A Parallel Attention-Augmented Bilinear Network for Early MRI-based Diagnosis of Alzheimer’s Disease - Supplementary Materials

Hao Guan\textsuperscript{a}, Chaoyue Wang\textsuperscript{a}, Jian Cheng\textsuperscript{b}, Jing Jing\textsuperscript{c}, Tao Liu\textsuperscript{b,d,}\textsuperscript{*}

\textsuperscript{a}School of Computer Science, Faculty of Engineering, The University of Sydney, Darlington, NSW 2008, Australia
\textsuperscript{b}Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing 100191, China
\textsuperscript{c}China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China
\textsuperscript{d}Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China

**Keywords:** Alzheimer’s disease, Convolutional Neural Network, Early diagnosis, Structural MRI, Visual Attention

Figure S1: MRI images of original size and cropped size. We choose the cropped size ($144 \times 176 \times 144$) to cover most areas of the brain, and reduce the useless computation in the background.

In Section 1, we present the examples of MRI images of original size and cropped size. In Section 2, we show the results of our method, the parallel attention-augmented bilinear network (pABN), with a deeper and wider backbone network. In Section 3, we show the models’ extents of overfitting by evaluating the trained models on the training dataset.

1. Examples of original MRI image and cropped image

We choose the cropped size ($144 \times 176 \times 144$) to cover the whole brain, and reduce the useless computation in the background. Here we show the original image and cropped image in the Figure S1. In addition, the cropped sizes are multiples of 8 to fit the down-sample operations of our neural network.

2. Results of pABN with a large backbone network

Our main goal is to devise an efficient and light-weight network, since the medical images are scarce and large networks are prone to overfitting on the small datasets. We provide the results of using a deeper and wider backbone network (BB) in Table S1. Compared to the BB used in the manuscript (3 residual units with 128 channels in maximum), the new BB uses an additional residual unit with 2 convolutional layers and 256 output channels, and we call this network as “BB256”. We test two networks on the basis of BB256: one uses BB256, followed by global average pooling and a fully-connected layer; the other one uses BB256 and pA-blocks, followed by bilinear pooling and a fully-connected layer. We also evaluate the 3D ResNet (Korolev et al., 2017) and nH-FCN (Lian et al., 2018) (both with 128 channels in maximum). The training procedures are the same as those used in the manuscript.

Table S1 shows that: 1) although the 3D ResNet (Korolev et al., 2017) and nH-FCN (Lian et al., 2018) have almost 4 times as many parameters as ours, our model still achieves the best performance; 2) increasing the number of layers and channels only brings marginal improvement to the “BB256 + GAP” model, but no contribution to the “BB256 + pA-blocks (A) + Bili” model; 3) Our pA-blocks with BB256 still largely surpass the other methods; 4) our methods with BB and BB256 show similar results, and the lower accuracy of the “BB256 + pA-blocks (A) + Bili” model might result from the overfitting (see the following part). The results suggest that our method is more efficient than the counterparts. Intuitively, large networks should work better, but only when enough training data are available.

3. Models’ extents of overfitting

We also show the models’ extents of overfitting by evaluating the trained models on the training dataset without data augmentation. The differences between the results on the training dataset and the test dataset can be used as a measure of overfitting. The bigger differences suggest the more overfitting of the model on the training dataset. Table S2 shows that our

\*Correspondence: Tao Liu, School of Biological Science and Medical Engineering, International Research Institute for Multidisciplinary Science, IRC 300, Beihang University, Beijing 100191, China. Email: tao.liu@buaa.edu.cn

Preprint submitted to Human Brain Mapping

September 15, 2021
Table S1: Results for AD classification (AD vs. NC) using baseline sMRI. The models were trained on the ADNI-1 dataset and evaluated on the ADNI-2 dataset.

| Model                  | Channels in max | Params (in millions) | ACC×100% (std.) | SEN×100% (std.) | SPE×100% (std.) | AUC (std.) |
|------------------------|-----------------|----------------------|-----------------|-----------------|-----------------|------------|
| 3D ResNet [Korolev et al., 2017] | 128             | 3.201 M              | 84.10 (0.33)    | 76.13 (2.91)    | 91.28 (2.27)    | 0.9209 (0.0054) |
| nH-FCN [Lian et al., 2018] | 128             | 3.129 M              | 86.36 (0.24)    | 85.94 (1.49)    | 86.75 (1.78)    | 0.9255 (0.0024) |
| BB + GAP               | 128             | 0.710 M              | 76.94 (0.63)    | 70.32 (4.72)    | 82.91 (4.54)    | 0.8384 (0.0073) |
| BB256 + GAP            | 256             | 1.661 M              | 87.53 (0.96)    | 78.97 (4.13)    | 78.14 (4.96)    | 0.8553 (0.0095) |
| BB256 + pA-blocks (A) + Bili | 256         | 1.738 M              | 88.56 (0.66)    | 87.10 (2.71)    | 89.88 (3.13)    | 0.9359 (0.0062) |
| BB + pA-blocks (A) + Bili (ours) | 128          | 0.751 M              | **90.70** (0.37) | **88.77** (1.20) | **92.44** (1.22) | **0.9358** (0.0049) |

“BB” refers to the backbone network; “BB256” refers to a deeper backbone network with 256 channels; “GAP” refers to global average pooling; “Params” refers to the number of convolutional weights (in millions).

Table S2: Results for AD classification on the training dataset (ADNI-1) and the test dataset (ADNI-2), and the differences of these two results.

| Model                  | Channels in max | Params (in millions) | ACC in Train | ACC in Test | ACC Diff | SEN in Train | SEN in Test | SEN Diff | SPE in Train | SPE in Test | SPE Diff | AUC in Train | AUC in Test | AUC Diff |
|------------------------|-----------------|----------------------|--------------|-------------|----------|--------------|-------------|----------|--------------|-------------|----------|-------------|-------------|----------|
| 3D ResNet [Korolev et al., 2017] | 128             | 3.201 M              | 95.47        | 84.10       | 11.37     | 96.16        | 76.13       | 10.33    | 94.87        | 91.28       | 3.59     | 0.9855      | 0.9209      | 0.0646   |
| nH-FCN [Lian et al., 2018] | 128             | 3.129 M              | 87.32        | 86.36       | 0.96      | 96.68        | 85.94       | 10.75   | 79.19        | 86.75       | -7.56    | 0.9689      | 0.9255      | 0.0434   |
| BB256 + GAP            | 256             | 1.661 M              | 90.62        | 78.53       | 12.09     | 99.88        | 78.97       | 20.91   | 82.54        | 78.14       | -4.40    | 0.9981      | 0.8553      | 0.1428   |
| BB256 + pA-blocks (A) + Bili | 256         | 1.738 M              | 98.48        | 88.56       | 9.91      | 98.91        | 87.10       | 11.81   | 98.09        | 89.88       | 8.21     | 0.9962      | 0.9339      | 0.0623   |
| BB + pA-blocks (A) + Bili (ours) | 128          | 0.751 M              | **90.48**    | 90.70       | **-2.22** | **96.37**    | **88.77**   | **7.60** | **81.62**    | **92.44**   | **-10.82** | 0.9682      | 0.9358      | 0.0324   |

“BB” refers to the backbone network; “BB256” refers to a deeper backbone network with 256 channels; “GAP” refers to global average pooling; “Params” refers to the number of convolutional weights (in millions); “Di” refers to the difference value between the results on the training and test datasets.

method (BB + pA-blocks (A) + Bili) has the lowest differences than all the counterparts. And the “BB256 + pA-blocks (A) + Bili” model has generally lower differences than the “BB256 + GAP” model. This suggests that the risk of overfitting is effectively reduced by our method. The “BB256 + pA-blocks (A) + Bili” model has bigger differences than the nH-FCN [Lian et al., 2018], which has 128 channels in maximum. We conjecture that the network with more channels are more easily overfitted. Our pA-blocks with BB256 has larger differences than the “BB + pA-blocks (A) + Bili” model, which suggests that the larger models are more easily overfitted on the small dataset, and produce inferior test results (see Table S1).

References

Korolev, S., Safiullin, A., Belyaev, M., Dodonova, Y., 2017. Residual and plain convolutional neural networks for 3d brain mri classification, in: Proc. IEEE 14th Int. Symp. Biomed. Imag. (IEEE-ISBI), pp. 835–838. doi:10.1109/isbi.2017.7950647

Lian, C., Liu, M., Zhang, J., Shen, D., 2018. Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri. IEEE Trans. Pattern Anal. Mach. Intell., 1–URL: https://www.ncbi.nlm.nih.gov/pubmed/30582529 doi:10.1109/TPAMI.2018.2889096