Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis using Structural MRI – *Supplementary Materials*

Chunfeng Lian[†], Mingxia Liu[†], Jun Zhang, and Dinggang Shen[*], *Fellow, IEEE*

1 Effectiveness of Voxel-Wise Correspondence for Location Proposals

As described in Section 4.1.1 of the main text, the definition of voxel-wise correspondence in the linearly-aligned image space is used for generating anatomically-consistent location proposals across different subjects in our network. To evaluate the effectiveness of this strategy, we implemented another version of our nH-FCN method, in which location proposals were directly determined in the template space, without further warping them onto the linearly-aligned image space.

TABLE S0
Results of AD classification obtained by our nH-FCN method without and with the definition of voxel-wise correspondence, respectively.

Method	ACC	SEN	SPE	AUC
Without voxel-wise corresp.	0.858	0.767	0.930	0.902
With voxel-wise corresp.	**0.878**	**0.805**	**0.935**	**0.938**

In Table S0, using the task of AD diagnosis as an example, we compare our nH-FCN method with its above variant that was implemented without the definition of voxel-wise correspondence. From Table S0, we can observe that more consistent location proposals (generated based on the definition of voxel-wise correspondence) indeed improves the diagnostic performance. It implies that, while CNNs could partially handle affine misalignment across different subjects (i.e., linearly-aligned brain sMRIs) during feature learning, more precise definition of consistent inputs are beneficial for improving the discriminative capacity of our network, especially the patch-level sub-networks.

C. Lian, M. Liu, J. Zhang, and D. Shen are with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. D. Shen is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea. (Email: dgshen@med.unc.edu).
∗ *Corresponding author.* † *Co-first authors.*
This work was supported by NIH grants (EB008374, AG041721, AG042599, EB022880). Data used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The investigators within the ADNI did not participate in analysis or writing of this study. A complete list of ADNI investigators can be found online.

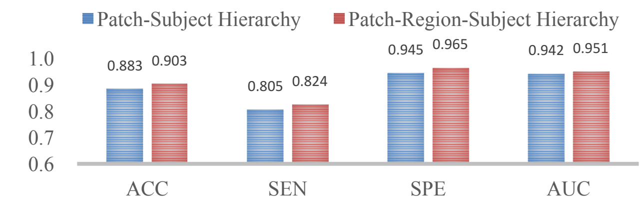

Fig. S1. Results of AD classification produced by our wH-FCN method with complete multi-scale architecture (i.e., Patch-Subject Hierarchy) and its variant without region-level sub-networks (i.e., Patch-Region-Subject Hierarchy).

2 Effectiveness of Hierarchical Architecture

To evaluate the effectiveness of our multi-scale hierarchy (i.e., integration of path-, region-, and subject-level sub-networks) strategy, we implemented another version of our wH-FCN, in which all patch-level sub-networks were combined and directly followed by the subject-level sub-network. In Fig. S1, this variant of wH-FCN (denoted as *patch-subject hierarchy*) is compared with the original wH-FCN (denoted as *patch-region-subject hierarchy*) on the task of AD diagnosis. From Fig. S1, we can observe that, integrating the region-level sub-networks between the path- and subject-level sub-networks effectively improves the final classification performance. These results indicate the importance of identifying multi-scale discriminative locations in brain sMR images, as we do in wH-FCN.

3 Influence of the Size of Regional Inputs

In the implementations of our H-FCN method, we integrated the outputs of spatially-nearest patches in a $2 \times 2 \times 2$ neighborhood to construct the subsequent region-level sub-networks. In this group of experiments, we tuned the size of regional inputs to evaluate its influence on the diagnostic performance. Specifically, we changed the architecture of our nH-FCN by grouping, respectively, $2 \times 2 \times 2$, $3 \times 3 \times 3$, and $4 \times 4 \times 4$ neighboring patches as a specific region. The quantitative results achieved by nH-FCN in the task of AD diagnosis are summarized in Fig. S2.

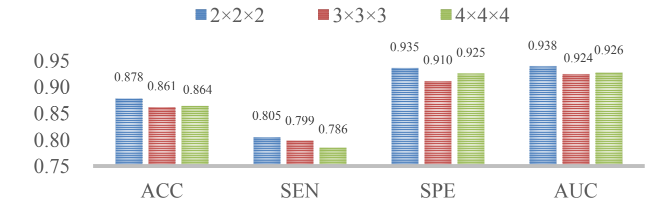

Fig. S2. Results of AD classification obtained by our nH-FCN method *in terms of* different sizes of inputs for the region-level sub-networks (i.e., $2 \times 2 \times 2$, $3 \times 3 \times 3$, and $4 \times 4 \times 4$ neighboring patches).

From Fig. S2, we can observe that constructing the regional inputs by combining the outputs for $2 \times 2 \times 2$ neighboring patches leads to relatively better performance, while further increasing the size of the regional inputs cannot bring additional improvement. It is perhaps due to 1) too large initial regions may increase the difficulty for identifying discriminative locations (at both the patch- and region-levels), considering uninformative location proposals are repeatedly counted in this way to amplify their negative influence (even when the network pruning step is followed), and 2) too large region may be not sensitive to subtle structural changes induced by the early stage of AD. On the other hand, it is worth noting that an appropriate size of regional inputs is also correlated with the size of image patches.

4 Generation of Voxel-wise AD Heatmaps

As mentioned in Section 5.12 of the main text, the proposed hierarchical fully convolutional network (H-FCN) can automatically identify hierarchical discriminative locations of brain atrophy at both the patch-level and region-level. Besides patch-level locations shown in Fig. 9 in the main text, we further analyze structural abnormalities discovered by our H-FNC method at a finer scale (i.e., at the voxel-level).

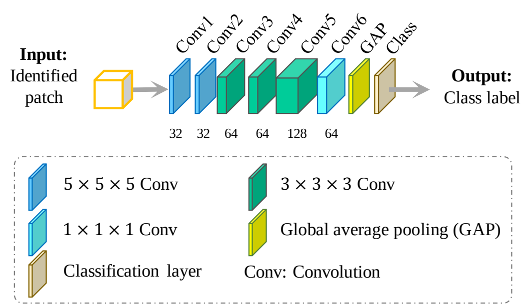

Fig. S3. Illustration of the 3D FCN for the generation of AD heatmaps based on the discriminative patch locations automatically identified by our H-FCN method.

To verify the feasibility of this claim, we extracted image patches (size: $25 \times 25 \times 25$) at the identified patch locations from the ADNI-1 dataset to train a 3D FCN, and then adopted the method proposed in [1] to generate the AD heatmaps for the corresponding patches extracted from the ADNI-2 dataset. The architecture of this 3D FCN is shown in Fig. S3. Specifically, it contains six convolutional (Conv) layers, including two $5 \times 5 \times 5$ layers (i.e., Conv1 and Conv2), three $3 \times 3 \times 3$ layers (i.e., Conv3 to Conv5), and a $1 \times 1 \times 1$ layer (i.e., Conv6). The number of channels for Conv1 to Conv6 is 32, 32, 64, 64, 128, and 64, respectively. To ensure that intermediate features have the same spatial resolution with the input patches for voxel-wise discriminative localization, all Conv layers are implemented with zero-padding and there is no pooling operations between them. Finally, the outputs of Conv6 are fed into a global average pooling layer (GAP), followed by a classification layer *without bias* (i.e., Class) to yield the classification score.

The network was trained on ADNI-1 with the Adam optimizer using the binary cross-entropy loss. After that, based on the weights of Class_P and the feature maps of Conv6, we adopted the operation proposed in [1] to generate high-resolution AD heatmaps for subjects from ADNI-2. Several illustrative examples have been presented in Fig. 10 of the main text, which suggested that, based on the discriminative patches localized by our H-FCN method, we could effectively identify more finer-scale discriminative locations at the voxel-level.

References

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 2921–2929.