Refining Intraocular Lens Power Calculation: A Multi-modal Framework Using Cross-layer Attention and Effective Channel Attention

Qian Zhou[1], Hua Zou[1, ✉], Zhongyuan Wang[1], Haifeng Jiang[2], and Yong Wang[2]

[1] School of Computer Science, Wuhan University, Wuhan, China
[2] Aier Eye Hospital of Wuhan University, Wuhan University, Wuhan, China
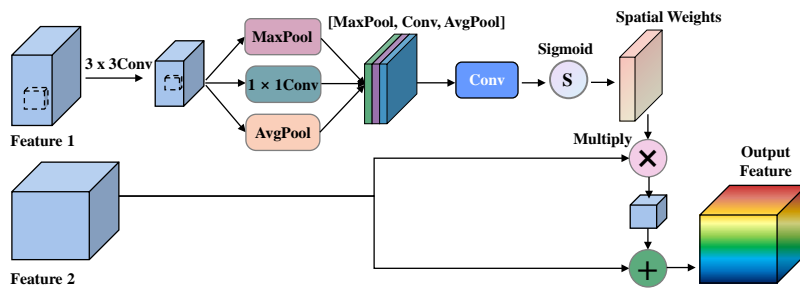zouhua@whu.edu.cn

Fig. 1. Detailed structure of cross-layer attention module (CLA). "Feature 1" is the features from the last layer, which is used to generate spatial weights. "Feature 2" represents the features from the current layer.
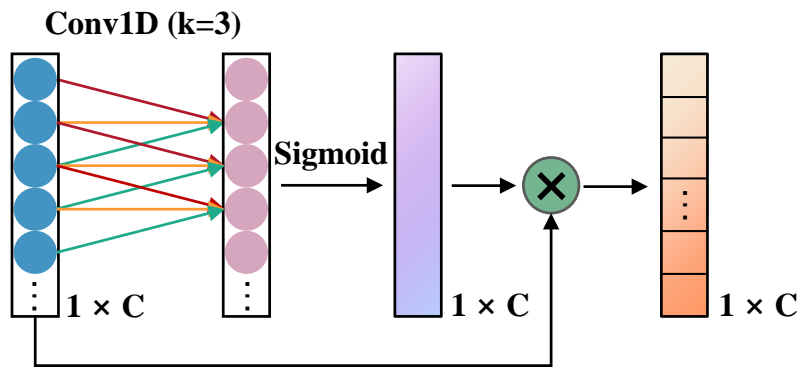

Fig. 2. Diagram of efficient channel attention (ECA) module. The input features and out features are 1D vectors.
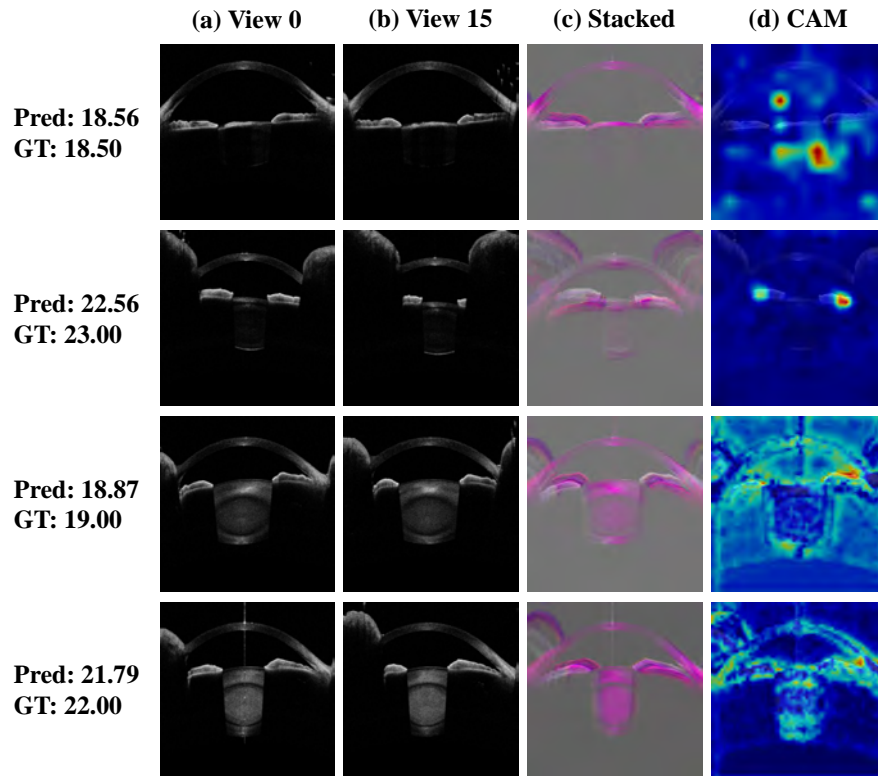
Fig. 3. Visualizations of input images and class activation maps (CAMs). (a) is the OCT images from view 0 and (b) is from view 15. (c) is the output of a Conv 3×3 before the backbone to stack 16 multi-view OCT images. (d) is the CAMs on view 0. Predict means the model's output, while GT is the ground truth.