2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

CAPPU-Net: A Convolutional Attention Network with Pyramid Pooling for Segmentation of Middle and Inner Ear Structures in CT Images

Geonsoo Kim[†], Bo-Soung Jeoun [†], Su Yang, Jin Kim, Sang-Jeong Lee, and Won-Jin Yi[*]

***Abstract*—We proposed CAPPU-Net, a modified U-Net with convolutional attention and pyramid pooling for the fully automated segmentation of facial nerve, cochlea, and ossicle in CT images. CAPPU-Net achieved higher accuracy in dice similarity coefficient, and predicted more precise masks than U-Net.**

I. INTRODUCTION

Accurate segmentation of facial nerve, cochlea, and ossicle in CT images is critical in surgical planning for middle or inner ear as providing the shape and relative location of the structures. However, manual segmentatioin of the structures requires considerable time and effort. In this study, we proposed a CAPPU-Net for the fully automated segmentation of facial nerve, cochlea, and ossicle in CT images.

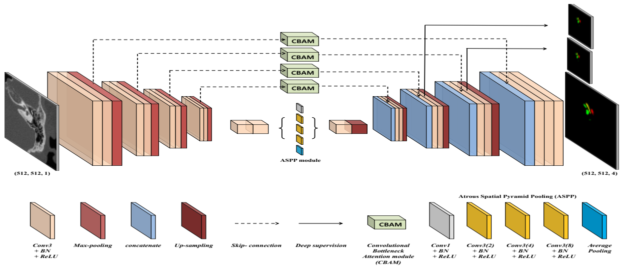

Fig. 1: CAPPU-Net architecture

II. METHODS

Data preparation. For this study, we used a total of 16346 axial CT images from 416 patients for training and testing. The number of training, validation, and test sets was randomly split into 10378, 2595, and 3373 images, respectively. Data augmentation was performed with rotation $(-10° - 10°)$, width/height shift$(0 - 10\%)$, and horizontal flipping.

Network architecture. The CAPPU-Net inspired by the vanilla U-Net employed advanced feature extraction blocks - convolutional bottleneck attention module (CBAM) and atrous spatial pyramid pooling (ASPP). The CBAM modules on each skip connection path refined intermediate feature maps effectively using channel and position attention maps [1]. The ASPP between the encoder and decoder exploited multi-scale features using dilated convolution [2]. In addition, auxiliary supervision branches were attached to the intermediate decoders to improve the convergence rate.

Training setup. We modified the loss function as a combination of categorical cross-entropy and logarithmic dice coefficient. To make the loss function more sensitive when the model weights are closer to saturation, we set the loss weights 0.75 and 0.25 on logarithmic dice coefficient and categorical cross-entropy, respectively. We employed Adam optimizer with a learning rate of 10^{-3} , and trained the proposed network for 150 epochs with a mini-batch size of 16.

Network	**DSC**		
	Facial nerve	**Cochlea**	**Ossicle**
U-Net	0.721 ± 0.231	0.841 ± 0.154	0.822 ± 0.207
CAPPU-Net	0.740 ± 0.214	0.848 ± 0.149	0.835 ± 0.194

TABLE 1: Comparison of U-Net and CAPPU-Net in DSC

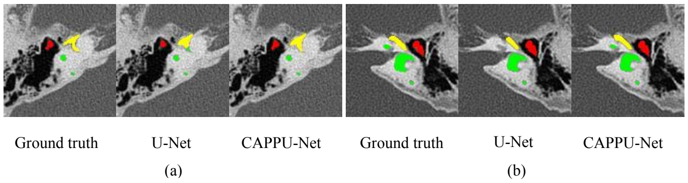

Fig. 2: Examples of ground truth and segmentation results.

III. RESULTS & CONCLUSION

We used the dice similarity coefficient (DSC) as the performance metric. The CAPPU-Net outperformed the U-Net for all three classes in the DSC (Table 1), and prediction masks. In Fig. 2. (a), the ossicle (red) is not fully recognized by U-Net, and a part of the facial nerve (yellow) is perceived as cochlea (green). Also, the U-Net losses a part of the cochlea in Fig. 2. (b). On the other hand, our network predicts the shape of structures more accurately with fewer false positives and false negatives. Therefore, CAPPU-Net achieved higher accuracy in dice similarity coefficient, and predicted more precise masks than U-Net.

References

[1] Woo, S., et al. "CBAM: Convolutional block attention module." *Proceedings of the European conference on computer vision (ECCV).* 2018.

[2] Chen, L, C., et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017): 834-848.

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (1711138289,KMDF-PR-20200901-0147)

G. Kim is with Medical Image Innovation Laboratory, Seoul National University, Seoul, Korea

B. S. Jeoun is with the Interdisciplinary Program in Bioengineering, Graduate School, Seoul National University, Seoul, Korea

W.J. Yi is with the Dep. of Oral and Maxillofacial Radiology, School of Dentistry, and the Dep. of Biomedical Radiation Sciences, GSCST, Seoul National University, Seoul, Korea (corresponding author to provide phone: (+82)2-2072-3049; e-mail: wjyi@snu.ac.kr).

[†]: Both authors contributed equally to this work.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE