Cross-dimension Affinity Distillation for 3D EM Neuron Segmentation Supplementary Material

Xiaoyu Liu[1] Miaomiao Cai[1] Yinda Chen[1] Yueyi Zhang[1,2] Te Shi[2]
Ruobing Zhang[3,2] Xuejin Chen[1,2] Zhiwei Xiong[1,2]

[1]MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China

[2]Institute of Artificial Intelligence, Hefei Comprehensive National Science Center

[3]Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences

liuxyu@mail.ustc.edu.cn, zwxiong@ustc.edu.cn

1. Architecture of the 2D Y-shape network

We present the detailed architecture of the 2D Y-shape network in Fig. 1, as introduced in Section 3 of our main paper. This network comprises an encoder and two decoder branches. The encoder initially extracts features independently from the two slices and then combines them using the concatenation operation, effectively capturing the inter-slice dependencies. The decoder subsequently decodes the fused features into two separate feature spaces, allowing for efficient affinity generation and cross-dimension distillation information interaction. This architecture design enables effective feature extraction and information fusion, enhancing the performance of our network in modeling 3D affinities using 2D networks.

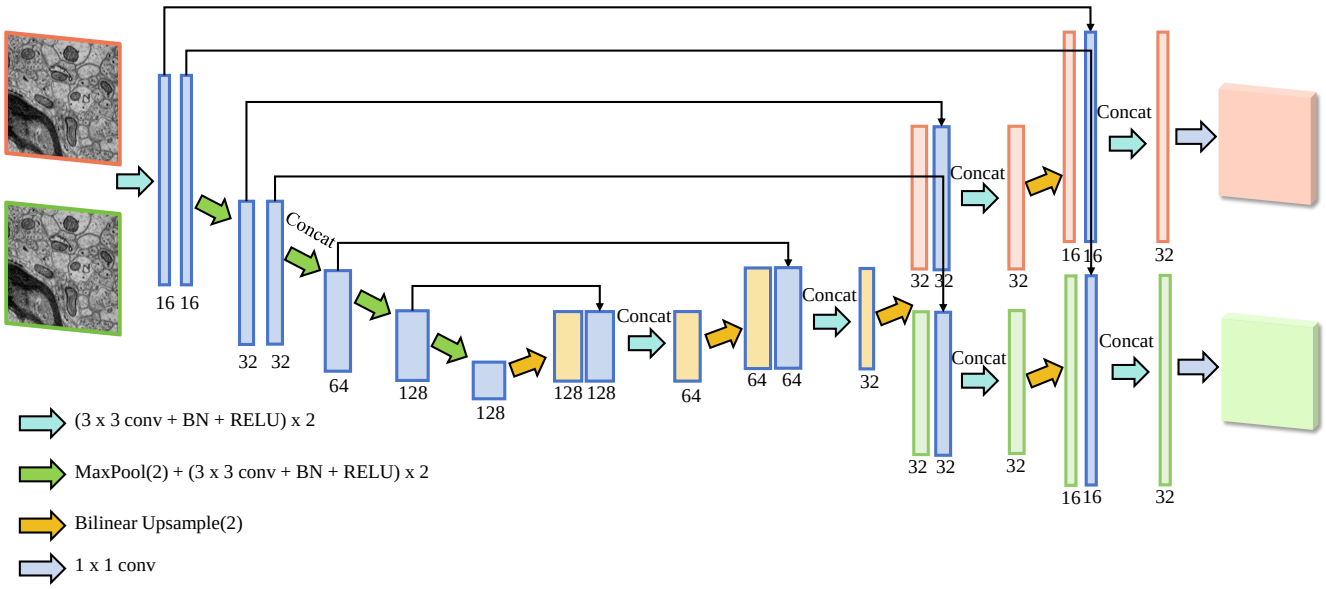

Figure 1. The detailed architecture of the 2D Y-shape network.

2. More qualitative results

To further demonstrate the effectiveness and reliability of our method in practical applications, we present the 3D visual comparison results of our built Wafer4 dataset in Fig. 2. These figures clearly illustrate that our proposed method accurately preserves the intricate 3D structures of neurons, surpassing the performance of other existing methods.

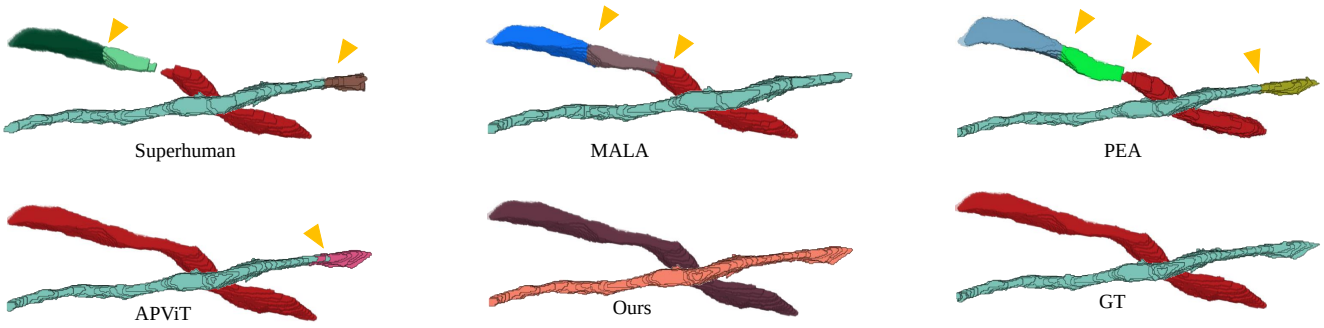

Figure 2. The 3D visual results on the Wafer4 dataset. The arrows indicate the segmentation errors in the 3D structure.

3. Analysis on the visualization of the embedding maps

As depicted in Fig. 3, we employ the PCA technique to project the embeddings of an EM section from a high-dimensional space onto a 3-dimensional RGB color space. The two embedding maps shown correspond to the outputs of the 3D CNNs and 2D CNNs, respectively.

When using 3D networks, a limitation arises due to their requirement of multiple 3D divided patches as input. This poses challenges when processing data on the 2D section plane. To overcome this limitation, we adopt a sliding window approach to extract 3D patches from the volume, which are then stitched together to obtain predictions for the entire section. However, this sliding window approach introduces additional errors, and the stitching process may result in artifacts at the boundaries of the patches. Moreover, predictions from different patches may not align seamlessly, adversely affecting the accuracy and coherency of the final predictions.

By projecting the embeddings onto a 3D RGB color space, we can observe that the embedding map generated by the 3D CNNs exhibits noticeable checkerboard-like artifacts. In contrast, our predicted embedding map using 2D CNNs demonstrates a clearer response at the target boundary and is not affected by the issues associated with sliding windows.

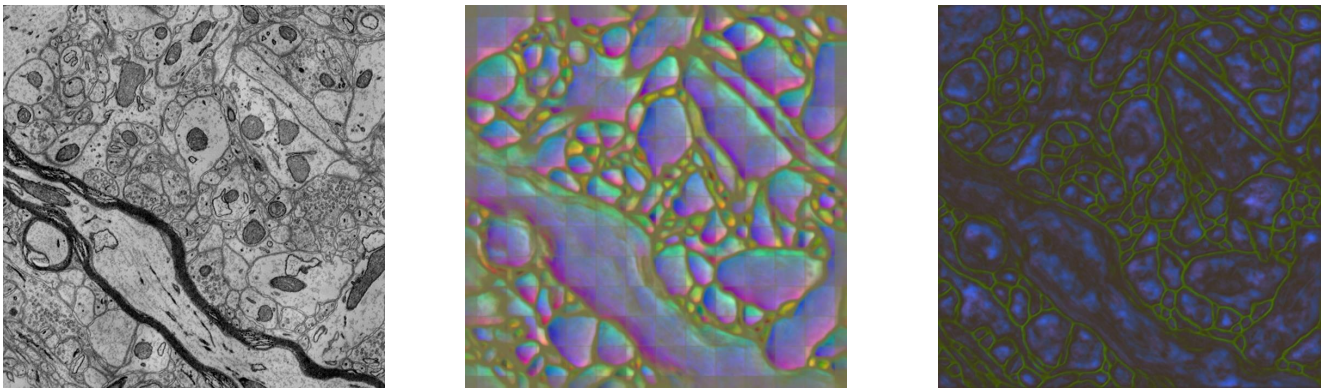

Figure 3. Embedding maps comparison between 3D CNNs and 2D CNNs