EMF-former: An Efficient and Memory-Friendly Transformer for Medical Image Segmentation(Supplementary material)

Zhaoquan Hao, Hongyan Quan[(✉)], and Yinbin Lu

School of Computer Science and Technology, East China Normal University, Shanghai, China
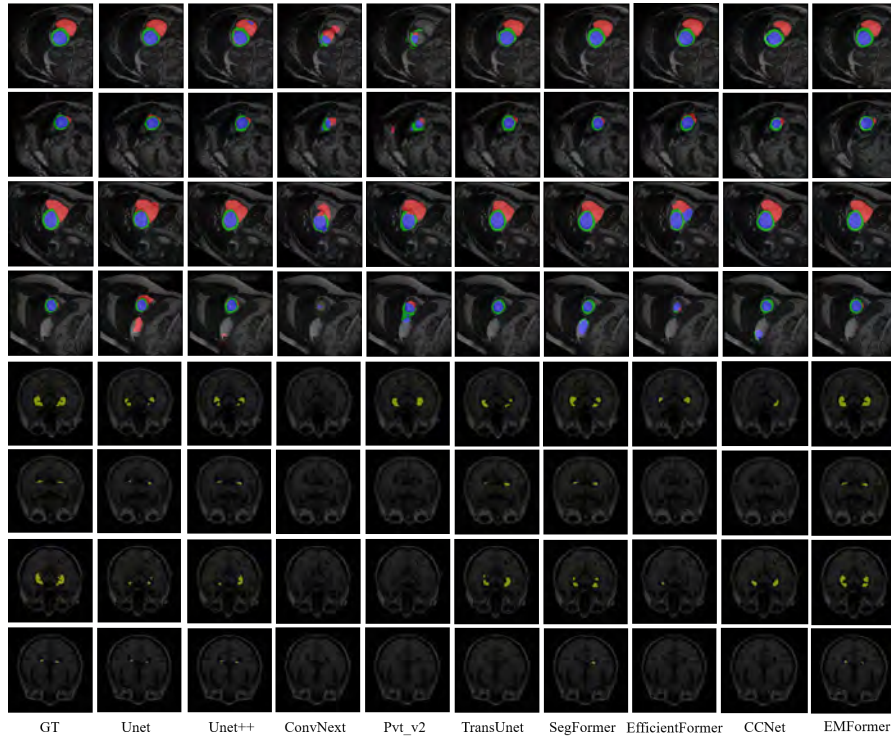hyquan@cs.ecnu.edu.cn

Fig. 1. Details of visual comparison. We selected two pairs of samples in both two dataset, each pair involving two segmentation results for large target and small target.
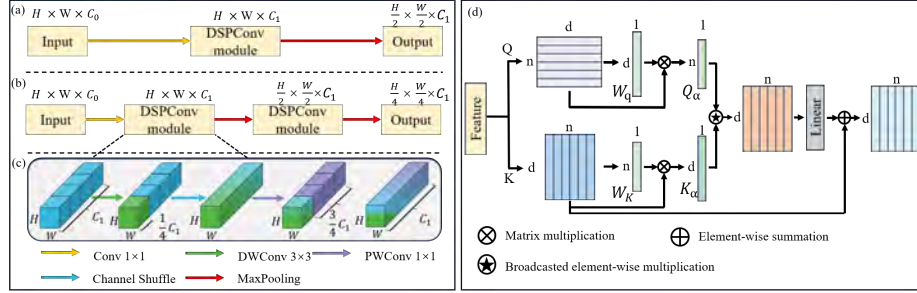
Fig. 2. Details of DSPConv Modules and VAA. (a) The DSPConv Merge structure is used in stage2~4. (b) The DSPConv Stem structure is used in stage 1. (d) The query and key matrix are multiplied by the learnable weights to produce a vector. The two vectors are then element-wise multiplied after the broadcast.
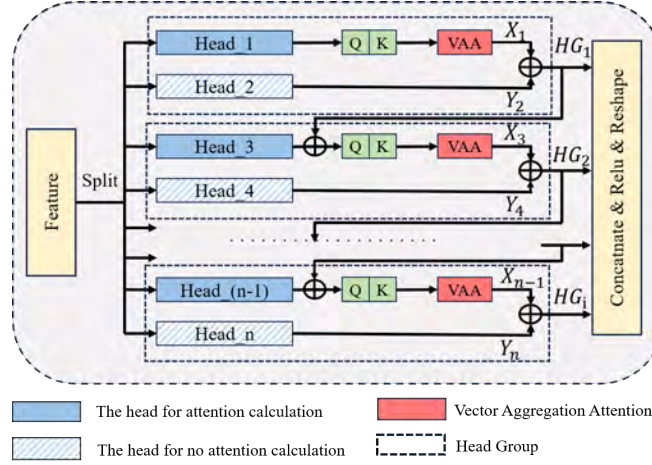

Fig. 3. Details of S-MHA. Multi-head are split from the features(We set the maximum value of n equal to 8). The head for attention calculation and the head for no attention calculation together form a Head Group (e.g., head_1 and head_2 in the figure).