Supplementary Material for CVPR 2022 paper #9084

Jinseong Jang Dosik Hwang*
School of Electrical and Electronic Engineering, Yonsei University

1. More studies for AD classification

This paper is supplementary material for CVPR 2022 paper #9098, which title is 'M3T: three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer'. We further introduce other experimental results of our proposed method including ROC curve, activation map of normal subjects, and other 3D medical image classification performance for Alzheimer's Disease (AD).

1.1. Performance Comparison of Receiver Operating Characteristic(ROC) curve

We compared M3T with conventional 3D classification methods and visualize the ROC curve plot with the area under curve (AUC) values. The conventional methods include 3D ResNet (50, 101, 152) [5], 3D DenseNet121 [6], , I3D [2] , MRNet [1], and MedicalNet [4] used in our main material. We also combined some 3D CNN networks with a transformer like the model used in the main material.

The ROC graphs are presented in Fig. 3. The results show that our proposed M3T (black line) achieves the highest performance of AUC value in all test datasets. Especially, the performance differences in AIBL and OASIS between M3T and the other models are more than those in the ADNI dataset, which indicates the proposed model is strong against overfitting to the training dataset. Furthermore, the curves of our method are closed to the ideal graphs where the AUC value is 1, which means that it has higher sensitivity and specificity values than the other methods in classifying Alzheimer's Disease.

1.2. Visualization results of both AD and Normal subjects

Using the same method with the main material [3], we also visualize the activated area of our M3T network. Fig. 1 shows an AD and normal control (NC) related heatmap in 3D MRI template images. The heatmap from both cases focuses mainly on the hippocampus area of the coronal plane and the ventricle region of the axial domain. However, compared to the AD cases, the heatmap from NC cases activates a wider region of the brain. It can be seen that the analysis for the NC cases is made by focusing a wide area around the AD-related brain region.

*Corresponding author.

Model Name	AUC of ADNI	AUC of AIBL	AUC of OASIS
R152+T	0.9330±0.0123	0.8875±0.0103	0.8601±0.0170
D201+T	0.9567±0.0120	0.9038±0.0106	0.8629±0.0084
MRNet	0.9456±0.0091	0.8942±0.0114	0.8643±0.0131
I3D	0.9162±0.0095	0.8608±0.0293	0.8406±0.0168
MedicalNet	0.9557±0.0104	0.8992±0.0138	0.8488±0.0067
M3T	0.9639±0.0055	0.9276±0.0097	0.8903±0.0059

Table 1. Comparisons of cross-validated trained models.

Model	ResNet152+T	ResNet101-T	I3D	MedicalNet
PN	122.96M	90.75M	12.30M	46.19M
FLOPs	194G	144G	145G	208.4G
Model	DenseNet201+T	MRNet	M3T	M3T-Small
PN	30.95M	24.75M	29.12M	28.96M
FLOPs	119G	107G	717G	151G

Table 2. Comparisons of model complexity. T: Transformer.

1.3. Visualization results comparison to baseline model

Fig. 2 shows the average activation map of all AD cases of the baseline (ResNet152 + Transformer). The heatmap of the baseline network does not strongly focus hippocampus which is one of the most important areas to analyze AD. Furthermore, the areas not significantly related to AD are activated in the axial image. We will add the results of various baseline models to the supplementary material.

1.4. Cross Validation

Table 1 shows the cross-validation results of M3T and other baseline networks. From the experiments, M3T achieves the best average performance and lowest deviation values in most values.

1.5. Computational complexity

Table 2 shows the parameter numbers and FLOPs of baseline models and M3T. The M3T has relatively high Flops compared to other baseline models. For a fair comparison, we newly designed the M3T-Small model with convolutional and channel number of 3D CNN block. The FLOPs of M3T-Small was similar to other baseline models. The differences in performance with the M3T model were slightly low, less than 0.01. This result shows that M3T

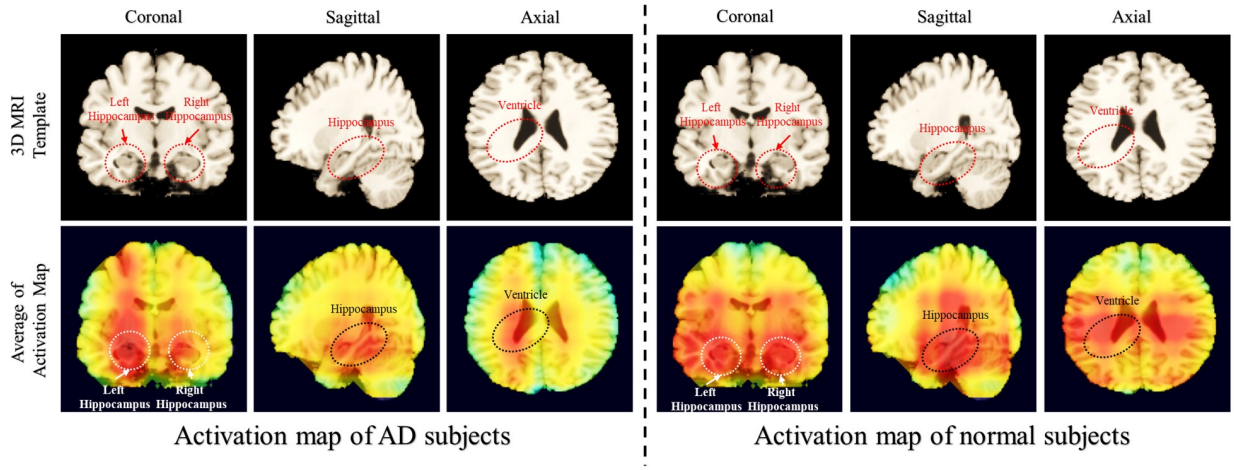

Figure 1. 3D MRI template images (first row) and average activation visualization map of all cases (second row). They consist of the visualization map of AD cases (left) and Normal cases (right)

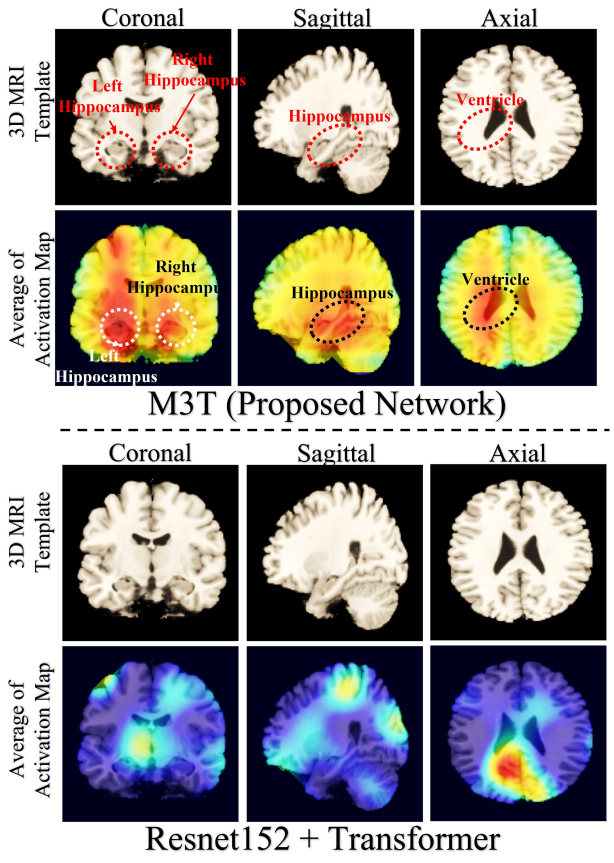

Figure 2. 3D MRI template images (first row) and average activation visualization map of all cases (second row) of M3T network, and the template images (third row) and activation map (fourth row) baseline (ResNet152 + Transformer) network

has sufficient competitiveness in terms of complexity if the model parameters are slightly changed.

1.6. Additional Ablation Study

2D ResNet-50 used in our 2D CNN network of M3T consists of 4 block levels [5]. To evaluate the degree to which the block depth of 2D CNN of the M3T network affects the performance, we compared the original M3T model with 3 models as follows: 1) M3T only using two block levels, 2) M3T using three-block levels and 3) M3T using whole blocks. The results show that M3T with the whole level ResNet Blocks has a higher performance than the other models. It represents that using entire models is very important to classify AD cases in the 3D MRI images. Using only some layers of the pre-trained 2D CNN network can degrade the classification performance. Furthermore, We experimented with freezing various layers of the pre-trained 2D CNN model, but learning all parameters on all layers achieved the best performance.

2. Feasibility of 3D CT classification

We also performed an additional study to check the feasibility of another task in another modality. We applied our methods to the classification of COVID-19 related abnormalities in 3D CT images. We have acquired a training dataset from MosMed-1111 dataset [7]. The number of the total training dataset is 889, including 224 normal control (NC) and 685 COVID-19 related cases. The test dataset includes a total of 221 cases which consist of 50 NC and 171 COVID-19 cases.

We performed pre-process to normalize and standardize CT images which consist of two processes: resizing process to same matrix size ($128 \times 128 \times 128$) and image intensity normalization of all the voxels using the zero-mean

AI Model		ADNI		AIBL		OASIS	
Name	Params	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
2 block levels	6.01M	0.9192	0.8762	0.8552	0.8947	0.8091	0.7807
3 block levels	12.57M	0.9455	0.9015	0.9008	0.9155	0.8439	0.8166
M3T (All block levels)	**29.12M**	**0.9634**	**0.9321**	**0.9258**	**0.9327**	**0.8961**	**0.8526**

Table 3. Quantitative comparison of AD classification using 3 different models of M3T to evaluate the depth of 2D CNN network.

AI Model		MosMed CT	
Name	Params	AUC	Accuracy
3D ResNet50	46.23M	0.7041	0.7738
3D ResNet50+TF	51.65M	0.7271	0.7828
3D ResNet101	85.33M	0.7328	0.7919
3D ResNet101+TF	90.75M	0.7054	0.8009
3D ResNet152	117.54M	0.7611	0.7964
3D ResNet152+TF	122.96M	0.7625	0.8009
3D DenseNet201	25.60M	0.7832	0.8100
3D DenseNet201+TF	30.95M	0.7887	0.8009
3D ViT	33.87M	0.6419	0.7964
MRNet	24.75M	0.7757	0.8009
I3D	12.30M	0.7439	0.8145
MedicalNet	46.19M	0.7804	0.8054
M3T (Ours)	**29.12M**	**0.8269**	**0.8190**

Table 4. Comparison with various 3D classification networks on 3D CT images for COVID-19 related abnormality.

unit-variance method. Only the two preprocessing methods were simply used to evaluate the classification performance of the deep learning algorithms. All of the algorithm and evaluation processes are the same as those of the AD classification in the main material.

The quantitative performance is presented in Table 4 which shows AUC, Accuracy values of COVID-19 related abnormality classification. M3T achieves the highest values of the metrics compared to the other methods. Like the results in the main material, the 3D ViT has lower performance than that of the other algorithms. Although the network achieves high performance in the experiments using a very large database, the pure-transformer networks obtain low performance in our experiment with a small amount of data. On the other hand, Our proposed M3T using a hybrid network achieves competitive performance in the low amount of 3D medical images.

References

[1] Nicholas Bien, Pranav Rajpurkar, Robyn L Ball, Jeremy Irvin, Allison Park, Erik Jones, Michael Bereket, Bhavik N Patel, Kristen W Yeom, Katie Shpanskaya, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet. *PLoS medicine*, 15(11):e1002699, 2018. 1

[2] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308, 2017. 1

[3] Hila Chefer, Shir Gur, and Lior Wolf. Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 782–791, 2021. 1

[4] Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019. 1

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2

[6] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 1

[7] SP Morozov, AE Andreychenko, NA Pavlov, AV Vladzymyrskyy, NV Ledikhova, VA Gombolevskiy, Ivan A Blokhin, PB Gelezhe, AV Gonchar, and V Yu Chernina. Mosmeddata: Chest ct scans with covid-19 related findings dataset. *arXiv preprint arXiv:2005.06465*, 2020. 2

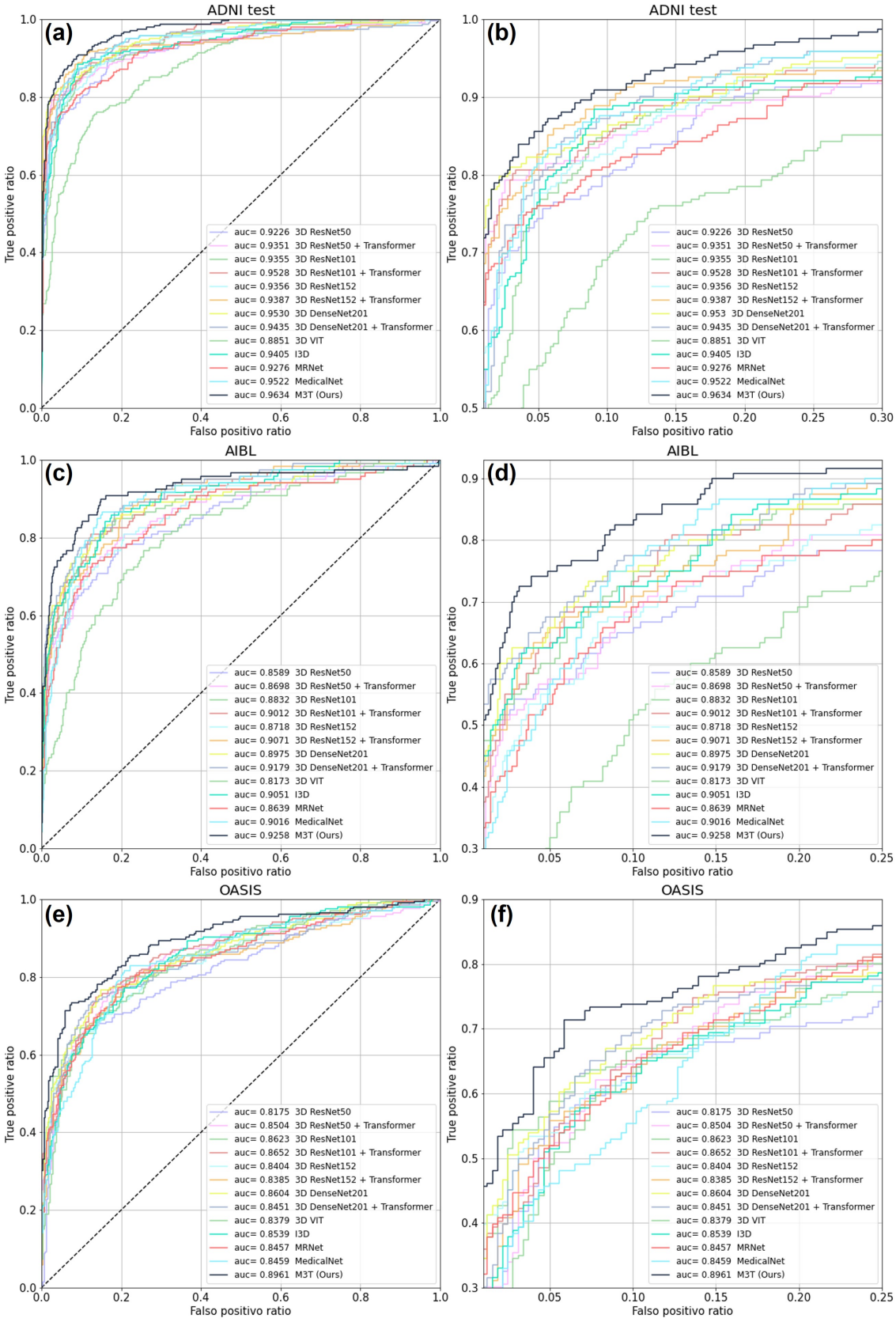

Figure 3. The Receiver Operating Characteristic (ROC) curve comparison graphs of the 3D classification model for Alzheimer's Disease. (a) : the roc curves and area under curve (AUC) on the ADNI test set. (b) : the tight graph of (a). (c) : the roc curves and area under curve (AUC) on the AIBL dataset. (d) : the tight graph of (c). (e) : the roc curves and area under curve (AUC) on the AIBL dataset. (f) : the tight graph of (e).