3-D CNN-Based Multichannel Contrastive Learning for Alzheimer's Disease Automatic Diagnosis

Jiaguang Li, Ying Wei, *Member, IEEE*, Chuyuan Wang, Qian Hu, Yue Liu, and Long Xu

***Abstract*— Alzheimer's disease (AD) is a common progressive neurodegenerative disease in the elderly. Mild cognitive impairment (MCI) is the symptomatic predementia stage of AD. Accurately distinguishing AD and MCI patients from normal people is the first step of the disease diagnosis. Several studies have demonstrated the potential of deep learning in the automatic diagnosis of AD and MCI using T1-weighted magnetic resonance imaging (MRI) images. In this article, we proposed an automatic classification method of AD versus normal control (NC) and MCI versus NC based on MRI images. This method used the 3-D convolutional neural network and took the whole 3-D MRI image as the input, which can obtain image information to the greatest extent. In addition, the multichannel contrastive learning strategy based on multiple data transformation methods (e.g., add noise) can combine the supervised classification loss with the unsupervised contrastive loss, which can further improve the classification accuracy and generalization ability of the network. To verify the effectiveness of our method, a large number of experiments were implemented on the ADNI dataset. The results show that our method can achieve excellent performance in accurate diagnosis of AD and MCI; the multichannel contrastive learning strategy can greatly improve the classification accuracy (AD versus NC: 4.19%; MCI versus NC: 4.57%) and generalization ability of the network.**

***Index Terms*— 3-D U-Net, Alzheimer's disease (AD), automatic diagnosis, contrastive learning, convolutional neural network (CNN), data transformation, deep learning, mild cognitive impairment (MCI), T1-weighted magnetic resonance imaging (MRI).**

I. Introduction

STUDYING brain function has become an important issue in neuroscience, especially for Alzheimer's disease (AD) [1]. Alzheimer's disease is recognized by the World Health Organization as a global public health priority [2], which mainly affects the elderly over 65 years old [3]. It is a progressive and degenerative disease of the brain that is the most common cause of dementia worldwide [4], [5] and is quickly becoming one of the most expensive, lethal, and burdening diseases of this century [6]. It starts with changes in the brain, which are unnoticeable to the person affected. Only after years of brain changes do individuals experience noticeable symptoms, such as memory loss and language problems [7]. Mild Cognitive Impairment (MCI) is the symptomatic predementia stage of AD, characterized by cognitive and functional impairment not severe enough to fulfill the criteria for dementia [8]. 10%–15% of MCI patients will progress to AD every year [9]. In the pathological spectrum of AD, MCI may be the optimal stage that clinical treatments or interventions can be effectively administered to prevent or delay the decline to severe dementia [10]. Accurately identification of AD and MCI from NC (normal control) is the first step in the disease diagnosis and treatment. To reduce the burden on doctors and improve the diagnosis efficiency, we proposed an automatic classification method of AD versus NC and MCI versus NC.

T1-weighted magnetic resonance imaging (MRI) is an established imaging technique for Alzheimer's disease diagnosis [11] due to its sensitivity to morphological changes caused by brain atrophy [12]. Fig. 1 shows the 3-D brain MRI images (after preprocessing) of patients at different stages of the disease. We can see the changes in the brain's anatomical structures with the development of the disease. Therefore, we used MRI images to study the computer-aided diagnosis (CAD) method of AD and MCI. We can also see from the figure that the differences between AD, MCI, and NC are not so obvious compared with natural scene images, which is one of the difficulties in AD and MCI diagnoses.

Some researchers have used deep learning methods to automatically identify AD, and MCI and achieved good performance [12]–[15]. The advantage of deep learning over traditional machine learning is that it does not need to design features manually. The deep learning method can automatically extract image features for classification tasks. Deep learning-based feature extraction methods are mainly divided into four categories: 2-D-slice-based [16]–[18], 3-D-patch-based [15], [19]–[21], ROI-based (region of interest) [22]–[24], and 3-D-overall-based [25]–[27]. The 2-D-slice-based method extracts the most informative 2-D slices from 3-D images for feature extraction, which can reduce the interference of huge back-

Manuscript received December 1, 2021; revised January 24, 2022; accepted March 13, 2022. Date of publication April 6, 2022; date of current version April 13, 2022. This work was supported in part by the National Nature Science Foundation of China under Grant 61871106 and Grant 61370152, in part by the Key Research and Development Projects of Liaoning Province, China, under Grant 2020JH2/10100029, and in part by the Open Project Program Foundation of the Key Laboratory of Opto-Electronics Information Processing, Chinese Academy of Sciences under Grant OEIP-O-202002. The Associate Editor coordinating the review process was David Ayllon. *(Corresponding author: Ying Wei.)*

Jiaguang Li, Chuyuan Wang, Qian Hu, and Long Xu are with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China (e-mail: 2190055@stu.neu.edu.cn; 1910312@stu.neu.edu.cn; 2010336@stu.neu.edu.cn; 1810342@stu.neu.edu.cn).

Ying Wei is with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China, and also with the Information Technology Research and Development Innovation Center, Peking University, Shaoxing 312035, China (e-mail: weiying@ise.neu.edu.cn).

Yue Liu is with the College of Information Science and Engineering, Northeastern University, Shenyang 110819, China, and also with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37240 USA (e-mail: 1610303@stu.neu.edu.cn).

Digital Object Identifier 10.1109/TIM.2022.3162265

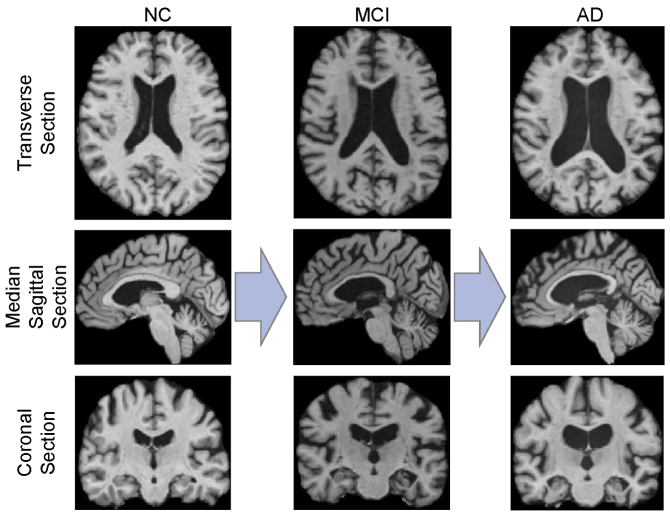


Fig. 1. 3-D brain MRI images of patients at different stages of the disease.

ground areas. However, this method will lose the relevance information between different slices. The 3-D-patch-based method cuts out several small 3-D patches from the original 3-D images and uses these patches to train the feature extractor. This method can reduce the amount of calculation and improve the training speed, but it loses the spatial information between patches. The ROI-based method is to segment-specific anatomical structures (such as gray matter, white matter, and hippocampus) from the original images and uses the segmentation results to train the network. This method makes the network pay more attention to the areas that have a great impact on the classification task. However, this method will lose information on other anatomical structures of the image. The 3-D-overall-based method can preserve the spatial information and detailed information of the image to the greatest extent. We believe that there is a greater potential to use the overall image as the input of the network.

In this article, we proposed multichannel contrastive learning based on the 3-D convolutional neural network (3-D CNN) for Alzheimer's disease automatic diagnosis. Our work is quite different from traditional self-supervised contrastive learning [28]–[30]. Self-supervised contrastive learning is for unsupervised network pretraining. It first makes two different transformations on large-scale unlabeled data and then obtains the supervision information according to the class invariance of the transformed data. The supervision information is used for network pretraining. The pretrained network is used for downstream tasks. Our proposed multichannel contrastive learning strategy aims to provide additional supervision information for supervised classification tasks, rather than unsupervised network pretraining. In addition, the proposed multichannel strategy can provide more supervision information than the two-channel strategy [31], [32].

The main contributions of this article are given as follows.

1) The multichannel contrastive learning strategy was added to the classification task of AD versus NC and MCI versus NC, which provided more additional supervision information for the task and greatly improved the classification accuracy and network generalization ability. We also analyzed the impact of the number of channels on the network performance.

2) 3-D CNN was used to extract image features, which can better obtain image detail information and spatial information.

3) We proposed a variety of data transformation methods, such as adding noise, histogram equalization, and edge sharpening. We also analyzed the impact of different data amplification methods on network performance.

4) A dynamic weighting factor was designed to balance supervised learning and contrastive learning.

5) We carried out a large number of experiments on the large-scale Alzheimer's Disease Neuroimaging (ADNI) [33] dataset. Experimental results show that our method can achieve excellent performance in AD and MCI diagnoses.

II. Related Works

In this section, we briefly review the previous studies on 3-D CNN-based Alzheimer's disease diagnosis methods using 3-D MRI images and self-supervised contrastive learning.

A. 3-D CNN-Based Alzheimer's Disease Diagnosis Methods

The advantage of the deep learning method over the traditional machine learning method is that it does not need to design features manually. Some studies have used CNN to extract image features for AD diagnosis and achieved good performance. CNN-based AD and MCI diagnosis methods can refer to [34].

When the 3-D MRI image is used as the input of the network, 3-D CNN can extract the correlation information between 2-D slices better than 2-D CNN. Recently, the 3-D CNN network has been used by a few researchers to extract the features from 3-D MRI images. For example, Fan *et al.* [27] proposed the 3-D U-Net network, which can achieve higher accuracy than other basic classification networks in AD diagnosis. Wang *et al.* [25] and Hedayati *et al.* [35] proposed an ensemble of 3-D convolutional networks for AD and MCI diagnoses. This method integrates the outputs of multiple 3-D CNN networks to obtain the final classification results. Korolev *et al.* [36] used residual and plain 3-D CNN architectures to skip feature extraction steps. Asl *et al.* [37] proposed 3-D deeply supervised adaptable CNN, which outperforms other CNN-based methods and conventional classifiers by accuracy and robustness.

We used 3-D CNN as the basic network in our method due to the excellent performance of the 3-D CNN network in AD diagnosis. To better obtain image features, we use the whole 3-D MRI image as the input of our network.

B. Self-Supervised Contrastive Learning

Self-supervised learning was proposed to learn visual features from large-scale unlabeled images or videos by constructing pretext tasks without using any human annotations [38]. As a popular solution, contrastive learning constructs positive samples and negative samples through two different data transformation methods. After the transformations, images from the same original image are regarded as positive

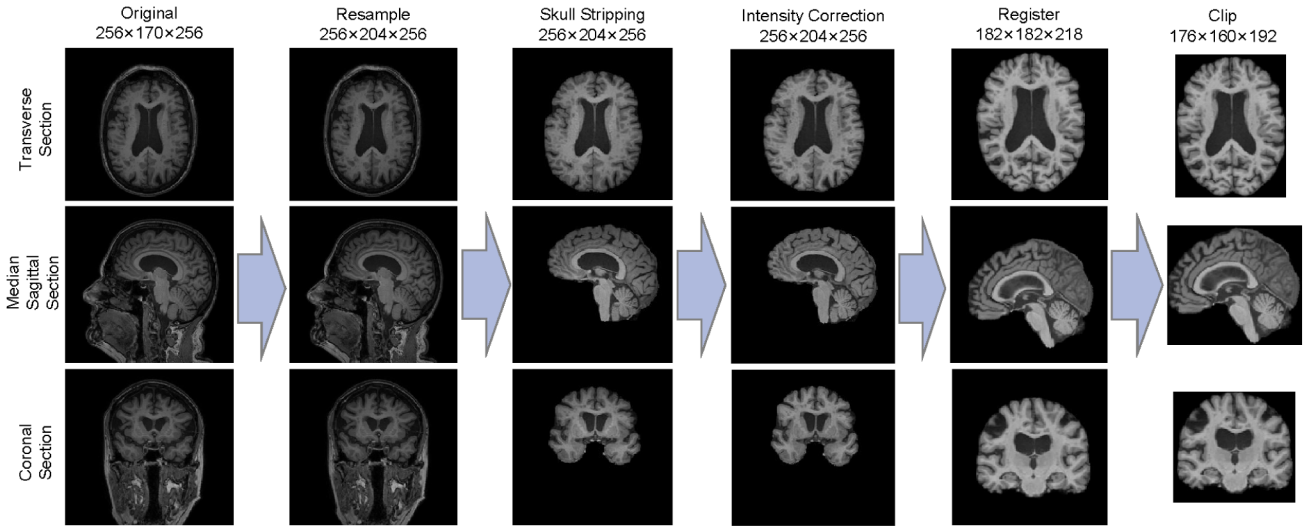


Fig. 2. Data preprocessing methods. The first row is the transverse section, the second row is the median sagittal section, and the third row is the coronal section. Each column represents the image after one processing step.

samples, whereas images from different original images are regarded as negative samples. Recently, Chen *et al.* [28] proposed the SimCLR, which trained the network by reducing the cosine distance between positive samples and increasing the cosine distance between negative samples. SimCLR achieved better performance than the previous self-supervised learning methods. However, it relies on large batch size and consumes huge computing resources. Wu *et al.* [39] constructed a memory bank to store output vectors of the network, which can greatly reduce the dependence of the network on the large batch size. He *et al.* [29] proposed the momentum contrast (MoCo) for unsupervised visual representational learning. They build a dynamic dictionary with a queue and a moving-averaged encoder to facilitate the training process. Chen *et al.* [30] proposed a simple Siamese (SimSiam) network and reported surprising empirical results that the network can learn meaningful representations even using none of the following: 1) negative samples; 2) large batch size; and 3) momentum encoders.

Inspired by SimSiam, our contrastive learning strategy does not need negative sample pairs or a large batch size. We only reduce the cosine distance between positive samples to improve the training results. In addition, multichannel contrastive learning can provide more additional supervision information than other contrastive learning methods, which is especially useful for small-scale data.

III. Materials

In this section, we introduce the dataset and data preprocessing methods in this work.

A. Dataset

We obtained the 3-D brain T1-weighted structure MRI images from the public ADNI datasets. ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease. Participants in ADNI are recruited across North America. We obtained 3746 MRI images from 928 subjects in the ADNI dataset. The 928 subjects are separated into 330 NC, 299 AD, and 299 MCI subjects. To prevent data leakage, we divided the data into the training set (60%), the validation set (20%), and the test set (20%) according to the subjects. Table I shows the demographic information of the subjects included in our datasets, including categories, the number of subjects, gender (male/female), age (mean $\pm$ standard deviation), and the number of MRI images.

TABLE I
Demographic Information of the Subjects Included in the Studied Datasets

Dataset	Category	Subjects	Gender	Age	Number
ADNI	NC	330	170 / 160	76.2 $\pm$ 6.2	1692
	MCI	299	110 / 189	76.4 $\pm$ 7.3	1115
	AD	299	135 / 164	75.8 $\pm$ 7.5	939

B. Data Preprocessing

Our data preprocessing method follows the following five steps.

1) Resample: Because the distance between 2-D slices in different MRI images is different, the resolution of different images is different. Therefore, the resampling method was used to make the distance between 2-D slices the same (1 mm) [40].

2) Skull Stripping: Since Alzheimer's disease will not change the skull, in order to prevent the skull from affecting the model training, the skull stripping technique [41] was used to turn the skull area into the background.

3) Intensity Correction: We used the intensity correction technique [42] to solve the gray inconsistency problem of MRI images caused by the uneven magnetic field.

4) Register: The MRI images were registered to the Montreal Neurological Institute standard brain template (MNI152, resolution of $1 \times 1 \times 1$ mm^3) [43], so we can ensure that the 3-D images have the same direction.

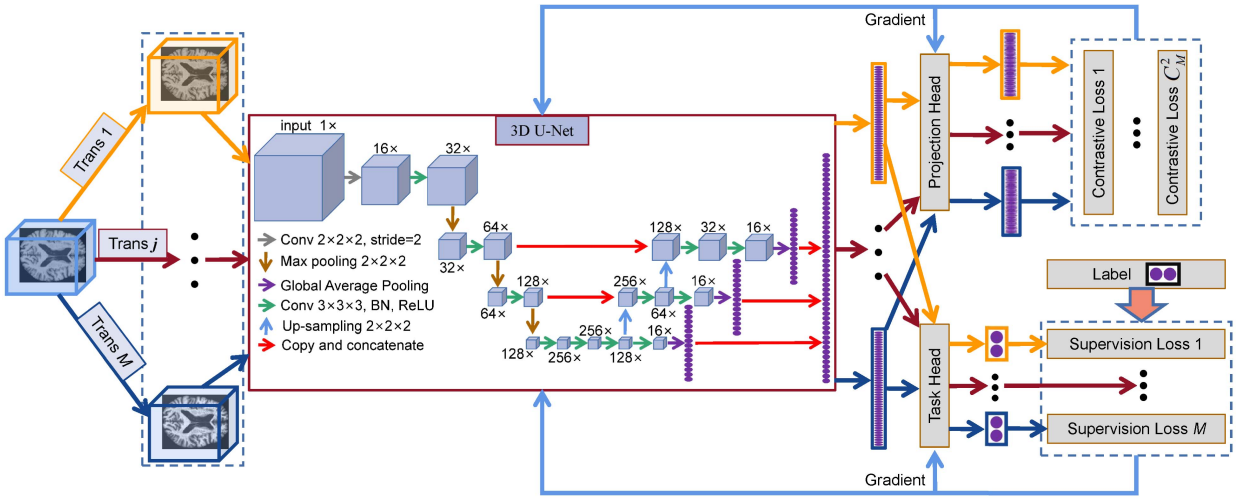


Fig. 3. Overall framework. The figure shows M channels. Each channel is represented by a separate color. Trans j represents the data transformation method used in the jth channel. The transformed images of all channels are used as the input of 3-D U-Net. The output vectors of 3-D U-Net are used as the input of Task Head and Projection Head. The output vectors of Task Head and Projection Head are used to calculate the supervision loss and contrastive loss respectively. The supervision loss is used to update the parameters of the 3-D U-Net and Task Head. The contrastive loss is used to update the parameters of the 3-D U-Net and Projection Head. Gradient means that the Adam optimization algorithm is used to update the network parameters θ, α, and β.

5) Clip: The images were clipped to remove slices that contain only the background.

To intuitively understand the preprocessing process, we give the images of the process in Fig. 2. Finally, we normalize the images according to (1) so that the pixels value are distributed between 0 and 1 [27]

$$\text{Nor} = \frac{\text{OI} - \min(\text{OI})}{\max(\text{OI}) - \min(\text{OI})} \tag{1}$$

where OI represents the set of all the pixels value of an original image and Nor represents the set of all the pixels value of the image after normalization.

IV. Methods

In this section, we introduce in detail our proposed multichannel contrastive learning method based on a 3-D U-Net network. Section A introduces the overall framework, the supervised learning, the contrastive learning, the loss function, and the 3-D U-Net structure. Section B describes the data transformation methods.

A. Overall Framework

Fig. 3 summarizes the overall framework of our method, including the supervised learning strategy, the multichannel contrastive learning strategy, and the 3-D U-Net network structure. First, we take N preprocessed 3-D MRI images as the original inputs of the framework. Second, M data transformation methods are used to transform the input images, so $N \times M$ transformed images are obtained. Third, the $N \times M$ images are input into our 3-D U-Net, respectively, and then, the corresponding $N \times M$ network output vectors are obtained. Let $x_{i,j}$ represent the jth transformed image from the ith original input image, where $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, M\}$. Thus, $F_\theta(x_{i,j})$ represents the corresponding output vector of the 3-D U-Net, where F represents the 3-D U-Net and θ represents the network parameters. These vectors will be used for supervised learning and contrastive learning.

1) Supervised Learning: Input these $N \times M$ vectors into the Task Head network to obtain $N \times M$ 2-D output vectors. Th Task Head network consists of a dropout layer, a full connection layer, and a softmax layer [see Fig. 4(a)]. The input dimension of Task Head is 48. Let $[T_\alpha(F_\theta(x_{i,j})), 1 - T_\alpha(F_\theta(x_{i,j}))]$ represent the 2-D output vectors and $[y_i, 1 - y_i]$ represent the corresponding labels, where T represents the Task Head network and α represents the network parameters. Then, we can calculate the supervised cross-entropy loss according to (2)

$$\begin{aligned} &L_{\text{sup}}(\theta, \alpha) \\ &= -\frac{1}{N \times M} \sum_{i=1}^{N} \sum_{j=1}^{M} \big(y_i \log T_\alpha\big(F_\theta\big(x_{i,j}\big)\big) \\ &\qquad + (1 - y_i) \log\big(1 - T_\alpha\big(F_\theta\big(x_{i,j}\big)\big)\big)\big). \end{aligned} \tag{2}$$

2) Contrastive Learning: Take the $N \times M$ output vectors of 3-D U-Net as the input of Projection Head; then, we can get $N \times M$ 128-D output vectors. The Projection Head is a multilayer perceptron (MLP) network [see Fig. 4(b)]. The input dimension of the Projection Head is 48, and the output dimension is 128. These 128-D vectors are used to calculate the contrastive loss according to (3)

$$\begin{aligned} &L_{\text{con}}(\theta, \beta) \\ &= -\frac{1}{N \times C_M^2} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{k=1}^{M} \big(I(j,k) \\ &\qquad \times \text{sim}\big(P_\beta\big(F_\theta\big(x_{i,j}\big)\big), P_\beta\big(F_\theta\big(x_{i,k}\big)\big)\big)\big) \end{aligned} \tag{3}$$

where P represents the Projection Head, β represents the parameters of the Projection Head, $P_\beta(F_\theta(x_{i,j}))$ represents the 128-D output vectors, C_M^2 represents the number of combinations that can be obtained by pairwise combination in M

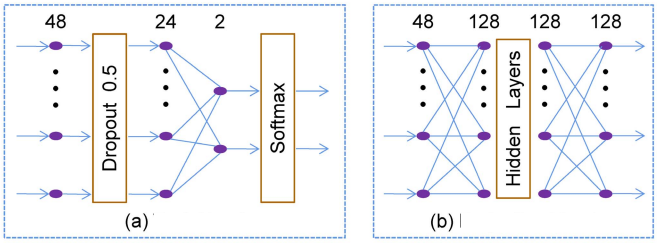


Fig. 4. Task head network and projection head network. (a) Task Head. (b) Projection Head.

samples, $I(j,k)$ is defined as follows:

$$I(j,k) = \begin{cases} 1, & j < k \\ 0, & j \geq k \end{cases} \tag{4}$$

and the sim function in (3) is used to calculate the cosine similarity between two vectors, which is defined as follows:

$$\text{sim}(a,b) = \frac{a^T b}{\mid a \mid \cdot \mid b \mid}. \tag{5}$$

3) Loss Function and Network Optimization: The final loss function consists of supervision loss and contrastive loss, which is defined as follows:

$$L(\theta,\alpha,\beta) = L_{\text{sup}}(\theta,\alpha) + \lambda L_{\text{con}}(\theta,\beta) \tag{6}$$

where λ is used to balance L_{sup} and L_{con}. We found that, in the initial stage of training, the contrastive loss is much less than the supervision loss (10^{-3} s and 10^{-1} s). This will cause the network to ignore the contrastive loss. Therefore, the weight factor λ should be set to a larger value to improve the network's response to contrastive loss. However, in the initial stage of training, the network does not have good classification ability. If the weight of contrastive loss is too large, the network will converge in an abnormal direction. The correct approach should first make the network converge in the right direction. With the increase in training epochs, the classification ability of the network will gradually improve. Then, we can increase the weight of contrastive loss to fine-tune the convergence direction of the network. Therefore, we designed a dynamic weighting factor that varies with epochs, which is defined as follows:

$$\lambda = 100 \times \frac{\text{Epoch}}{\text{Total Epochs}}. \tag{7}$$

Finally, the U-Net model parameters θ, the Task Head parameters α, and the Projection Head parameters β can be estimated by solving the following optimization problem:

$$\theta^*, \alpha^*, \beta^* = \arg\min_{\theta,\alpha,\beta} L(\theta,\alpha,\beta). \tag{8}$$

The adaptive moment estimation (Adam) optimization algorithm [44] was used to solve this optimization problem.

During the testing period, we will discard Projection Head and use 3-D U-Net and Task Head to predict the test data. The output of Task Head is the final prediction result.

4) The 3-D U-Net Network: The U-Net [45] network is a kind of CNN. 3-D U-Net has been proven to be able to achieve excellent performance in AD and MCI automatic diagnoses [27], so we used 3-D U-Net as the basic network of our framework. We have tried several CNNs (i.e., 3-D ResNet), but the results are not as good as 3-D U-Net (selecting the basic network is not the focus of this article, so we do not discuss it in depth). Our 3-D U-Net structure is shown in Fig. 3. First, we do the convolution with the kernel of $2 \times 2 \times 2$ and stride of 2 for the input image to reduce the image volume to $1/8$ and reduce the amount of calculation. We combine a $3 \times 3 \times 3$ convolution layer, a batch normalization (BN) layer, a rectified linear unit (ReLU) layer, and a $2 \times 2 \times 2$ max-pooling layer as the encoder structure to extract image features. The decoder structure includes a $3 \times 3 \times 3$ convolution layer, a BN layer, an ReLU layer, and a $2 \times 2 \times 2$ upsampling layer. The copy and concatenate layers between encoder and decoder can fuse deep features with shallow features. We use convolution, BN, and ReLU to make the outputs of different scales in the decoding process have 16 channels. Finally, the global average pooling layer is used to make the outputs of different scales become 16-D vectors. These 16-D vectors are concatenated to form a new 48-D vector, which is the final output of the network.

B. Data Transformation

This section introduces the data transformation methods that we designed in detail, including histogram equalization, sharpening, smoothing, nonlinear transformation, gamma transformation, clipping and resizing, adding noise, and flipping. Fig. 5 shows the transformed images.

1) Histogram Equalization: Histogram equalization can enhance the image contrast ratio. We first multiply the pixels value of the image by b and round it to make the pixels value be in $\{0,\ldots,b\}$ because the background occupies a large number of pixels, which will have a bad impact on equalization. Therefore, we only equalize the pixels with values in $\{1,\ldots,b\}$. Here, we set b to 255. Histogram equalization follows the following formula:

$$y_i = \frac{b-1}{A} \sum_{j=1}^{i} n_j \tag{9}$$

where $i \in \{1,\ldots,b\}$ represents pixel value, A represents the total number of pixels in the image, n_j represents the number of pixels with value j, $j \in \{1,\ldots,i\}$, and y_i represents the transformed pixels value of pixels whose value is i before the transformation. Finally, we divide the image by 255 so that the pixels' values are distributed in [0, 1].

2) Sharpening and Smoothing: Sharpening can make the image edge clearer, and smoothing can reduce noise interference. We use the fixed $3 \times 3 \times 3$ convolution kernel to convolute the original image for sharpening or smoothing. The convolution kernel is shown in Fig. 6(a). Here, we set a in Fig. 6(a) to 0.25 for smoothing and -0.25 for sharpening.

3) Nonlinear Transformation: The nonlinear transformation method that we designed can improve the contrast ratio

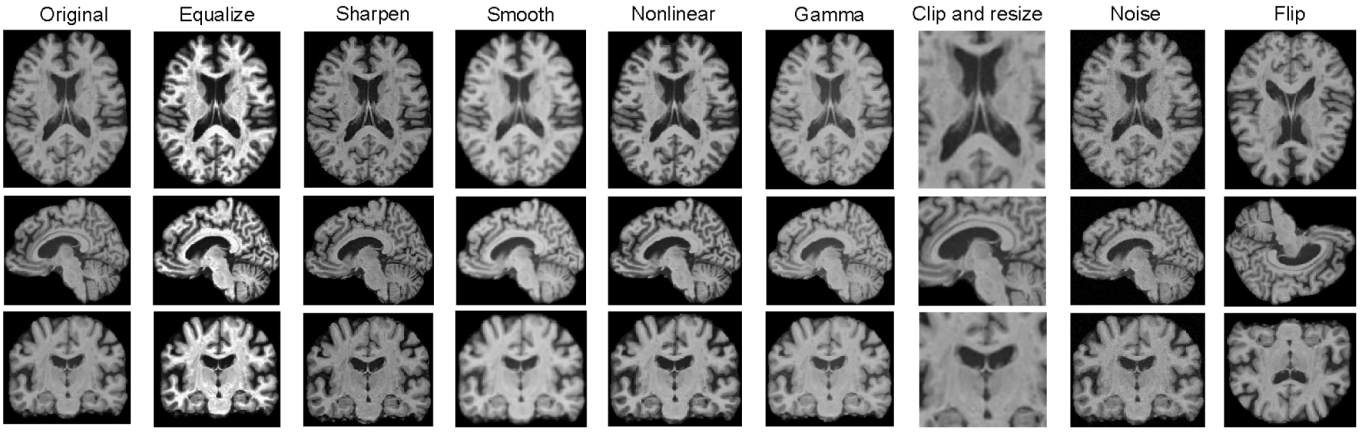


Fig. 5. Data transformation methods. The first row is the transverse section, the second row is the median sagittal section, and the third row is the coronal section. Each column represents the image transformed by a kind of data transformation method.

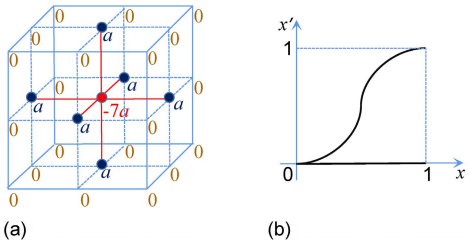


Fig. 6. Fixed convolution kernel for sharpening and smoothing, and the nonlinear transformation method. (a) Convolution kernel. (b) Nonlinear transform.

between gray matter and white matter in brain images. The nonlinear transformation follows:

$$x' = \begin{cases} (2x-1)^c \times 0.5 + 0.5, & x \geq 0.5 \\ (2x-1)^{\frac{1}{c}} \times 0.5 + 0.5, & x < 0.5 \end{cases} \tag{10}$$

where x and x' are pixel values before and after transformation and c is used to adjust the intensity of the transformation. Here, we set c to 0.75. The transformation formula is visualized in Fig. 6(b).

4) Gamma Transformation: Because most of the images are a little dark, we use gamma transformation to improve the brightness of the images. Gamma transformation is defined as

$$x' = x^{\gamma} \tag{11}$$

where x and x' are pixel values before and after transformation and γ is used to adjust the intensity of the transformation. Here, we set γ to 0.75.

5) Clipping and Resizing: We cut the image with the size of 88 × 80 × 96 from the central area of the original image. Then, the image size is restored to 176 × 160 × 192 by trilinear interpolation.

6) Adding Noise and Flipping: Adding Gaussian noise with a mean of 0 and variance of 0.0008 to the image can be used as a data transformation method. Random flipping along different dimensions of the image can also be a way of data transformation.

V. Experiments and Analyses

In this section, we implemented a large number of experiments to verify the effectiveness of the important components of our framework, including the contrastive learning strategy, the multichannel strategy, the data transformation methods, and the Projection Head. We also analyzed the influence of the weight factor λ between the supervision loss and the contrastive loss on the performance of the framework. Besides, we used Grad-CAM [46] to visualize the feature map learned by our model. Then, we analyzed the convergence of the training. Finally, the state-of-the-art methods are compared with ours.

A. Experimental Settings

Our method was validated on both tasks of AD diagnosis (i.e., AD versus NC) and MCI diagnosis (i.e., MCI versus NC). The training process stopped when the number of epochs reached 80. We gave the initial learning rate of 0.001. The learning rate decay follows: $0.001 \times 0.97^{\text{epoch}}$, where epoch $\in \{0, \ldots, 79\}$. We used the He initialization [47] method to initialize network parameters. The training batch size was eight. The number of hidden layers in Projection Head was set to 0, and the weight factor λ between L_{sup} and L_{con} was set to $1.25 \times$ epoch. We used a threefold cross-validation method in our experiments (except the Grad-CAM visualization experiment), and the average values were displayed. For the Grad-CAM visualization experiment, we randomly selected one result from the threefold cross-validation as the final result. In the test period, the data transformation methods and the Projection Head network did not work, and the number of channels was one. The input images did not make other changes after preprocessing during testing. The output of Task Head was the final output. The experiments were conducted on four NVIDIA RTX 2080Ti GPUs.

B. Evaluation Metrics

We used accuracy (ACC), sensitivity (SEN), specificity (SPC), precision (PRE), area under receiver operating characteristic (ROC) curve (AUC), and F1 score to evaluate the classification performance of this method. These metrics are defined as ACC = (TP + TN/TP + TN + FP + FN), SEN = (TP/TP+FN), SPC = (TN/TN+FP), PRE = (TP/TP+FP), and $F1 = (2 \cdot \text{SEN} \cdot \text{PRE}/\text{SEN} + \text{PRE})$, where TP, TN, FP,

TABLE II

COMPARISON BETWEEN SUPERVISED LEARNING ONLY (BASELINE), SUPERVISED LEARNING WITH DATA AUGMENTATION (BASE+TRANS), AND SUPERVISED LEARNING WITH MULTICHANNEL CONTRASTIVE LEARNING (OURS) (%)

Task	Method	ACC	SEN	SPC	PRE	AUC	F1	Mean improve
AD vs. NC	Baseline	90.87	**98.53**	76.88	88.62	87.71	93.31	
	Base+Trans	92.02 (+1.15)	96.47 (-2.06)	83.87 (+6.99)	91.62 (+3.00)	90.17 (+2.46)	93.98 (+0.67)	(+2.04)
	Ours	**93.16** (+2.29)	95.00 (-3.53)	**89.78** (+12.9)	**94.44** (+5.82)	**92.39** (+4.68)	**94.72** (+1.41)	(+3.93)
MCI vs. NC	Baseline	77.38	63.79	**86.19**	**75.00**	74.99	68.94	
	Base+Trans	79.16 (+1.78)	75.45 (+11.66)	81.65 (-4.54)	73.45 (-1.55)	78.55 (+3.56)	74.44 (+5.50)	(+2.74)
	Ours	**80.44** (+3.06)	**83.18** (+19.39)	78.59 (-7.60)	72.33 (-2.67)	**80.89** (+5.90)	**77.38** (+8.44)	(+4.42)

and FN represent the true positive, true negative, false positive, and false negative values, respectively.

C. Ablation Experiments

In this experiment, we compared supervised learning only (Baseline), supervised learning with data transformation (Base+Trans), and supervised learning with multichannel contrastive learning (Ours). For both Ours and Base+Trans methods, the number of channels was set to 2. The original images were directly used as the input of the first channel, whereas the input images of the second channel were both histograms equalized and randomly flipped. There was only one channel for the Baseline method, and the original images were used as the input. For Baseline and Base+Trans methods, Projection Head had no effect, and only the supervision loss L_{sup} was used as the loss function. The results were summarized in Table II. The red brackets in the table represent the improvement compared with the Baseline, whereas the black brackets represent the regression. The last column represents the average improvement of the six evaluation metrics.

It can be seen from Table II that, for our method, the most important evaluation metrics ACC, AUC, and F1 increased significantly compared with the Baseline and Base+Trans methods, both in AD versus NC and MCI versus NC tasks. This shows that the contrastive learning strategy can, indeed, provide additional supervision information for the classification task, so as to improve the performance of AD and MCI automatic diagnoses. Although SEN decreased a little in AD diagnosis, and SPC and PRE decreased a little in MCI diagnosis, it was insignificant compared with the increase in other evaluation metrics.

D. Effects of the Number of Channels

This experiment discussed the influence of the number of channels of the model on classification performance. We took the original images without any transformation as the input of the first channel. The data transformation methods used in this experiment included histogram equalization, sharpening, adding noise, and flipping. Each original image went through these four transformations. We randomly selected one of the four transformed images as the input of the jth channel (duplicate selection was not allowed), where $j \in \{2, \ldots, M\}$ and M was the number of channels. With the increase in the number of channels, the amount of data participating in network training also increased, which will greatly increase the training time. Therefore, we set M to 5 here. Fig. 7 shows the change in model classification performance as the number

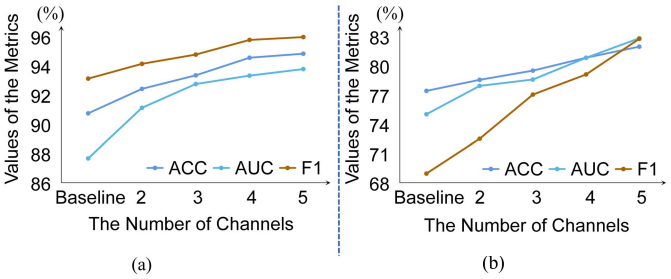


Fig. 7. Effects of the number of channels. The abscissa represents the number of channels, and the ordinate represents the value of the evaluation metrics (%). (a) AD vs. NC. (b) MCI vs. NC.

of channels increases. Here, three important evaluation metrics were used to evaluate the performance of the model, including ACC, AUC, and F1 scores. As can be seen from Fig. 7, as the number of channels increases, the performance of the model for both AD and MCI classification tasks is improving. For the case where the number of channels is 5, the ACC, AUC, and F1 metrics in AD diagnosis achieved 95.06%, 93.98%, and 96.23% respectively, and improved by 4.19%, 6.27%, and 2.92% compared to the Baseline. In the diagnosis of MCI, these three metrics achieved 81.90%, 82.76%, and 82.70%, respectively, and improved by 4.52%, 7.77%, and 13.76% compared to the Baseline. Therefore, we can conclude that the multichannel strategy that we designed can provide more supervision information for the model than the dual-channel strategy and further improve the classification ability of the model.

E. Effects of Different Data Transformation Methods

We analyzed the performance of the eight data transformation methods shown in Fig. 5 through the AD versus NC classification task. In this experiment, the number of channels of the model was fixed at 2. The input of the first channel was the original images without any transformation. The input images of the second channel were transformed by one of the above eight data transformation methods. Table III compared the classification performance when different data transformation methods were used in the second channel. We selected the representative ACC, AUC, and F1 as the evaluation metrics of this experiment. As can be seen from the table: 1) histogram equalization is the most important data transformation method; 2) sharpening and flipping can also greatly improve the performance of the model; 3) gamma transformation, adding noise, nonlinear transformation, and smoothing can partly improve the AD diagnosis ability of the model; and 4) however, the clipping and resizing method

TABLE III
PERFORMANCE OF DIFFERENT DATA AMPLIFICATION METHODS (%)

Transformation	ACC	AUC	F1	Mean improve
Baseline	90.87	87.71	93.31	
Flip	92.78 (+1.91)	91.25 (+3.54)	94.52 (+1.21)	(+2.22)
Sharpen	93.16 (+2.29)	92.39 (+4.86)	94.72 (+1.41)	(+2.79)
Gamma	91.83 (+0.96)	89.66 (+1.95)	93.88 (+0.57)	(+1.16)
Noise	91.44 (+0.57)	89.85 (+2.14)	93.51 (+0.20)	(+0.97)
Equalize	**93.92** (+3.05)	**92.62** (+4.91)	**95.38** (+2.07)	(+3.34)
Clip and resize	89.16 (-1.71)	89.55 (+1.84)	91.32 (-1.99)	(-0.62)
Nonlinear	91.25 (+0.38)	88.73 (+1.02)	93.5 (+0.19)	(+0.53)
Smooth	91.73 (+0.86)	90.01 (+2.30)	93.75 (+0.44)	(+1.20)

TABLE IV
EFFECTS OF THE WEIGHT FACTOR BETWEEN SUPERVISION LOSS AND CONTRASTIVE LOSS (%)

Weight Factor λ	ACC	AUC	F1	Mean
1	92.97	95.29	88.71	92.32
10	93.35	96.76	87.10	92.40
100	94.11	95.59	**89.60**	93.10
$10 * \frac{Epoch}{Total\ Epochs}$	94.30	97.94	87.63	93.29
$100 * \frac{Epoch}{Total\ Epochs}$	**94.49**	**98.24**	87.63	**93.45**
$10 * (\frac{Epoch}{Total\ Epochs})^2$	93.83	97.65	86.83	92.77
$100 * (\frac{Epoch}{Total\ Epochs})^2$	93.35	95.88	88.71	92.65

has a bad effect on the performance of the model; we think that this is because this method leads to the loss of a large amount of information helpful for classification (e.g., some areas containing gray matter and white matter).

F. Effects of the Weight Factor Between Supervision Loss and Contrastive Loss

In this experiment, we analyzed the impact of weight factors λ between the supervision loss and the contrastive loss on network performance through the AD versus NC classification task. The number of channels was set to 4. The first channel used the original images, the second channel used histogram equalization, the third channel used sharpening, and the fourth channel used flipping. Table IV compares the performance of the model with different weight factors λ. It can be seen from the table that the network performance is the best when $\lambda = 100 * (\text{Epoch/Total Epochs})$. This proves that the dynamic weight factor changing with the epoch is better than the fixed weight factor. In the initial stage of training, there should be a small λ to make the network converge in the right direction. In the later stage of training, there should be a large λ to fine-tune the convergence direction of the network.

G. Effects of the Projection Head

In this experiment, we analyzed the effects of the number of hidden layers of Projection Head through the AD versus NC classification task. The number of channels was set to 4. The first channel used the original images, the second channel used histogram equalization, the third channel used sharpening, and the fourth channel used flipping. We compared four cases: without Projection Head and Projection Head containing zero, one, and two hidden layers. The results were shown in Table V. As can be seen from the table, the best results can be obtained when Projection Head contains no hidden layer.

TABLE V
EFFECTS OF THE PROJECTION HEAD (%)

Projection Head	ACC	AUC	F1	Mean
No Projection Head	93.92	92.37	95.40	93.90
0 Hidden Layer	**94.49**	**92.93**	**95.84**	**94.42**
1 Hidden Layer	93.73	92.59	95.21	93.84
2 Hidden Layers	93.16	91.17	94.87	93.07

H. Visualization

To display the effectiveness of our method in feature extraction, we provided visualizations of the features learned by the Baseline method and our method with five channels using Grad-CAM in Fig. 8. The color indicates the importance of regions for classification. As can be seen, for both AD diagnosis and MCI diagnosis, our method can pay more attention to areas with a large amount of information and learn more useful features. This shows that our method can provide more supervision information for feature extraction, so as to improve the performance of the model.

I. Convergence Analysis

In order to intuitively explain the fitting performance of our model and the convergence performance of training, we show the loss graphs for both training and validation in Fig. 9. The experimental settings are the same as those in ablation experiments (see Section V-C). It can be seen from the figure that, for both AD and MCI diagnoses, the training loss of our method and the Base+Trans method converges faster before the 30th epoch, which shows that our method and the Base+Trans method can fit the training data more quickly. In addition, the training loss of the three methods finally converges to 0, indicating that the three methods can perfectly fit the training data. For the validation loss, our method can converge to a smaller value after the 30th epoch, which shows that our method can better fit the validation data. A common problem in this field is that the convergence performance of validation loss is far worse than that of training loss. How to solve this overfitting problem is the direction of our future efforts.

J. Comparison Results With State-of-the-Art Methods

The proposed model is compared with the state-of-the-art algorithms using MRI images in Table VI. At present, the number of studies on AD versus NC classification is far more than that on MCI versus NC classification. Due to

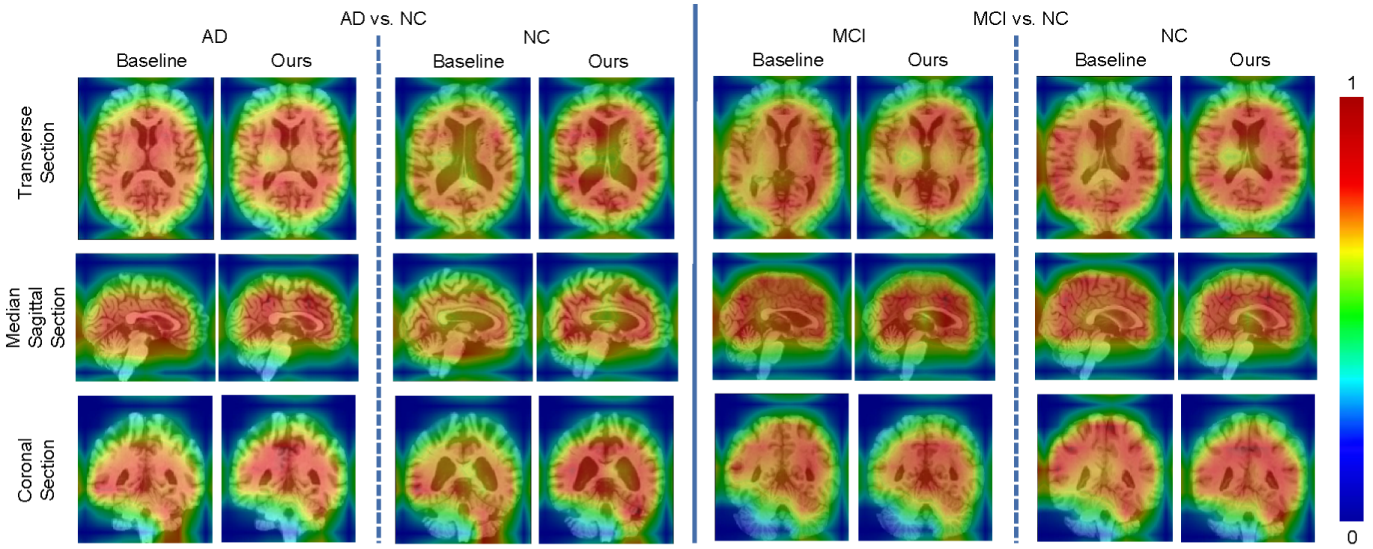


Fig. 8. Grad-CAM visualizations of the features learned by the Baseline method and our method. The first and second columns are an AD image in the AD versus NC classification task, the third and fourth columns are an NC image in the AD versus NC classification task, the fifth and sixth columns are an MCI image in the MCI versus NC classification task, and the seventh and eighth columns are an NC image in the MCI versus NC classification task.

TABLE VI

COMPARISON RESULTS WITH STATE-OF-THE-ART METHODS (%)

Method	Year	Data (Subject)	AD vs. NC			MCI vs. NC		
			ACC	AUC	F1	ACC	AUC	F1
C. Lian *et al.* [12]	2020	ADNI (787)	90.00	95.00				
M. Liu *et al.* [15]	2018	ADNI (787)	90.56	95.74	89.10			
M. Liu *et al.* [19]	2018	ADNI (787)	91.09	95.86	89.74			
W. Lin *et al.* [26]	2021	ADNI (670)	82.00	81.76				
M. Liu *et al.* [48]	2016	ADNI (225)	93.83					
K. Kwak *et al.* [49]	2018	ADNI (319)	94.17			85.89		
C. Feng *et al.* [50]	2018	ADNI (319)	94.29	96.23		64.47	67.14	
Z. Zhang *et al.* [51]	2021	ADNI (1139)	92.00	96.20	90.90	80.10	85.00	78.60
J. H. So *et al.* [52]	2019	ADNI (150)	85.00			75.00		
X. Jiang *et al.* [53]	2020	modularbrain [54] (91)				76.89		
Ours		ADNI (928)	95.06	93.98	96.23	81.90	82.76	82.70

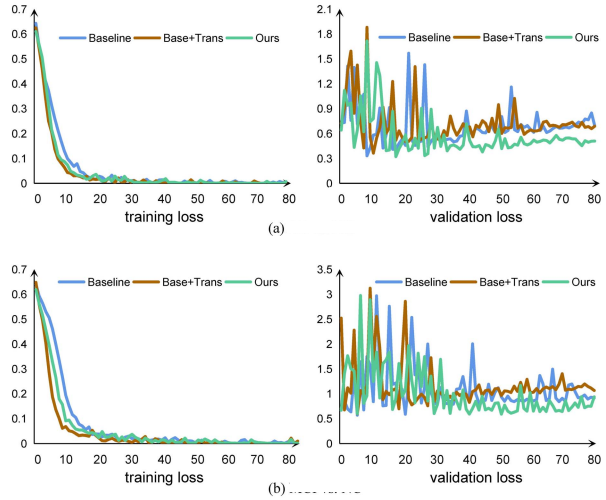


Fig. 9. Convergence analysis. The left two graphs are the training loss, the right two graphs are the validation loss, the top two graphs are the AD diagnosis, and the bottom two graphs are the MCI diagnosis. The ordinate of each graph is the loss value, and the abscissa is the epoch number. (a) AD vs. NC. (b) MCI vs. NC.

the utilization of different data, the results obtained by these methods should not be compared directly. These methods can only be used as rough references. Some methods report better performance than our method, but the results obtained from the small-scale data they use are not representative. In other words, small-scale data are more likely to cause overfitting of the network. This overfitting phenomenon will enable the network to obtain higher performance on the small-scale data, but it is difficult to generalize the network to other large-scale data. The data that we use have a large scale (928 subjects and 3746 MRI images), which is more challenging, but the results are more representative. It can be seen from the table that, although we used challenging large-scale data, the performance was still higher than most other studies. This proves that our proposed method is competent for the automatic diagnosis of AD and MCI.

VI. CONCLUSION

This article proposed an automatic diagnosis method of AD and MCI using T1-weighted MRI images. This method used a 3-D U-Net network and takes the whole 3-D MRI image as the input, which can obtain image features to the greatest extent. This method combined supervised learning with unsupervised contrastive learning, which can provide additional supervision information for the network, so as to improve the performance of the network. What is more, for

the contrastive learning part of the method, the multichannel strategy can provide more additional supervision information for the network than the dual-channel strategy. We designed a variety of data transformation methods and compared the performance of different methods. The experiment results show that histogram equalization, sharpening, and flipping can greatly improve the network performance; gamma transformation, adding noise, nonlinear transformation, and smoothing can partly improve the network performance; and clipping and resizing cannot improve the network performance. In addition, the performance of the dynamic weight factor λ between the supervision loss and the contrastive loss is better than that of the fixed weight factor.

The traditional self-supervised contrastive learning benefits from a large amount of data (millions). However, a large amount of medical image data is difficult to obtain. The proposed multichannel strategy can provide a new idea for the application of self-supervised contrastive learning on small datasets because it can mine more supervision information from the data itself.

REFERENCES

[1] L.-D. Liao *et al.*, "A novel 16-channel wireless system for electroencephalography measurements with dry spring-loaded sensors," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1545–1555, Jun. 2014.

[2] C. A. Lane, J. Hardy, and J. M. Schott, "Alzheimer's disease," *Eur. J. Neurol.*, vol. 25, pp. 59–70, Jan. 2018.

[3] C. L. Masters, R. Bateman, K. Blennow, C. C. Rowe, R. A. Sperling, and J. L. Cummings, "Alzheimer's disease," *Nature Rev. Disease Primers*, vol. 1, p. 15056, Oct. 2015.

[4] M. S. Kamal, A. Northcote, L. Chowdhury, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–7, 2021.

[5] C. Trigona *et al.*, "Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 4, pp. 971–980, Apr. 2018.

[6] P. Scheltens *et al.*, "Alzheimer's disease," *Lancet*, vol. 397, pp. 1577–1590, Apr. 2021.

[7] J. Wiley, "Alzheimer's disease facts and figures," *Alzheimers Dement*, vol. 17, pp. 327–406, Mar. 2021.

[8] F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 17–26, Jan. 2020.

[9] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1077–1089, Apr. 2016.

[10] H. T. Shen *et al.*, "Heterogeneous data fusion for predicting mild cognitive impairment conversion," *Inf. Fusion*, vol. 66, pp. 54–63, Feb. 2021.

[11] P. Scheltens *et al.*, "Alzheimer's disease," *Lancet*, vol. 388, pp. 505–517, Aug. 2016.

[12] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[13] X. Zhu, H. Suk, S. Lee, and D. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 607–618, Mar. 2015.

[14] X. Zhu, H. I. Suk, L. Wang, S.-W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Med. Image Anal.*, vol. 38, no. 3, pp. 205–214, May 2017.

[15] 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 Inform.*, vol. 22, no. 5, pp. 1476–1485, Sep. 2018.

[16] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105242.

[17] A. Valliani and A. Soni, "Deep residual nets for improved Alzheimer's diagnosis," in *Proc. 8th ACM Int. Conf. Bioinf., Comput. Biol., Health Informat.*, Aug. 2017, p. 615.

[18] M. Hon and N. M. Khan, "Towards Alzheimer's disease classification through transfer learning," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 1166–1169.

[19] 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, Jan. 2018.

[20] B. Shi, Y. Chen, P. Zhang, C. D. Smith, and J. Liu, "Nonlinear feature transformation and deep fusion for Alzheimer's disease staging analysis," *Pattern Recognit.*, vol. 63, pp. 487–498, Mar. 2017.

[21] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, and S. Li, "Prediction of Alzheimer's disease progression with multi-information generative adversarial network," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 711–719, Mar. 2021.

[22] R. Cui and M. Liu, "Hippocampus analysis based on 3D CNN for Alzheimer's disease diagnosis," *Proc. SPIE*, vol. 10806, Aug. 2018, Art. no. 108065O.

[23] N. Amoroso *et al.*, "Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge," *J. Neurosci. Methods*, vol. 302, pp. 3–9, May 2018.

[24] K. Aderghal, J. Benois-Pineau, and K. Afdel, "Classification of sMRI for Alzheimer's disease diagnosis with CNN: Single Siamese networks with 2D+ε approach and fusion on ADNI," in *Proc. ACM Int. Conf. Multimedia Retr.*, Jun. 2017, pp. 494–498.

[25] H. Wang *et al.*, "Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease," *Neurocomputing*, vol. 333, pp. 145–156, Mar. 2019.

[26] W. Lin *et al.*, "Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease," *Frontiers Neurosci.*, vol. 15, p. 357, Apr. 2021.

[27] Z. Fan *et al.*, "U-Net based analysis of MRI for Alzheimer's disease diagnosis," *Neural Comput. Appl.*, vol. 33, pp. 13587–13599, Apr. 2021.

[28] T. Chen, S. Kornblith, M. Norouzi, and GA. Hinton, "Simple framework for contrastive learning of visual representations," *Proc. 37th Int. Conf. Mach. Learn.*, vol. 119, 2020, pp. 1597–1607.

[29] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[30] X. Chen and K. He, "Exploring simple Siamese representation learning," 2020, *arXiv:2011.10566*.

[31] J. Li *et al.*, "Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19," *Pattern Recognit.*, vol. 114, Jun. 2021, Art. no. 107848.

[32] X. Chen, L. Yao, T. Zhou, J. Dong, and Y. Zhang, "Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107826.

[33] *ADNI*. Accessed: Nov. 1, 2021. [Online]. Available: http://adni.loni.usc.edu/

[34] J. Wen *et al.*, "Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101694.

[35] R. Hedayati, M. Khedmati, and M. Taghipour-Gorjikolaie, "Deep feature extraction method based on ensemble of convolutional auto encoders: Application to Alzheimer's disease diagnosis," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102397.

[36] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 835–838.

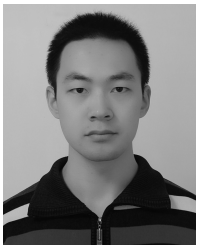
[37] E. H. Asl *et al.*, "Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network," *Frontiers Biosci.*, vol. 23, no. 3, pp. 584–596, 2018.

[38] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021.

[39] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3733–3742.

[40] 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.

[41] Y. Wang, J. Nie, P.-T. Yap, F. Shi, L. Guo, and D. Shen, "Robust deformable-surface-based skull-stripping for large-scale studies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2011, pp. 635–642.
[42] S. Denis, B. Esther, L. Boudewijn, S. Marion, K. Stefan, and S. Marius, "Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease," *Frontiers Neuroinform.*, vol. 7, Jan. 2014, Art. no. 50.
[43] J. Mazziotta *et al.*, "A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM)," *Philos. Trans. Roy. Soc. London B, Biol. Sci.*, vol. 356, no. 1412, pp. 1293–1322, Aug. 2001.
[44] D. P. Kingma and L. B. Jimmy, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 7–9.
[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Nov. 2015, pp. 234–241.
[46] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[47] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[48] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1473–1482, Jul. 2016.
[49] K. Kwak, H. J. Yun, G. Park, and J.-M. Lee, "Multi-modality sparse representation for Alzheimer's disease classification," *J. Alzheimer's Disease*, vol. 65, no. 3, pp. 807–817, Sep. 2018.
[50] C. Feng, A. Elazab, P. Yang, T. Wang, B. Lei, and X. Xiao, "3D convolutional neural network and stacked bidirectional recurrent neural network for Alzheimer's disease diagnosis," in *Proc. Int. Workshop Predictive Intell. Med.*, in Lecture Notes in Artificial Intelligence, 2018, pp. 138–146.
[51] Z. Zhang *et al.*, "THAN: Task-driven hierarchical attention network for the diagnosis of mild cognitive impairment and Alzheimer's disease," *Quant. Imag. Med. Surg.*, vol. 11, no. 7, pp. 3338–3354, Jul. 2021.
[52] J.-H. So, N. Madusanka, H.-K. Choi, B.-K. Choi, and H.-G. Park, "Deep learning for Alzheimer's disease classification using texture features," *Current Med. Imag.*, vol. 15, no. 7, pp. 689–698, Aug. 2019.
[53] X. Jiang, L. Zhang, L. Qiao, and D. Shen, "Estimating functional connectivity networks via low-rank tensor approximation with applications to MCI identification," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 7, pp. 1912–1920, Jul. 2020.
[54] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, and D. Shen, "Estimating functional brain networks by incorporating a modularity prior," *NeuroImage*, vol. 141, pp. 339–407, Nov. 2016.



Jiaguang Li received the B.S. and M.S. degrees in mechanical engineering and automation from Northeastern University, Shenyang, China, in 2019 and 2021, respectively, where he is currently pursuing the Eng.D. degree with the College of Information Science and Engineering.

His research lies at the intersection of deep learning and image processing. His current research works are about medical images and satellite images based on deep learning.

Ying Wei (Member, IEEE) received the B.Sc. degree from the Harbin Institute of Technology, Harbin, China, in 1990, and the M.Sc. and Ph.D. degrees from Northeastern University, China, in 1997 and 2001, respectively.

Her research interests include image processing and pattern recognition, medical image computation and analysis, computer vision, computer-aided diagnosis/detection (CAD), deep learning, and so on. She is now a full-time professor with Northeastern University, China. She has more than 50 journal papers and two granted patents in her research fields.

Dr. Wei is an elected committee of branch Association of Chinese Automation, and committee of Association of Artificial Intelligence, Liaoning Province, China. She served as a Session Chair for CCDC 2016, CCDC 2015, CCDC 2012, CCDC 2011, and so on.

Chuyuan Wang received the B.S. and M.S. degrees from Northeastern University, China, in 2016 and 2019, respectively, where she is currently pursuing the Ph.D. degree in control science and engineering.

Her research interests include medical image processing, brain image segmentation, and auxiliary diagnosis of Alzheimer's disease.

Qian Hu received the B.S. and M.Sc. degrees from Northeastern University, China, in 2018 and 2020, respectively, where she is currently pursuing the Ph.D. degree with the Faculty of Information Science and Engineering.

Her research interests are machine learning, pattern recognition, medical image processing, and brain image segmentation.

Yue Liu received the B.S. degree in electronic and information engineering from Northeastern University, Shenyang, China, in 2016, where she is currently pursuing the Ph.D. degree with the College of Information Science and Engineering.

Her research lies at the intersection of deep learning and image processing. Her current research work is about medical image analysis and cortical surface registration based on deep learning.

Long Xu received the B.S. degree from Inner Mongolia University, China, in 2016, and the M.S. degree from Northeastern University, China, in 2018, where he is currently pursuing the Ph.D. degree in pattern recognition.

His research interests include machine learning and visual object tracking.