dementia, J. Alzheimers Assoc.*, vol. 7, pp. 257–262, 2011.

[3] H. I. Suk, S. W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Funct.*, vol. 220, pp. 841–859, 2015.

[4] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, pp. 856–867, 2011.

[5] P. Cao *et al.*, "Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures," *Comput. Biol. Med.*, vol. 91, pp. 21–37, 2017.

[6] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, 2017.

[7] J. Zhang, M. Liu, and D. Shen, "Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4753–4764, Oct. 2017.

[8] M. Liu, J. Zhang, E. Adeli, and D. Shen, *Deep Multi-Task Multi-Channel Learning for Joint Classification and Regression of Brain Status*. New York, NY, USA: Springer, 2017.

[9] M. Liu, J. Zhang, D. Nie, P. T. Yap, and D. Shen, "Anatomical landmark based deep feature representation for MR images in brain disease diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1476–1485, Sep. 2018.

[10] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, "Ensembles of deep learning architectures for the early diagnosis of Alzheimer's disease," *Int. J. Neural Syst.*, vol. 26, 2016, Art. no. 1650025.

[11] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2017, pp. 835–838.

[12] O. Lindberg *et al.*, "Hippocampal shape analysis in Alzheimer's disease and frontotemporal lobar degeneration subtypes," *J. Alzheimers Dis.*, vol. 30, pp. 355–365, 2012.

[13] J. Lötjönen *et al.*, "Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease," *Neuroimage*, vol. 56, pp. 185–196, 2011.

[14] M. F. Beg, P. R. Raamana, S. Barbieri, and L. Wang, "Comparison of four shape features for detecting hippocampal shape changes in early Alzheimer's," *Statist. Methods Med. Res.*, vol. 22, pp. 439–462, 2013.

[15] K. K. Leung *et al.*, "Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease," *Neuroimage*, vol. 51, pp. 1345–1359, 2010.

[16] L. Xin, X. Hong, Z. Zhen, and L. Tong, "3D texture analysis of hippocampus based on MR images in patients with Alzheimer disease and mild cognitive impairment," *J. Beijing Univ. Technol.*, vol. 38, pp. 942–948, 2012.

[17] C. Platero and M. C. Tobar, "A fast approach for hippocampal segmentation from T1-MRI for predicting progression in Alzheimer's disease from elderly controls," *J. Neurosci. Methods*, vol. 270, pp. 61–75, 2016.
[18] M. Chupin *et al.*, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus*, vol. 19, pp. 579–587, 2009.
[19] X. Tang, D. Holland, A. M. Dale, L. Younes, and M. I. Miller, "Baseline shape diffeomorphometry patterns of subcortical and ventricular structures in predicting conversion of mild cognitive impairment to Alzheimer's disease," *J. Alzheimers Dis.*, vol. 44, pp. 599–611, 2015.
[20] H. C. Achterberg *et al.*, "Hippocampal shape is predictive for the development of dementia in a normal, elderly population," *Hum. Brain Mapping*, vol. 35, pp. 2359–2371, 2014.
[21] O. B. Ahmed, J. Benois-Pineau, M. Allard, C. B. Amar, and G. Catheline, "Classification of Alzheimer's disease subjects from MRI using hippocampal visual features," *Multimedia Tools Appl.*, vol. 74, pp. 1249–1266, 2015.
[22] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imag.*, vol. 17, no. 1, pp. 87–97, Feb. 1998.
[23] Y. Wang, J. Nie, P. T. Yap, F. Shi, L. Guo, and D. Shen, "Robust deformable-surface-based skull-stripping for large-scale studies," *Med. Image Comput. Comput. Assist. Interv.*, vol. 14, pp. 635–642, 2011.
[24] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, pp. 2261–2269, 2017.
[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
[27] R. Li *et al.*, "Deep learning based imaging data completion for improved brain disease diagnosis," *Med. Image Comput. Comput. Assist. Interv.*, vol. 17, pp. 305–312, 2014.
[28] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
[29] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.
[30] M. Liu, D. Zhang, and D. Shen, "Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1463–1474, Jun. 2016.