CNNs Based Multi-Modality Classification for AD Diagnosis

Danni Cheng, Manhua Liu*
Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, Shanghai, China
*mhliu@sjtu.edu.cn

***Abstract*—Accurate and early diagnosis of Alzheimer's disease (AD) plays a significant part for the patient care and development of future treatment. Magnetic Resonance Image (MRI) and Positron Emission Tomography (PET) neuroimages are effective modalities that can help physicians to diagnose AD. In past few years, machine-learning algorithm have been widely studied on the analyses for multi-modality neuroimages in quantitation evaluation and computer-aided-diagnosis (CAD) of AD. Most existing methods extract the hand-craft features after image preprocessing such as registration, segmentation and feature extraction, and then train a classifier to distinguish AD from other groups. This paper proposes to construct multi-level convolutional neural networks (CNNs) to gradually learn and combine the multi-modality features for AD classification using MRI and PET images. First, the deep 3D-CNNs are constructed to transform the whole brain information into compact high-level features for each modality. Then, a 2D CNNs is cascaded to ensemble the high-level features for image classification. The proposed method can automatically learn the generic features from MRI and PET imaging data for AD classification. No rigid image registration and segmentation are performed on the brain images. Our proposed method is evaluated on the baseline MRI and PET images from Alzheimer's Disease Neuroimaging Initiative (ADNI) database on 193 subjects including 93 Alzheimer's disease (AD) subjects and 100 normal controls (NC) subjects. Experimental results and comparison show that the proposed method achieves an accuracy of 89.64% for classification of AD vs. NC, demonstrating the promising classification performance.**

Keywords- Multi-modality Classification, Convolutional neural networks (CNN), Cascaded CNNs, Alzheimer's disease diagnosis, Image classification.

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that results in impairment of the memory and cognitive functions [1]. Currently there is no effective cure for AD, but it is of great interest to develop treatments that can delay its progression [2]. Thus, accurate and early diagnosis of AD is not only challenging but also important for patient care and development of treatments. Neuroimaging technique, such as magnetic resonance images (MRI) and positron emission tomography (PET) are providing powerful imaging modalities to help observe the anatomical and functional neural changes related to AD [1, 2]. Magnetic resonance image (MRI) is a non-invasive medical imaging modality used to imaging the internal body structures. Structural MRI are often used to capture the regional brain atrophy and help understand the brain anatomical changes. Thus, it proved as a promising representative of Alzhimer's disease progression [2]. Positrons Emission Tomography (PET) is a functional biomedical modality that can help physicians to diagnose AD. A positron-emitting radionuclide (tracer) with a biologically active molecule, such as (18)F-fluorodeoxy-glucose, is introduced in the body. Concentrations of this tracer are imaged using a camera and tissue metabolic activities are indicated by the absorption of glucose in the corresponding region [3]. Fluorodeoxy-glucose positron emission tomography (FDG-PET) offers a powerful functional imaging biomarker to assist AD diagnosis.

In recent years, various pattern recognition methods have been investigated in analyzing multi-modality neuroimages to identify those patterns related to AD and release the disease states for computer-aided-diagnosis (CAD) [2-8]. Previous studies have shown that it is very important to extract the influential biomarkers for image classification. The region based method was proposed to extract features for classification of AD with MRI and PET images [4-6]. In this method, brain images are mapped into many anatomical regions of interest (ROIs) and the first four moments and the entropy of the histograms of these regions are computed as the regional features. Zhang et al. proposed a multi-kernel support vector machine (SVM) to ensemble those multimodal brain features such as tissue volumes extracted from 93 ROIs for disease classification [4]. Recently, deep learning networks were also explored to extract the latent features from measurements of ROIs with different image modalities for AD classification [5-6]. Liu et al. [5] extracted a set of latent features from 83 ROIs of the MRI and PET scans and trained a multi-layered neural network consisting of several auto-encoders to combine multimodal features for classification. Suk et al. [6] employed a stacked autoencoder to study the latent high-level features separately from 93 ROIs of the MRI and PET images and then these features is combined use a multi-kernel SVM to improve the classification performance.

To capture the rich neuroimaging information, voxel-wise features were extracted after registering all neuroimaging data to associate each voxel with a vector of scalar measurements for AD diagnosis [7-9]. The brain volume is segmented into gray matter (GM), white matter (WM), and CSF parts, and the voxel-wise tissue density maps are leveraged to calculate the

Sponsors: NSFC No. 61375112 & 61773263 and SJTU

regional GM loss for AD classification [7,8]. The voxel-wise features of MRI and PET including the GM density map of MRI and the intensity values of PET are combined with a sparse regression classifier, and a CNN based framework was proposed to estimate the missing PET scans for multimodal classification of AD in [9].

Despite multimodal neuroimaging analysis has promising results in above literature, there are still some limitations in the above methods. Feature dimension reduction and robust representations privilege are revealed in ROI-based method, but some small disturbance might be neglected. In addition, the ROIs are segmented by prior hypotheses, therefore, those abnormal brain regions relevant to AD might not be suitable using the pre-defined ROIs, which will lead to extracted features lacking representation ability. The voxel-wise features can alleviate this problem, but they have huge dimensionality, far more features than training subjects, which may lead to poor classification performance for the sake of 'curse of dimensionality'[4]. In addition, the correctness of extracted features highly depends on image preprocessing steps such as segmentation and registration, which requires the domain expert knowledge.

To alleviate these problems, convolutional neural networks (CNNs) have recently been a hot spot for neuroimage analysis [10,11]. Hosseini-Asl et al. [10] proposed a deep 3D-CNNs based method to learn generic features capturing AD biomarkers and predict AD using the structural MRI scans. In this method, deep 3D-CNNs are built upon the 3D autoencoder and the convolutional filters are pre-trained to capture anatomical shape variations of structural MRI. The deep CNNs were explored to extract the imaging features from both the structural MRI and functional MRI for AD classification [11]. The above methods concentrated on the AD diagnosis based on MRI. In the case of multimodal neuroimages, further investigations is required to determine their ability to diagnose AD.

Motivated by the success of CNNs in image classification, this paper raises a novel classification framework based on the cascaded CNNs to learn the multi-level and multi-modality imaging features using MRI and PET images, which are combined to classify AD and NC subjects. First, a deep 3D CNN model is constructed to hierarchically and gradually transform the whole brain image into more compact and discriminative features for each modality. Second, a high-level 2D CNN is cascaded to combine the multi-modality features learned from multiple CNNs for image classification. The lower 3D CNN are individually learned for each modality image and the upper 2D convolution layers are fine-tuned to combine the multi-modality features for image classification. The proposed method works well to automatically extract the generic features from the high-dimensional imaging data and combine the multi-modality features for image classification. Our experimental results on ADNI database demonstrate the effectiveness of the proposed method for AD diagnosis.

The rest of this paper is organized as follows. Section 2 presents the proposed multi-modality classification method based on CNNs. Experiments and comparison are provided in Section 3. Section 4 concludes this paper.

II. Proposed Multi-modality Classification Method

Motivated by the success of CNNs in computer vision, this paper proposes to construct cascaded CNNs to learn the generic features from MRI and PET brain images for multi-modality classification. Figure 1 shows the overview of the proposed classification method. The framework of our classification method consists of two main steps: feature extraction with deep CNNs for a single modality and multi-modality classification for AD diagnosis, as detailed below.

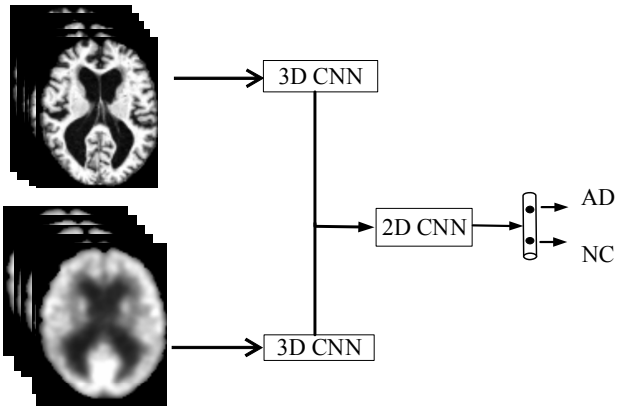


Figure 1. The overview of the proposed classification method based on cascaded CNNs.

A. Feature extraction with deep CNN for a single modality

In this section, we will present the proposed multi-modality classification method by deep CNNs in detail. If all voxel intensities of each image are directly used for classification, there are 82,000 features input to the classifier. Since the voxel intensity features are of big dimensionality, much more features than the number of training subjects, which may lead to low classification performance. In addition, no rigid registration is performed for each image so that the features are sensitive to the transformation variance. Convolutional neural networks (CNNs) proves to be a powerful tool for various applications such as image classification and object detection [14-16]. CNNs can extract visual patterns directly from pixel images, which takes almost no preprocessing step. They has been successfully applied to those occasion with extreme variability, likes handwritten characters recognition, largescale image classification and so on, which shows robustness in distortions, geometric transformations and etc.

To efficiently encode the richer spatial information of 3D brain images, the 3D convolution kernel is employed to build the deep 3D CNN architecture for each modlaity neuroimage in this work. The deep CNN is built to hierarchically and gradually learn the scale and shift invariant features of multi-modality brain images at low-high level. There are several variations on the CNN architecture. Typically, a deep CNN for

feature extraction alternatively stacks several convolutional and sub-sampling layers followed by fully connected and softmax layers. In this work, the CNN architecture for 3D brain image classification is composed of convolutional, maxpooling, fully connected and softmax classification layers as shown in Figure 2.

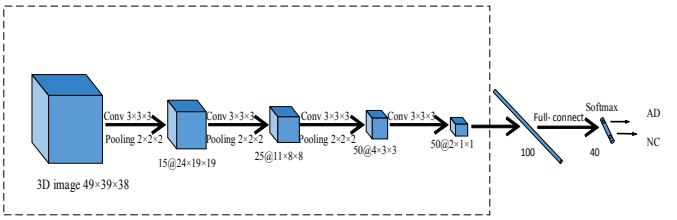


Figure 2. The architecture of deep CNNs denoted with the sizes of input, convolution, max pooling and output layers and the numbers and sizes of generated feature maps.

The first layer of the deep CNNs is the input layer which accepts a 3D image patch of fixed size (49×39×38 voxels). The second type is the convolutional layers. A typical convolutional layer usually convolves the last input features with the learned kernel filters, followed by adding a bias term and applying a non-linear activation function, and finally produce a feature map for each filter. By using 3D kernel to consider the spatial correlations of three dimensions, the 3D CNN can take privilege of the volumetric contextual features. After convolution, *tanh* is adopted as activation function. In this way, convolutional layer can capture the important features of neuroimage.

A pooling layer followed after each convolutional layer which used to downsample the feature maps. There are several formation of pooling, such as averaging, maximum, and linear combination of block neural values. In our framework, the maximum of the pooling block is taken to obtain more compact and efficient features. Maxpooling can downsample the input feature map along the spatial dimensions by replacing each non-overlapping block of fixed voxels with their maximum. This function can reduce the dimension and keep the most influential features for distinguishing disease. Through maxpooling, the features become more compact and efficient from low to higher layer, which can obtain the high-level features robust to some variations.

The forth type of layer is the fully connected layers which consist of a number of output neurons. After a series of convolutional and max-pooling layers appearing alternately, those 3D feature maps are flattened into a 1D feature vector, then a fully connected layer is connected. A fully connected layer give connected to all feature elements in the previous layer and output layer. It consists of a number of output neurons. Each neuron of fully connected layer outputs the learned linear combination of all neural values in previous layer and passed through a nonlinearity. The inputs and outputs of fully connected layers are concatenated into one-dimensional vector and they are not spatially located anymore.

At the end of CNN, a softmax classification layer is appended which is fine-tuned back-propagation with negative log-likelihood to predict class probability. The softmax function is a generalization of logistic function that maps feature vectors to output the probabilities of each output classes. The softmax layer has been extensive used at the last of CNN architecture to predict a probabilistic classification score for each class label. The output of each node, in fact, ranges from 0 and 1, and the sum of all the nodes is always 1.

In our implementation, the 3D deep CNN is built by stacking 4 convolutional layers, 3 max-pooling layers, 1 fully connected layers and a final softmax layer. The sizes of all convolution filters are set to 3×3×3 and the filter numbers are set to 15, 25, 50, 50 for 4 convolution layers. Max pooling is applied for each 2×2×2 region. *Tanh* function is adopted as the activation function in these layers for its good performance. The 3D convolutional kernels are initialized use xavier uniform initializer method. Those parameters of the our proposed network are tuned using the back-propagation with Adadelta by minimizing the cross entropy loss. In addition, the dropout strategy is implemented by randomly and temporarily disconnected those input and output neurons, which can reduce overfitting problem and improve generalization capability of our model [17].

B. Multi-modality classification by cascaded CNNs

Inspired by microcolumns of neurons in the cerebral cortex, we combine the multiple deep CNNs trained for each single modality neuroimage to form a multi-modality classification. While the lower layers of the 3D-CNN can extract discriminative features, the upper layers are trained for task-specific classification with these features. Thus, training of the proposed multi-modality classification consists of pre-training of individual CNN, and task-specific fine-tuning for final classification. Initially, we train a deep CNN for each modality separately by directly mapping the outputs of the fully connected layer to the probabilistic scores of all class labels. Different from the conventional combination methods by averaging the class probabilistic scores, we propose to learn a 2D CNN to combine the multi-modality features and make the final classification. The output feature maps by 3D CNN are flatten to one dimension, and then 1D feature vectors of MRI and PET are combined into two dimensional feature map to conduct 2D CNN.

In the learning process of 2D CNN, the initial-trained parameters of 3D CNNs are used to fix the first three convolutional and three pooling layers of each 3D CNN, while the parameters of the last convolutional layer and upper CNN layers are fine-tuned jointly to combine the multi-modality features with a softmax top-most output layer for the task-specific classification. There are two advantages by finetuning the last few layers for final classification. First, by fixing the first several layers, the knowledge learnt from each modality can be preserved to extract the specific features. By fine-tuning the last few layers, the models can be more adapted to the global classification task. So the knowledge in both imaging variations and classification task can be integrated and help improve the classification accuracy. Second, comparing with fine-tuning the whole network, fine-tuning the last few layers significantly reduce the computational cost and the overfitting problem.

There are three main advantages to apply deep CNNs for our task. First, the deep complex architecture can extract the low- to high-level features learning from a large number of training images. Second, since different neuroimages have different anatomical and functional characteristics, deep CNNs can take full advantage of the image spatial relationships and learn those local 3D filters effectively for the final classification task. Finally, by stacking multiple modality CNNs, our modal can extract a hierarchy of more complex features representing the states of Alzheimer's diseases, finally providing the global class prediction probabilities.

III. Experimental Results

Our proposed method is evaluated on MRI and FDG-PET neuroimages. The T1-weighted MRI are widely obtainable, non-invasive modalities, thus it is often seen as the first biomarker in AD diagnosis. In addition, FDG-PET images provides a powerful functional imaging biomarker to help inspect the neural changes of AD diagnosis. All the test data used in this work comes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). In ADNI, volumetric 3D MPRAGE with 1.25×1.25 mm2 in-plane spatial resolution and 1.2 mm thick sagittal slices the T1-weighted are conducted to obtain MRI data. Most of these images were collected with 1.5T scanners, while few parts were obtained using 3T scanners. We employ T1-weighted MR and FDG-PET imaging data from 193 ADNI participants of the baseline visits. Detailed information about MRI and FDG-PET acquisition procedures is available at the ADNI website.

We conduct pre-processing steps of MRI images before feature extraction and classification. Particularly for MRI, nonparametric nonuniform intensity normalization algorithm are firstly conducted to correct the intensity inhomogeneity, then skull-stripped and cerebellum-removed [12, 13]. The PET images were processed to make the images from different systems appears analogous. Then, intensity normalization and isotropic resolution uniform with 8 mm FWHM are performed as in [3]. The voxel intensities of each PET image are used for classification. The voxels with the zeros mean gray values are removed from the image analysis and final images used are of size 98×78×76 voxels. We further down-sampled the neuroimage to 49×39×38 voxels by 2 to reduce the requirements of computation and memory costs for the experiments.

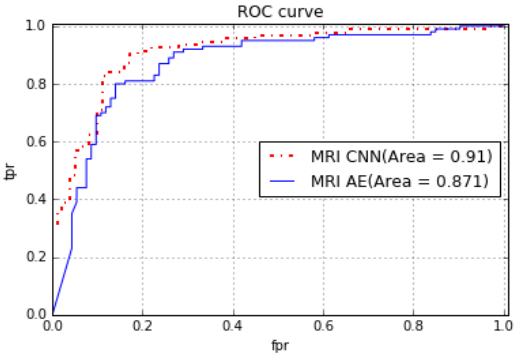
In this section, we conduct some experiments to test the proposed classification algorithm. MRI and PET images are taken from 193 subjects including 93 AD, and 100 NC from ADNI for our experiments. Ten-fold cross-validation is used to avoid random factors affecting results. Each time, one fold of the image set is used for testing, another fold used for validation while the left eight folds were used for training. The validation part is for parameter tuning. To increase training data, augmentation is performed by shift, sampling and rotation to generate additional images for the training set. The proposed algorithm is implemented with the Keras library in Python. The experiment is conducted on PC with GPU NVIDIA GTX1080.

In the low level, two deep CNNs are independently trained to extract the features with the output of the prediction scores. The Adadelta gradient descent algorithm [16] is used to train the local deep CNNs. To avoid overfitting, dropout, L1 and L2 regulation are adopted in building our CNN models [17]. The batch size is set to 64, and the model begins to converge after 15~30 epoch. To evaluate the classification performance, we compute the classification accuracy (ACC), the sensitivity (SEN), the specificity (SPE), and the area under receiver operating characteristic curve (AUC) in the experiments.

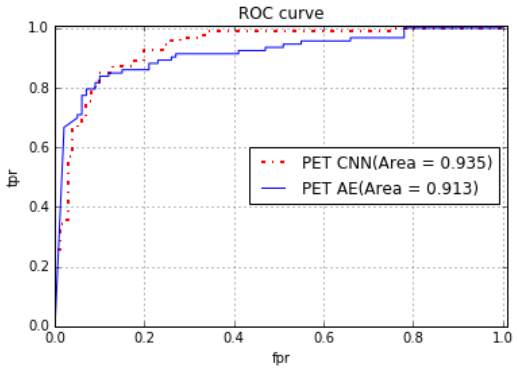
The first experiment is to compare the proposed deep 3D CNNs to the method by 3D autoencoder [10]. For fair comparison, we downloaded the source codes released by the authors of [10] in the website and implemented with our best effort by using a single modality of our data set. Table 1 shows the comparison of classification results by 3D Auto-Encoder method [10] and our proposed 3D CNN method using the MRI and PET images. Figure 3 compares the ROC (receiver operating characteristic) curves of 3D Auto-Encoder method [10] and our proposed 3D CNN method using the MRI and PET images for classification of AD vs. NC. From these results, the proposed CNN method performs better than 3D Auto-Encoder method. The multi-modality by cascaded CNNs can further improve the classification accuracy.

TABLE I. Compaison of the classification results by Auto-Encoder method and our proposed CNN method

Method	Modality	ACC(%)	SEN(%)	SPE(%)	AUC(%)
Auto-Encoder [10]	MRI	81.87	81.00	82.80	87.09
	PET	84.97	84.95	85.00	91.34
Proposed CNN Method	MRI	85.47	83.87	90.00	90.98
	PET	87.13	87.10	84.00	93.49
Muti-modality		89.64	87.10	92.00	94.45



(a) MRI



(b) PET

Figure 1. Comparison of ROC curves for classification of AD vs. NC with MRI and PET images.

In addition, we compared the performances of our proposed multi-modality fusion appoarch with those results by directly summing or averaging of the prediction scores produced by multi-modality neuroimages. Table II shows Comparison of the classification results by averaging fusion method and the proposed 2D CNN method. These results further confirm the effectiveness of our proposed algorithm for Alzheimer's disease classification.

TABLE II. COMPAISON OF THE CLASSIFICATION RESULTS BY AVERAGING FUSION METHOD AND THE PROPOSED 2D CNN METHOD.

Method	ACC(%)	SEN(%)	SPE(%)	AUC(%)
Averaging	88.60	87.10	90.00	93.90
Proposed 2D CNN	89.64	87.10	92.00	94.45

IV. CONCLUSIONS

This paper has put forward a multi-modality classification method based on the cascaded CNNs to classify AD vs.NC using MRI and PET images. Two deep CNNs are built on different modality images to learn the specific discriminative features. Then a high-level CNN is cascaded to combine the features learned from different modalities for image classification. No segmentation and rigid image registration are performed on the MRI and PET scans. Our experimental results and comparison on ADNI datasets confirm the performance improvement of the proposed method for AD diagnosis. There may be great potential for extending our study to other biomedical areas.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (NSFC) under grants No. 61375112 and 61773263 and SJTU Excellent Young Faculty program.

REFERENCES

[1] L. Minati, T. Edginton, M. G. Bruzzone, and G. Giaccone, "Reviews: current concepts in Alzheimer's Disease: a multidisciplinary review," American Journal of Alzheimer's Disease and Other Dementias, 2nd ed, vol. 24, 2009, pp.95-121.

[2] R. Cuingnet, E. Gerardin, J. Tessieras et al., "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database," NeuroImage, 2nd ed, vol. 56, 2011, pp.766-781.

[3] M Silveira, J Marques, "Boosting Alzheimer Disease Diagnosis Using PET Images", International Conference on Pattern Recognition, pp.2556-2559, 2010.

[4] D. Zhang, Y. Wang, L. Zhou, H. Yuan and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment", Neuroimage, 3td ed, vol. 55,2011,pp. 856-867.

[5] S. Liu, W. Cai, H. Che et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease", IEEE Trans. on Biomedical Engineering, 4th ed, vol. 62, pp.1132-1140, 2015.

[6] H. I. Suk, S. W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis", Brain Structure and Function, 2nd ed, vol. 220, 2015, pp. 841-849.

[7] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer et al., "Automatic classification of MR scans in Alzheimer's disease", Brain A Journal of Neurology, 3rd ed, vol. 131, 2008, pp. 681-689.

[8] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung and S. C. Johnson, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset", Neuroimage, 1st ed, vol. 48, 2009, pp. 138-149.

[9] RJ Li, WL Zhang, H. I. Suk, W Li, L. Jiang and DG Shen et al., "Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis", Med Image Comput Comput Assist Interv. 3rd ed, vol. 17, 2014, pp. 305-312.

[10] E. Hosseiniasl, R. Keynto, A. Elbaz, "Alzheimer's Disease diagnostics by adaptation of 3D convolutional network", IEEE International Conference on Image Processing, Phoenix, Arizona, USA, Sept. 25-28, 2016.

[11] S. Sarraf, G. Tofighi, "DeepAD: Alzheimer's Disease classification via deep convolutional neural networks using MRI and fMRI", biorxiv.org.

[12] Y. Wang, J. Nie, P. T. Yap, F. Shi, L. Guo and D. Shen, "Robust deformable-surface-based skull-stripping for large-scale studies", International Conference on Medical Image Computing and Computer-Assisted Intervention, 3rd ed, vol. 14, pp. 635-642, 2011.

[13] J. G. Sled, A. P. Zijdenbos, A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data", IEEE Trans Med Imaging, 1st ed, vol. 17, pp.87-97, 1998.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition". Proceedings of the IEEE, 11th ed, vol.86, pp.2278–2324, 1998.

[15] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks". International Conference on Neural Information Processing Systems, 2nd ed, vol.25, pp.1097-1105, 2012.

[16] M. D. Zeiler, "ADADELTA: An adaptive learning rate method", arXiv:1212.5701, 2012.

[17] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting", Journal of Machine Learning Research, 1st ed, vol.15, 2014, pp.1929-1958.