Performance of Ensemble Methods with 2D Pre-trained Deep Learning Networks for 3D MRI Brain Segmentation

Sang-il Ahn, Toan Duc Bui, Hyekyoung Hwang, and Jitae Shin

***Abstract*—Ensemble method has been shown a great success for 2D image segmentation, while 3D brain segmentation has received less attention using 2D pre-trained model. In this work, we present various 2D ensemble methods to utilize the 2D pre-trained models for the brain MRI segmentation task using given small medical 3D data. We perform a series of experiments by comparing several 2D single pre-trained models to build and analyze the various 2D ensemble methods. We evaluate the ensemble methods against 3D single scratch model in terms of accuracy, time, and crop size. In addition, we investigate the relationship between different compositions of train data and performance for semantic segmentation using MRBrainS18 train dataset. Experimental results demonstrate a significant improvement of the proposed ensemble method in comparison with existing methods using 3D CNN models for brain MRI segmentation.**

***Index Terms*—2D ensemble, pre-trained models, 3D small medical data, various composed train data, brain segmentation.**

I. Introduction

Brain segmentation plays an important role in medical image analysis. It aims to assign each pixel in the image into a class. However, it is a challenging task due to image artifact such as noisy, inhomogeneous, and low contrast among tissues. Nowadays, magnetic resonance imaging (MRI) technique considers as a potential way to solve the problem, because it provides a high dimensional data (i.e. 3D data), high tissue contrast.

In the past several years, many researches have been proposed to improve segmentation accuracy in brain MRI segmentation. With the success of deep learning, the deep learning-based methods [1]-[3] become a promising way for accurate segmentation in the brain MRI segmentation. However, these methods based on 3D convolutional neural network (CNN) often requires a training from scratch, so it is time-consuming and can't get the advantage of pre-trained model. In addition, 3D CNN methods often utilize image patches (i.e. crop a small region in the original image) which cause overfitting problem for training with a limited dataset as brain MRI dataset, because these methods have a small receptive field.

An alternative way for training the network with the small dataset is to fine-tune the network pre-trained using a large labeled dataset like ImageNet dataset. Nima [2] demonstrated the usage of a pre-trained model with adequate fine-tuning outperformed 3D data training from scratch for medical image application

Already some research focus on brain segmentation using ensemble method. Jose Dolz [4] studied infant brain MRI segmentation using deep CNN ensembles. K. Kamnitsas [5] proposed ensemble of multiple models and architectures for brain tumour segmentation. Both of these research base on 3D CNN models.

However, the performance of ensemble from different 2D pre-trained model using small 3D medical data for brain MRI has not been investigated. In this paper, we propose some efficient schemes for 3D brain MRI segmentation by comparing various ensemble methods based on state-of-the-art pre-trained 2D models using different training datasets with 3D scratch models. In MRI segmentation tasks using a small number of high dimensional data, the proposed method allows to exploit good features from 2D pre-trained models and to integrate them together.

Manuscript received February 5, 2019; revised April 23, 2019.

Sang-il Ahn, Toan Duc Bui, Hyekoung Hwang, and Jitae Shin are with the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea (e-mail: il2s@skku.edu, toanhoi@skku.edu, ristar1234@skku.edu, jtshin@skku.edu).

II. Materials and Methods

A. Datasets

We use MRBrainS18 images to evaluate our schemes. All MRBrainS18 images have a voxel size of 0.958mm x 0.985mm x 3.0mm and consist of T1, T1-IR, T2-FLAIR for each subject having multi-modalities. MRBrainS18 challenge provides 7 subjects train data (240 x 240 pixels), but test data was not provided. Thus we randomly pick up 1 for validation data and 6 remaining subjects for training data. MRBRainS18 images have 11 Classes (0: Back-ground, 1: Cortical gray matter, 2: Basal ganglia, 3: White matte, 4: White matter lesions, 5: Cerebrospinal fluid in the extracerebral space, 6: Ventricles, 7: Cerebellum, 8: Brain, stem, 9: Infarction, 10: Other). But we excluded classes 0, 9, 10 in the evaluation.

Most 2D pre-trained model was conducted using ImageNet which consists of 3 channels of RGB images. Thus 2D pre-trained models can get only 3 channels input. But MRbrainS18 dataset (T1, T1-IR, T2-FLAIR) have 48 channels for each modality. We perform a pre-process step to slice T1, T1-IR, T2-FLAIR voxel images into one channel image.

B. Input Data Composition

To improve the brain segmentation performance, we employ ensemble method based on 2D pre-trained models where we have small medical dataset. In order to use 2D models which is already trained with 3 channels images (i.e., RGB channel), we need to conduct a data that can be used as

doi: 10.18178/ijiee.2020.10.3.3

input data in 2D pre-trained models from 3D voxel dataset. Thus we create three kinds of input data from the 1 channel data. We first conduct a module which consists of two convolutional layers and two ReLU layers in order to transform from 1 channel to 3 channels with keeping the size. Secondly, we copy each channel data which is made of three sliced data respectively to make 3 channels data as input. Finally, we concatenate the each one channel data to use as input.

C. Ensemble with 2D Network Models

To compare performance of 3D scratch models and 2D ensemble, we introduce two ensemble methods using 2D models as Fig. 1. Ensemble A uses three 2D models simultaneously in training and validation. The parameters of the three models in training share the loss value and are updated concurrently to achieve better segmentation performance. Since three models are also used simultaneously in validation, Ensemble A produces one prediction value. On the other hand, Ensemble B method consist of three models which are trained separately, validation data enter each models as an input value in validation. The models which are trained independently output the prediction value with the each pixel value and average the each pixel prediction value from each model to make the final averaging prediction value.

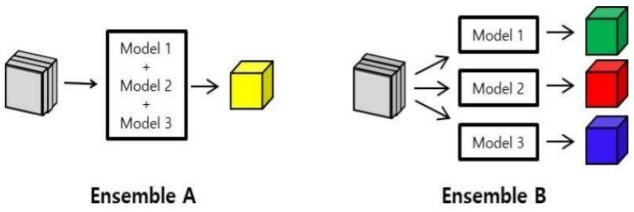

Fig. 1. Ensemble A uses three 2D models simultaneously in validation and produces one prediction value. Ensemble B uses three models separately in validation and generates three prediction values.

Therefore, Ensemble B can produce 3 output values in validation. This is similar to the way that several experts make a final decision after integrating each one's diagnosis.

D. Best Combination of Ensemble Methods

We experiment with several combinations by using three kinds of inputs for the ensembles in Fig. 2 to verify brain segmentation effect of input data shape. For Scheme 1, in training, we use three numbers of each channel data (i.e., each T1, T1-IR, T2-FLAIR slice) respectively. In order to make 3 channels input data, we input each one channel data into the module. And then, we can train 2D pre-trained models with the method of Ensemble A concurrently using output feature map of module.

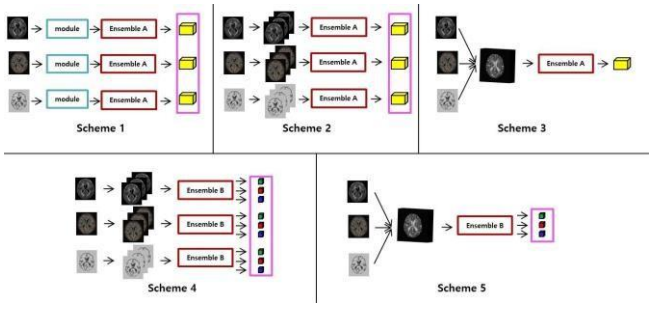

Fig. 2. Input data and ensemble combination.

In validation, since we also use three numbers of 1 channel data and the module, we can get three prediction values and compute segmentation performance by averaging three prediction values to make final prediction value. Scheme 2 and Scheme 3 also apply Ensemble A, but instead of adopting the module, Scheme 2 uses copied data and Scheme 3 uses concatenated data. In this process, Scheme 2 can make three prediction values and computes segmentation performance like Scheme 1, while Scheme 3 makes only one prediction value, thus Scheme 3 use the prediction value as final prediction value. In Scheme 4 and Scheme 5, we adopt Ensemble B with using copied data or concatenated data, respectively. The data enters the single 2D models as input value for each case to train the single models independently. After training, Scheme 4 and Scheme 5 can produce 9 and 3 prediction values where we use 3 copied and 1 concatenated validation data, respectively. Thus we predict the brain segmentation using each final prediction value which is averaged value of the 9 values or average of 3 numbers of each prediction value.

To compare with 3D models and 2D ensemble, we train 3D models using Adam with a learning rate of $5x10^{-4}$, momentum of (0.9, 0.99) method, mini-batch size of 4, and step size of 4000. The weight is initialized as in He *et al.* [3]. Each experiment was performed 7 times and averaged to guarantee reliability of the proposed method.

III. Experiment Results

A. Effects of Crop Size

In this subsection, we evaluate two size types (i.e., original size and 64x64 crop size) on 2D and 3D models to show the effect of the crop size. We use concatenated data as input and employ scratch model in 2D models, (i.e., no adoption from pre-trained model), to compare fairly with the 3D models in same environment. The batch size of the 2D models is set as same as the above 3D model. Dice coefficient is a metric to measure how much overlapped between prediction and ground truth segmentation. Table I show that no cropped image (i.e. full-resolution and original image) can consistently improve the performance comparing with scratch 2D models which use cropped image. Moreover, in some of the results, the 2D model with original image achieved better performance than the 3D models with cropped image. These experiments demonstrate that although we use 2D models and did not use image patch, overfitting problem did not occurred in limited medical dataset.

TABLE I: Segmentation Dice Accuracy According to Crop Size

Scratch \ Crop Size	Original (2D)	64x64(2D)	64x64(3D)
Resnet152	78.13	77.19	80.74
DenseNet121	80.37	79.05	80.26
Se-ResNet101	81.14	80.81	81.43
Xception	83.57	83.08	83.21
ResNext101(64x4)	82.94	82.35	82.65

B. 2D Pre-Trained Network

Before we demonstrate the application of the 2D ensemble

which consists of 2D pre-trained modes using 3D small dataset, we consider the performance of the 2D pre-trained models. In order to build the ensemble schemes in brain MRI segmentation, we validate several 2D pre-trained models(ResNet101, TarnausNet(Vgg16), DUC(ResNet152), DenseNet169, InceptionV4, Xception, ResNext101(64x4)) networks using pre-trained model and choose the networks of best top-three results in Table II. To evaluate each 2D pre-trained models performance, we build decoder to get same number of pixel values from model output with the number of pixel values of input because most 2D pre-trained model is built for classifier. Each of networks extracts feature maps composed of 1/2, 1/4, 1/8, 1/16, 1/32 of original input image size, performs upsampling to make feature maps of original input size, and concatenate all of them to make segmentation images. Each network was trained as same as above about 2D models except mini-batch size of 16, and step size of 2000.

TABLE II: SINGLE 2D PRE-TRAINED MODEL SEGMENTATION ACCURACY ON VALIDATION DATA

Network models	Dice Average	Training Time
ResNet101	82.24	11m 17s
DUC(ResNet152)	82.89	24m 28s
DenseNet169	84.52	13m 36s
Xception	84.87	22m 47s
ResNext101(64x4)	85.18	19m 56s
InceptionV4	85.39	15m 22s
TarnausNet(Vgg16)	86.04	10m 30s

C. Results of Various Ensemble Methods

Table III shows the comparison of different schemes in Fig. 2 via the segmentation results of 3D brain MRI. These results show that the ensemble scheme using the 2D pre-trained models can improve the performance in general. The result of Scheme 3 shows similar performance with the 3D top models, furthermore the Scheme 5 achieves a top performance where we use only single train data (concatenated 3 channels). In particular, we achieve more increased performance where we use copied data and concatenated data together such as (Scheme 2 +Scheme 3), (Scheme 4 +Scheme 5) in Table III. We note that although performance of some of the 2D ensemble method where we use only single train dataset (1 channel or copied 3 channels) was not increased as much as top performance of 3D models, while some 2D ensemble methods where we use the dataset (concatenated 3 channels and together with copied and concatenated) show better performance in the 3D brain segmentation in a relatively short time as we observe the Table III.

Especially in case of Scheme 3 or Scheme 5 which use concatenated 3 channels input, the performance was increased a lot, which shows that the ensemble using concatenated 3 channels input has more influence on the ensemble methods. This suggests that constructing input data with each channel which have different values effects further performance improvement when we can be able to build the input data.

TABLE III: SEGMENTATION ACCURACY ON VALIDATION DATA USING DICE COEFFICIENTS

Networks	Class1	Class2	Class 3	Class4	Class5	Class 6	Class7	Class 8	Dice Average	Training Time (hour-min-sec)	Crop Size
Scheme 1	83.12	83.45	83.37	80.66	78.93	92.61	93.11	85.85	85.14	1h 11m 52s	Ori
Scheme 2	82.96	84.84	83.44	82.79	79.82	91.97	93.03	85.77	85.58	1h 23m 22s	Ori
Scheme 3	83.48	86.39	84.36	87.32	82.65	92.82	92.72	79.08	86.10	23m 25s	Ori
Scheme 4	82.86	83.66	83.33	84.90	78.72	91.99	93.11	85.19	85.47	1h 35m 56s	Ori
Scheme 5	85.04	84.14	85.07	86.12	82.14	93.21	93.13	86.04	86.86	45m 48s	Ori
Scheme 2+Scheme 3	84.37	83.88	84.40	87.19	83.00	92.54	92.92	85.65	86.74	1h 46m 27s	Ori
Scheme 4+Scheme 5	85.43	85.34	84.20	87.24	83.56	92.15	93.27	85.45	**87.08**	2h 21m 44s	Ori
3D DenseSeg	83.91	82.61	82.82	85.79	83.37	92.80	93.37	87.64	86.54	4h 6m 18s	Ori
3D U-net[6]	85.05	81.80	85.05	86.63	81.35	88.31	93.85	88.67	86.34	4h 47m 23s	64x64
3D Xception	82.49	81.97	81.58	80.42	80.34	89.93	89.00	79.94	83.21	14h 52m 49s	64x64
3D esNext101(64x4)	79.10	79.04	79.20	78.07	75.74	91.60	91.77	86.66	82.65	5h 6m 3s	64x64

IV. CONCLUSION

In this paper we proposed the ensemble methods with 2D pre-trained model to improve the segmentation performance of 3D brain MRI on small number of 3D medical dataset. We showed that the 2D ensemble method can improve the brain MRI segmentation. Especially, by using that several 2D pre-trained models were used in the ensemble, we can get better and faster results than when using a single 3D scratch model.

For further study, we plan to get more robust medical segmentation performance with fewer data through various experiments.

ACKNOWLEDGMENT

This research was supported partly by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2017R1D1A1B03031752) and partly by MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2018-0-01798) supervised by the IITP (Institute for Information & communications Technology.

REFERENCES

[1] T. D. Bui, J. Shin, and T. Moon, "3D densely convolution networks for volumetric segmentation," arXiv preprint arXiv:1709.03199, 2017.

[2] N. Tajbakhsh, J. Y. Shin, S. R Gurudu, R. T. Hurst, C. B. Kendall, M. B Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1299-1312, 2016.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[4] L. Wang, J. Yuan, D. Shen, B. Ismail, A. J. Dolz, and C. Desrosiers, "Deep cnn ensembles and suggestive annotations for infant brain mri segmentation," arXiv:1712.05319, 2017.
[5] E. Ferrante, S, McDonagh, M, Sinclair, N, Pawlowski, M, Rajchl, M. Lee, B. Kainz, D. Rueckert, B. Glocker, K. Kamnitsas, and W. Bai, "Ensembles of multiple models and architectures for robust brain tumour segmentation," Brain Tumour Segmentation(BRATS) 2017 competition, 2017.
[6] C. Ozgun, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3 u-net: learning dense volumetric segmentation from sparse annotation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 424–432, 2016.

Sang-il Ahn received the B.S. degrees in information security engineering from the University of BaekSeok in 2018. He is currently an master student in the Department of Electronic, Electrical and Computer Engineering, College of Information and Communication Engineering, Sungkyunkwan University, Republic of Korea. His research interests include machine learning of computer vision.

Toan Duc Bui received a B.S. degree from Hanoi University of Science and Technology, Vietnam, in 2012. He received the M.S. and Ph.D. degrees in electrical and computer engineering from Sungkyunkwan University, the Republic of Korea, in 2014 and 2017, respectively. He worked as a post-doc in Sungkyunkwan University.

Since 2019, he is working as a post-doc in UNC-Chapel Hill, US. His research interests include image processing and machine learning, with a special focus on medical image segmentation, semantic segmentation, and deep learning.

Hyekyoung Hwang received the BS degree from Sungkyunkwan University, mathematics and electrical and electronic engineering in 2018. She is currently an integrated (MS leading to PhD) student in the Department of Electronic, Electrical and Computer engineering, College of Information and Communication Engineering, Sungkyunkwan University, Republic of Korea. Her research interests include machine learning of computer vision.

Jitae Shin received his B.S. degree from Seoul National University in 1986 and his M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST) in 1988. After working at Korea Electric Power Corp. (KEPCO) and the Korea Atomic Energy Research Institute (KAERI), he returned to study and received M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1998 and 2001, respectively. Since 2002, he is a professor in the School of Electronic and Electrical Engineering of Sungkyunkwan University, Suwon, Republic of Korea. His current research interests include image/video signal processing, video transmission over wireless/mobile communication systems, and deep learning applications.