3D convolutional neural network for feature extraction and classification of fMRI volumes

1st Hanh Vu
Department of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea
vuthihanh.bk@gmail.com

2nd Hyun-Chul Kim
Department of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea
hyunchul_kim@korea.ac.kr

3rd Jong-Hwan Lee
Department of Brain and Cognitive Engineering
Korea University
Seoul, Republic of Korea
jonghwan_lee@korea.ac.kr

Abstract— **Recently, deep learning (DL) techniques have been gaining interest in the neuroimaging community. In this study, we present 3D convolutional neural network (3D-CNN) as an end-to-end model to label a target task among four sensorimotor tasks for each functional magnetic resonance imaging (fMRI) volume. To the best of our knowledge, this is the first study that employs a single blood-oxygenation-level-dependent (BOLD) fMRI volume as the input of the 3D-CNN for task classification. We hypothesized that 3D-CNN has the capability to extract potentially shift-invariant features in local brain areas while preserving the overall spatial layout of the whole brain fMRI volume. We designed a 3D-CNN model by extending the LeNet-5 CNN for 2D image classification to 3D volume classification. The designed 3D-CNN model was thoroughly evaluated using BOLD fMRI volumes acquired from four sensorimotor tasks in terms of the classification performance and feature representations for each of the four sensorimotor tasks.**

Keywords—Convolutional neural network (CNN), deep learning, fMRI, 3D-CNN, sensorimotor

I. Introduction

Deep learning (DL), which has been proven to be a powerful sub-field of machine learning, is gaining popularity in various research fields including pattern recognition, computer vision, and natural language processing. For example, the convolutional neural network (CNN), one of the DL models, has been proven its capability for image classification [1, 2], face recognition [3], and video captioning [4, 5] tasks by automatically extracting representative features from input data without a need for a feature engineering step. More recently, the CNN has also been applied to automatically reconstruct magnetic resonance imaging (MRI) image from raw k-space data [6].

There have also been growing efforts to apply DL techniques to analysis of neuroimaging data including functional MRI (fMRI) to investigate the human brain functions [7-11]. For instance, Jang and colleagues conducted classification of fMRI volumes acquired from sensorimotor tasks using a fully-connected feedforward deep neural network (fcDNN) that was pretrained using a deep belief network (DBN) [9]. In this study, the estimated DNN weights showed spatial patterns that are remarkably task-specific, specifically in the higher layer that is closest to the output layer. In another study using 3D-CNN, spatial patterns from a sparse dictionary learning of whole-brain fMRI volumes were used as input for automatic recognition of resting-state functional networks [11].

To the best of our knowledge, however, there has been no study employing the 3D-CNN for raw blood-oxygenation-level-dependent (BOLD) fMRI volumes as input. We hypothesized that the 3D-CNN model with 3D volume input is potentially better suited for fMRI volume classification than the fcDNN model with 1D input vector [9]. This is because the convolutional operation followed by pooling operation may address (a) misalignment issues during spatial normalization of raw fMRI volumes [12, 13] to a standard space and (b) individual variability of local BOLD responses across sessions and subjects [14, 15]. To this end, we implemented a 3D-CNN model by extending the LeNet-5 CNN for 2D image classification [16] to 3D fMRI volume classification. The sensorimotor dataset that was used for previous study to evaluate the fcDNN [9] was also applied to our 3D-CNN model. Then, the 3D-CNN and 1D-fcDNN were evaluated in terms of their classification performance and features representation capability.

II. Materials and Methods

A. Data acquisition

1) Participants and experimental setup

Twelve young, healthy, right-handed subjects (age = 25.0 ± 2.0 years) participated in our experiments. Each subject performed one of four sensorimotor tasks including left-hand clenching (LH), right-hand clenching (RH), auditory attention (AD), and visual stimulus (VS) during each of the four fMRI runs (150 s per run; Fig. 1) [9, 17].

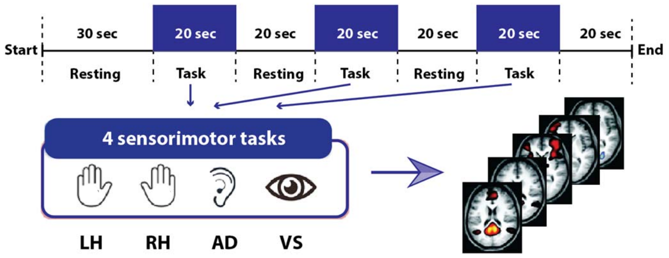

Fig. 1. Experiment paradigm of one fMRI run (150 sec).

This work was supported by the National Research Foundation (NRF) grant, MSIP of Korea (NRF-2015R1A2A2A03004462, NRF-2017R1E1A1A01077288, NRF-2016M3C7A1914450), in part by the grant of the Ministry of Education of the Republic of Korea and the NRF of Korea (NRF-2015S1A5B6036594), in part by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Korea (HI12C1847), and in part by the ICT R&D program of MSIP/IITP [R7124-16-0004, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding].

978-1-5386-6859-7/18/$31.00 ©2018 IEEE

A 3-T Tim Trio MRI scanner (Siemens, Erlangen, Germany) and 12 channel head coil were used to acquire the fMRI run with BOLD contrast by applying a standard gradient-echo echo-planar-imaging (EPI) pulse sequence (repetition time, or TR = 2000 ms; time of echo = 30 ms; field-of view = 240×240 mm^2; in-plane voxel = 64×64; 36 axial slices with 4 mm thickness without a gap; voxel size = 3.75×3.75×4 mm^3).

2) Data preprocessing

The BOLD fMRI volumes were preprocessed using the SPM8 software toolbox with standardized options including slice timing correction, motion correction, spatial normalization to the Montreal Neurological Institute space with a 3-mm isotropic voxel size, and spatial smoothing with an 8-mm full-width at half-maximum Gaussian kernel.

The BOLD signals of the preprocessed fMRI volumes within a brain area were changed to the percentage changes using average BOLD intensity in all rest periods as baseline. Then, each 3D fMRI volume with percentage BOLD intensities was used as an input sample of the 3D-CNN, whereas the 1D voxel-series within the brain area with percentage BOLD intensities was used as an input sample of the fcDNN.

The TR for each fMRI volume was 2 s and thus there were (a) 30 input samples for each of the four tasks in each subject and (b) a total of 1,440 input samples across the 12 subjects and across the four tasks.

B. 3D-CNN for 3D fMRI volume classification

1) Our 3D-CNN Architecture

Fig. 2 shows the network architecture of our 3D-CNN model which is a 3D extension of the LeNet-5 model [16]. Our 3D-CNN consisted of three 3D convolutional (Conv) layers (8 filters with 7×7×7 kernel size in the first convolutional layer, 16 filters with 5×5×5 kernel size in the second convolutional layer, and 32 filters with 3×3×3 kernel size in the third Conv layers), two fully-connected layers (128 hidden nodes in the hidden layer), and one output layer with four output nodes to classify each of the four tasks.

For example, considering that each 3D volume input was 53×63×46, the dimension of output pattern at the first Conv layer was 24×29×30 due to a stride of two after Conv operation and there were eight of them across 8 filters (i.e., channels); the dimension of output pattern at the second Conv layer was 10×13×8 from a stride of two and there were 16 of them across the 16 channels; the dimension of the output pattern at the third Conv layer was 4×6×3 from a stride of two and there were 32 of them across the 32 channels. Then, the output of the third Conv layer was changed into 1D vector with 2304 elements (= 4×6×3×32) and this was used as the input of the fully-connected layer.

2) Training of 3D-CNN

Parameters used to train our 3D-CNN were as follows: rectified linear units (ReLU) activation function for each node in the Conv layers as well as the FC layer; cross-entropy loss function at the output layer; stochastic gradient descent with an initial learning rate of 10^{-3} (without momentum) and annealing after 50 epochs with a minimum learning rate 10^{-6}; mini-batch size of 50; dropout with a probability of 0.5 in the third 3D Conv layer to minimize overfitting. Additionally, the hyperbolic tangent (Tanh) activation function was also applied for the 3D-CNN to compare with ReLU. The leave-one-subject-out cross-validation framework was used to evaluate the performance (i.e., 11 subjects were used for training the 3D-CNN and one remaining subject for test the trained 3D-CNN and repeat this process for each one of 12 subjects as test subject). The output of each convolutional layer was the learned feature maps of the input.

C. fcDNN for 1D voxel pattern classification

The fcDNN that was pre-trained using the deep belief network [9]was also employed to compare the performance and weight feature representations with those from our 3D-CNN. The fcDNN with three hidden layers (and 100 hidden nodes for each hidden layer) was adopted and the optimal sparsity level of weights of the fcDNN at the first layer (i.e., the target percentage of non-zero weights of 0.001) was used to minimize overfitting [9]. The weights of the first layer of the trained fcDNN were interpreted as weight feature map [8, 9].

In addition, the classification performance was dependent on an activation function [9] and thus the two scenarios of the

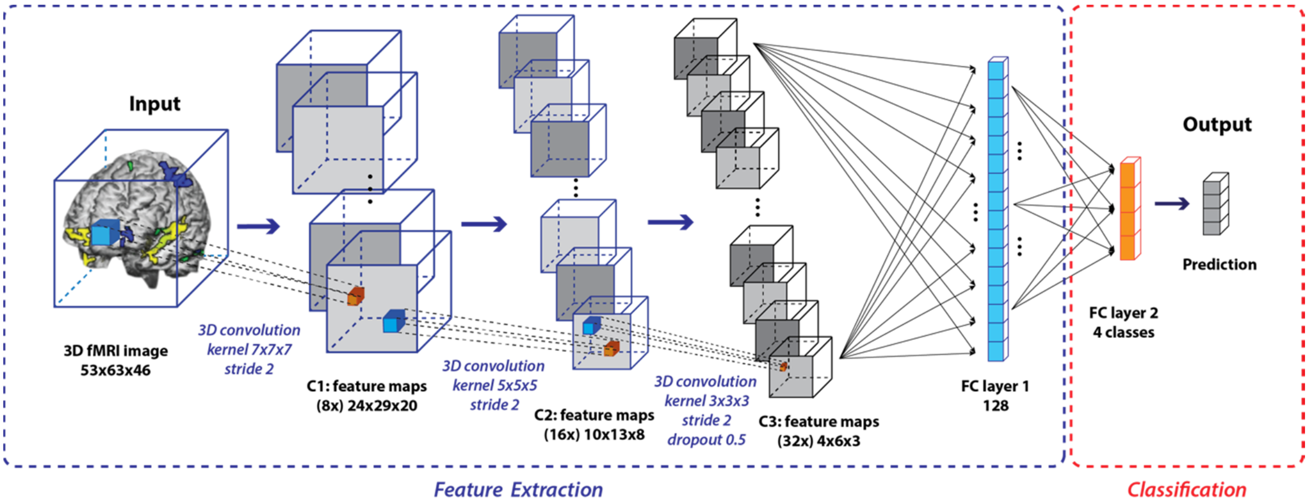

Fig. 2: The 3D-CNN model for 3D fMRI volume classification

fcDNN models including (a) Tanh and (b) ReLU as the node activation functions were considered and the performance as well as the weight feature maps were interpreted.

III. Results

Fig. 3 shows the results of error rates using fcDNN and 3D-CNN for each of the four sensorimotor classes that were averaged across all 12 subjects (with standard error of the mean as whisker). The dashed line indicates the averaged error rate of support vector machine (SVM) based classification using fMRI volume as the input [9]. The statistical significance was evaluated by the paired t-test. Overall, the 3D-CNN showed decreased error rates compared to the fcDNN and SVM. Particularly, 3D-CNN with ReLU activation function showed the lowest average error rate and notably, the error rate of the AD task (6.4 ± 2.7%) was significantly dropped from the 3D-CNN with ReLU than the fcDNN models. The error rates of the LH, RH and VS from the 3D-CNN and fcDNN with ReLU were comparable.

In terms of the comparison between the ReLU and Tanh, the ReLU activation function (3.8 ± 0.9%) in the fcDNN yielded a significantly low error rate compared to the Tanh activation function (5.3 ± 1.1%; $p < 0.05$). Also, averaged error rate across

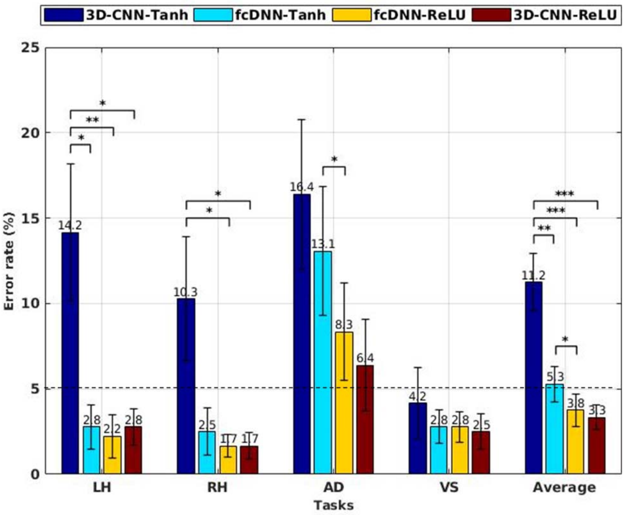

Fig. 3. Comparison of averaged classification error rates (mean ± standard deviation) of each sensorimotor task across all 12 subjects. The dashed line indicates the averaged error rate of SVM-based classification. The statistical significance was evaluated by the paired t-test (*, p < 0.05; **, p < 0.01; ***, p < 0.001). LH, left-hand clenching; RH, right-hand clenching; AD, auditory attention; VS, visual stimulus.

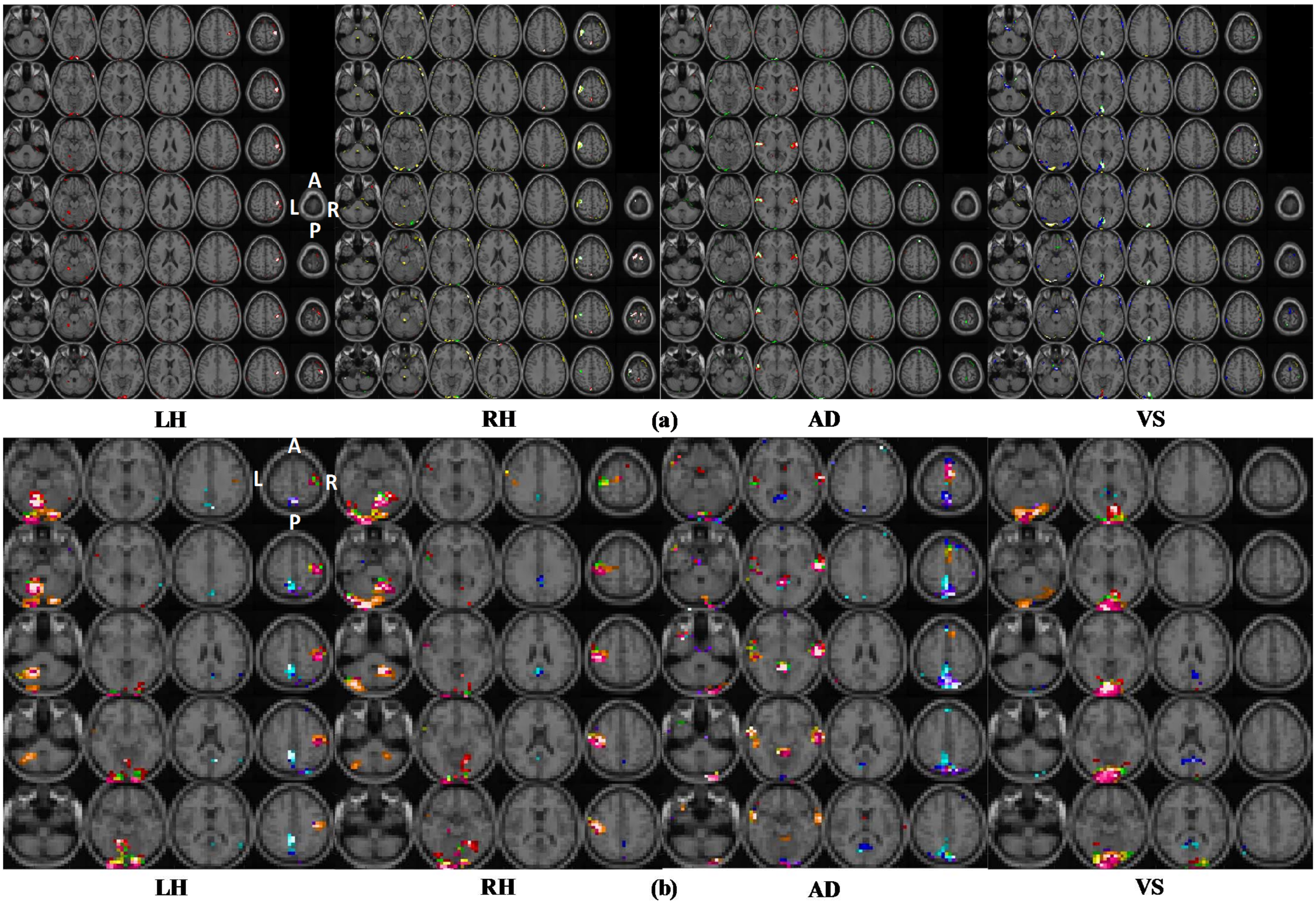

Fig. 4: (a) Visualization of the weight-feature map from the first layer of our DNN model in the case of using ReLU as activation function. The dimension of the weight-feature map was similar to the input's dimension (53×63×46). (b) Visualization of feature map from the first layer of our 3D-CNN model in the case of using ReLU as activation function. The dimension of feature map was reduced from 53×63×46 to 24×29×20 after the first 3D convolutional layer (stride = 2). The threshold to recognize activation patterns of brain areas was set to the top 1 percentile of brain voxels in both case (a) and (b). LH, left-hand clenching; RH, right-hand clenching; AD: auditory attention; VS: visual stimulus; L, left; R, right; A, anterior; P, posterior.

all 12 subjects was considerably decreased for the AD task classification (8.3 ± 2.9% for ReLU vs. 13.1 ± 3.8% for Tanh; $p < 0.05$). Using the 3D-CNN, ReLU activation function resulted in a markedly lower averaged error rate than Tanh activation function (3.3 ± 0.7% and 11.2 ± 1.7%, respectively; $p < 0.001$).

Fig. 4 shows the learned features from the 3D-CNN and the fcDNN, in which each feature map is represented by a different color and gradient of color to represent the intensity of weights (the brighter the color, the greater the intensity).

IV. Discussion

A. *Classification performance*

The 3D-CNN model showed a slightly higher performance to fcDNN possibly due to the capability of the 3D-CNN to handle the shift-invariant feature extraction in local brain areas while preserving the overall spatial layout of the whole brain fMRI volume.

The substantial improvement of the performance in both the case of fcDNN and 3D-CNN using ReLU activation function compared to to Tanh activation function was possibly due to the potential effect to deal with the vanishing gradient problem of the stochastic gradient descent learning algorithm by the ReLU function and/or additional regularization of hidden node output [18, 19]. The slightly improved classification performance from the 3D-CNN than fcDNN can potentially be pronounced from scenarios that the difference between fMRI volumes across classes are subtler in terms of spatial extent and their multivoxel patterns in local area compared to the sensorimotor data in this study.

The significantly lower error rates from fcDNN with Tanh compared to the 3D-CNN with Tanh would be due to the initialization of weights using the pretrained DBN as well as the fine-tuning of the L1-norm regularization via our weight sparsity optimization scheme [8, 9].

B. *Features representation*

3D-CNN has proven its superior performance in the automated feature extraction compared to fcDNN. In the fcDNN with ReLU, each of the weight-feature maps in the first layer is associated with one of the four tasks. Furthermore, the weight features associated a single task were significantly overlapping each other (Fig. 4a). On the other hand, the feature maps of a single task obtained from the 3D-CNN layers were localized in the task-relevant area and moreover, these distinct features were slightly shifted covering the task-relevant area for each of the four sensorimotor tasks (Fig. 4b).

V. Conclusion

We presented the utility of the 3D-CNN for the classification of a single fMRI volume. The 3D-CNN can be gainfully utilized to increase classification performance by automatically extracting shift-invariant features from the fMRI volumes despite of potential issues such as spatial misalignment during normalization and spatial variability of activation patterns across sessions and/or subjects.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[2] K. Korekado, T. Morie, O. Nomura, H. Ando, T. Nakano, M. Matsugu, *et al.*, "A convolutional neural network VLSI for image recognition using merged/mixed analog-digital architecture," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2003, pp. 169-176.

[3] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks,* vol. 8, pp. 98-113, 1997.

[4] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725-1732.

[5] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, 2014, pp. 568-576.

[6] B. Zhu, J. Z. Liu, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain transform manifold learning," *arXiv preprint arXiv:1704.08841,* 2017.

[7] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, *et al.*, "Deep learning for neuroimaging: a validation study," *Frontiers in neuroscience,* vol. 8, p. 229, 2014.

[8] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," *Neuroimage,* vol. 124, pp. 127-146, 2016.

[9] H. Jang, S. M. Plis, V. D. Calhoun, and J. H. Lee, "Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks," *Neuroimage,* vol. 145, pp. 314-328, Jan 15 2017.

[10] H. Huang, X. Hu, Y. Zhao, M. Makkie, Q. Dong, S. Zhao, *et al.*, "Modeling Task fMRI Data via Deep Convolutional Autoencoder," *IEEE Trans Med Imaging,* Jun 15 2017.

[11] Y. Zhao, Q. Dong, S. Zhang, W. Zhang, H. Chen, X. Jiang, *et al.*, "Automatic Recognition of fMRI-derived Functional Networks using 3D Convolutional Neural Networks," *IEEE Trans Biomed Eng,* Jun 15 2017.

[12] E. Dohmatob, G. Varoquaux, and B. Thirion, "Inter-subject registration of functional images: do we need anatomical images?," *Frontiers in Neuroscience,* vol. 12, p. 64, 2018.

[13] V. D. Calhoun, T. D. Wager, A. Krishnan, K. S. Rosch, K. E. Seymour, M. B. Nebel, *et al.*, "The impact of T1 versus EPI spatial normalization templates for fMRI data analyses," *Human brain mapping,* vol. 38, pp. 5331-5342, 2017.

[14] G. K. Aguirre, E. Zarahn, and M. D'Esposito, "The variability of human, BOLD hemodynamic responses," *Neuroimage,* vol. 8, pp. 360-9, Nov 1998.

[15] J. V. Haxby, J. S. Guntupalli, A. C. Connolly, Y. O. Halchenko, B. R. Conroy, M. I. Gobbini, *et al.*, "A common, high-dimensional model of the representational space in human ventral temporal cortex," *Neuron,* vol. 72, pp. 404-16, Oct 20 2011.

[16] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the Ieee,* vol. 86, pp. 2278-2324, Nov 1998.

[17] Y. H. Kim, J. Kim, and J. H. Lee, "Iterative approach of dual regression with a sparse prior enhances the performance of independent component analysis for group functional magnetic resonance imaging (fMRI) data," *Neuroimage,* vol. 63, pp. 1864-89, Dec 2012.

[18] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.

[19] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, 2013, p. 3.