SUPPLEMENTARY MATERIALS

Deep learning-based segmentation method

We performed the Desikan-Killiany atlas-based freesurfer segmentation on 388 patients of Yeouido St. Mary's Hospital dataset as well as public datasets such as HCP, ADNI, PPMI, AIBL, and IXI, and two expert performed manual correction to produced fine-turned ground truth dataset. In addition, hypo-intensity region was added. The dataset was separated into three sets: training, validation, and testing. We first randomly shuffled the dataset and separated 49 patients for testing. The remaining patient's data were used for training and validation (9.5:0.5). The training data was constructed by extracting the three-dimension patch image using uniform sampling (96×96×96) for the individual ground truth data (Figure 1).

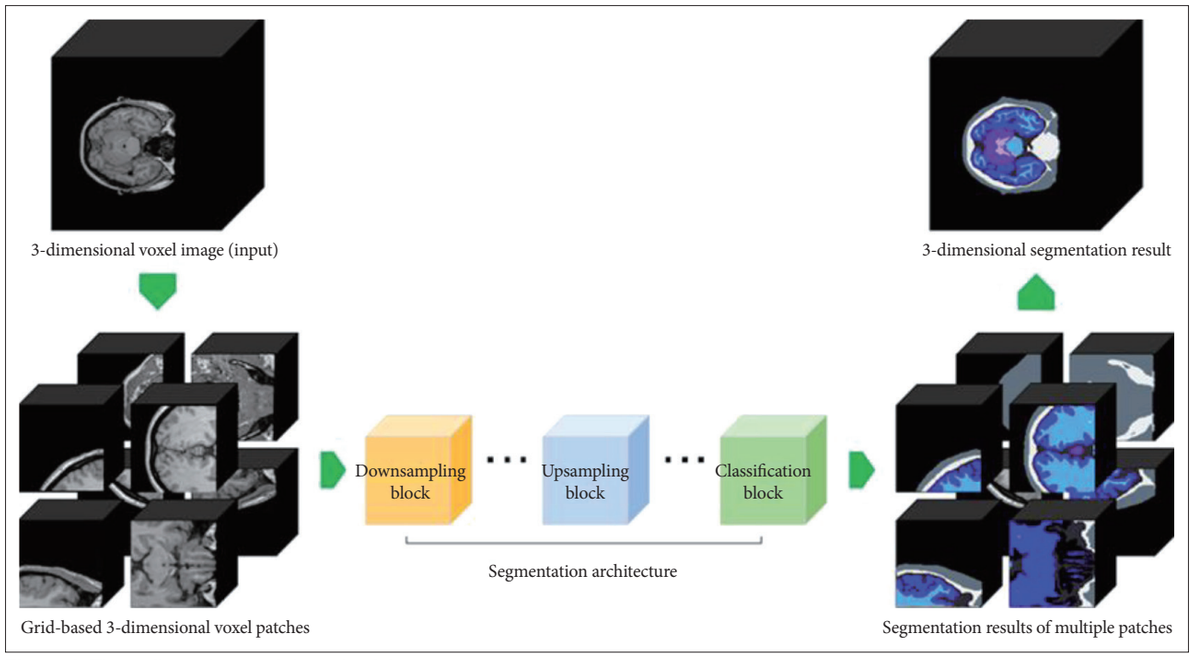

Figure 1. Three-dimension patch-based training.

We improved the UNet++ deep learning architecture with a three-dimension methodology to train about 104 labels. This algorithm has a convolutional layer in the skip path, which bridges the semantic gap between the encoder and decoder characteristic maps. There is a dense skip-connection in the skip path, which improves the gradient flow, has a deep supervision, which enables model pruning, improves performance, or at worst compares to using only one lossy layer. Performance can be achieved (Figure 2).

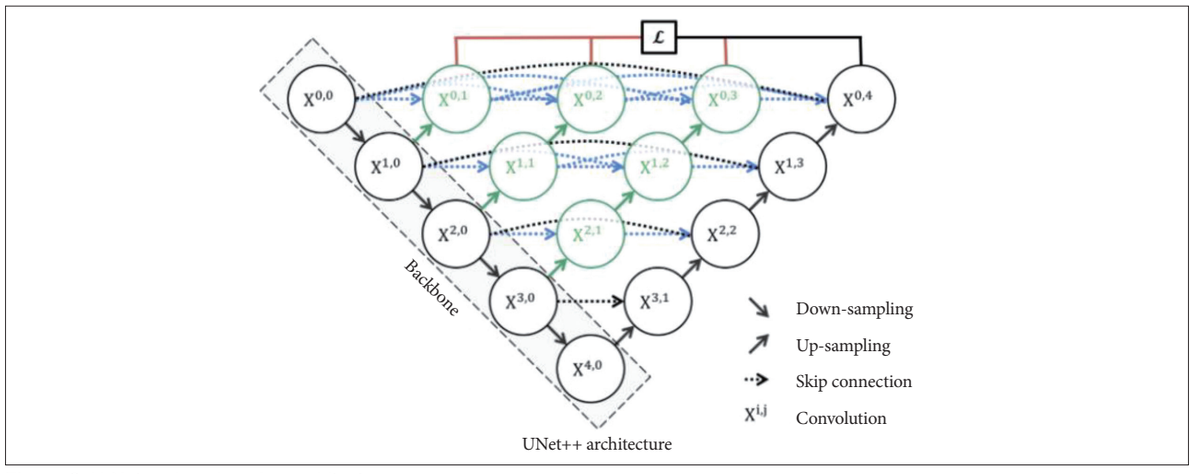

Figure 2. U-Net++ deep learning architecture.

Because the voxel by voxel segmentation learning method is used, the CrossEntropy loss function is used, and the learning rate for Adam optimizer is 0.0001. The total number of iterations is 300,000. Segmentation results are obtained by merging inference data using a three-dimension patch sliding aggregator. Figure 3 shows the segmentation result of brain sub-volumes.

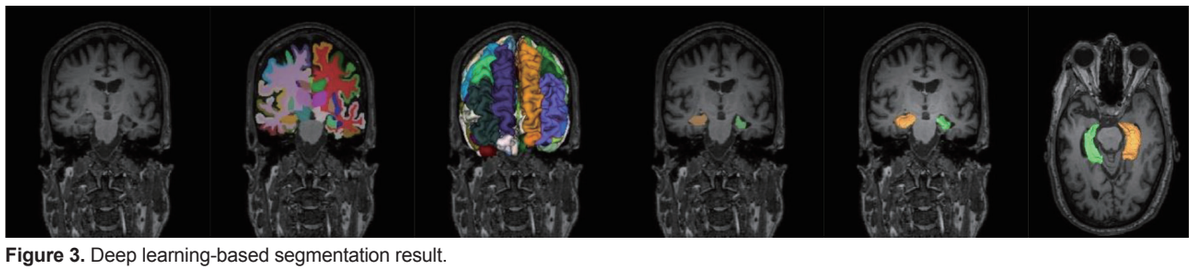

Figure 3. Deep learning-based segmentation result.

We perform the Dice overlap with the extra-validation set and, the average Dice coefficient is 0.840±0.083. Table 1 shows the whole Dice coefficients with the test set.

Table 1. Dice overlap result with the test set (49 case)

Case	Average dice 106 labels
0	0.706±0.262
1	0.791±0.186
2	0.866±0.085
3	0.873±0.067
4	0.857±0.068
5	0.877±0.066
6	0.822±0.100
7	0.857±0.070
8	0.856±0.072
9	0.765±0.113
10	0.768±0.099
11	0.766±0.108
12	0.779±0.088
13	0.758±0.106
14	0.829±0.098
15	0.886±0.052
16	0.874±0.053
17	0.886±0.057
18	0.853±0.084
19	0.873±0.065
20	0.876±0.063
21	0.875±0.068
22	0.876±0.061
23	0.877±0.060
24	0.872±0.061
25	0.881±0.058
26	0.863±0.102
27	0.871±0.071
28	0.883±0.058
29	0.84±0.075
30	0.861±0.064
31	0.825±0.067
32	0.837±0.097
33	0.851±0.080
34	0.839±0.097
35	0.822±0.103
36	0.867±0.065
37	0.865±0.060
38	0.794±0.101
39	0.758±0.109
40	0.864±0.060
41	0.789±0.089
42	0.878±0.059
43	0.883±0.060
44	0.752±0.102
45	0.855±0.092
46	0.859±0.069
47	0.863±0.064
48	0.854±0.059
Mean±std	0.840±0.083