1
-eps-converted-to.pdf

More Method Details for Frequency-Guided Network for Low-contrast Dental Plaque Segmentation Beyond Human Vision

We provide more method details as complementary to the paper's Method parts.

I. High-frequency augmentation algorithm

We conduct a sequence of operations in high-frequency augmentation, including hole-filling, opening, and dilation, to refine the HF-phase output teeth mask, as shown in Fig. 1.

The detailed principle of hole-filling. As teeth surfaces should not contain holes, hole-filling operation [1] helps fill in any holes segmented incorrectly within the teeth region. Hole-filling involves $Marker$, $Mask$ and a structuring element S_h. $Marker$ is the starting image of the transformation and will be continuously dilated until it converges. $Mask$ is used to constrain the dilation result, which means $Marker$ cannot dilate more than the $Mask$. The structuring element is used as the kernel of the dilation. The formula about $Marker$, $Mask$ and structuring element S_h is shown in Eq. 1.

$$Marker = (Marker \odot S_h) \cap Mask, \tag{1}$$

where $\odot$ represents dilation, S_h is structuring element, $\cap$ represents intersection operation.

The detailed principle of erosion. Erosion [1] can be formulated as follows:

$$\begin{aligned} (S_e)_x &= \{y \mid y = a + x, a \in S_e\}, \\ A \ominus S_e &= \{x \mid (S_e)_x \subseteq A\}, \end{aligned} \tag{2}$$

where A represents the initial mask, $(S_e)_x$ represents translating all elements of S_e by x units, $\ominus$ represents the erosion operator, and S_e is the structuring element. The erosion results in translation points make S_e still belong to A after translation. Erosion can also be regarded as calculating the minimum value of the pixel points in the area covered by structuring element S_e and assigning this minimum value to the pixel specified by the reference point.

The detailed principle of dilation. Dilation [1] can be formulated as follows:

$$A \odot S_d = \{x \mid (\hat{S_d})_x \cap A \subseteq A\}, \tag{3}$$

where A represents the initial mask, S_d is the structuring element, $(\hat{S_d})_x$ means the reversed $(S_d)_x$, $\odot$ represents dilation and $\cap$ represents intersection operation. Dilation can also be regarded as calculating the maximum value of the pixel points in the area covered by S_d and assigning this maximum value to the pixel specified by the reference point.

The detailed principle of opening operation. We perform an opening operation [1] to eliminate noise from the teeth region of the image. As the teeth area is typically a connected and sizable region, small noise could introduce redundant boundaries and adversely affect the LF phase. The opening operation involves erosion followed by dilation, where erosion can be regarded as assigning the minimum pixel value in the area covered by a structuring element to the specified reference point. We previously explained dilation in the hole-filling method. We perform the opening operation with a 30 × 30 structuring element to remove noise without distorting the main body of teeth. The opening operation can be formulated as follows:

$$A \circ S_o = (A \ominus S_o) \odot S_o, \tag{4}$$

where A represents the initial mask, $\circ$ represents the opening operator, and S_o is the structuring element.

II. Frequency-guided decoupling

The proposed method for disentangling the high-frequency part and providing independent supervision is illustrated in Fig.2.

III. Attention mechanism

The channel attention module. The channel attention maps [2] capture the inter-channel relationship of the features. The input feature map is subjected to global max pooling and global average pooling based on width and height, generating two feature maps with the size of 1×1, and the number of channels consists of the original feature map. Then, the two feature maps are sent to a shared two-layer neural network composed of multi-layer perception (MLP) with one hidden layer. The MLP contains two layers: $W_0 \in \mathbb{R}^{C/r \times C}$ and $W_1 \in \mathbb{R}^{C \times C/r}$, r is the reduction ratio, which is set to 16 in our network. The $ReLU$ activation function is followed by W_0. After that, the MLP output features are subjected to an element-wise addition, and then a $sigmoid$ function is performed to generate the final channel attention map.

The spatial attention module. The spatial attention module [2] generates the spatial attention map by using the inter-spatial relationship of the features. First, the input feature map is applied global max pooling and global average pooling based on the channel, yielding two feature maps in $\mathbb{R}^{1 \times H \times W}$. Then they are concatenated by channels, and the channel will be reduced to 1 after a convolution operation with a filter whose size is 7 × 7. At last, the module will generate the spatial attention feature through the $sigmoid$ function.

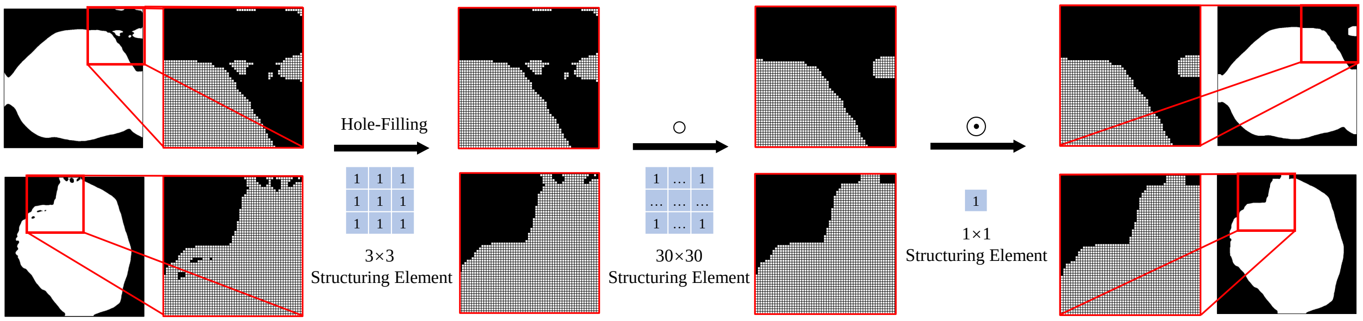

Fig. 1. The illustration of the high-frequency augmentation. We conduct the hole-filling operation, opening, and dilation sequentially. Of which, "○" denotes opening while "⊙" denotes dilation. Each small square in the figure represents a pixel.

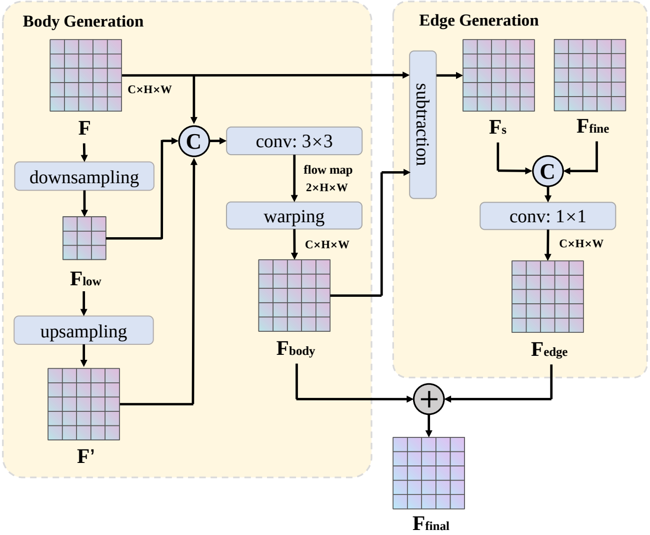

Fig. 2. The illustration of frequency-guided decoupling. Of which, "Ⓒ" denotes concatenation and "⊕" denotes element-wise addition.

Attention modules combination. Two possible methods of combining the channel and spatial attention modules are in parallel or in series. To illustrate their differences, we compare the parallel and series network structures in Fig. 3. After performing preliminary experiments, we include the channel and spatial attention modules as a residual block in parallel.

References

[1] R. C. Gonzalez, R. E. Woods *et al.*, "Digital image processing," 2002.
[2] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *ECCV*, 2018, pp. 3–19.

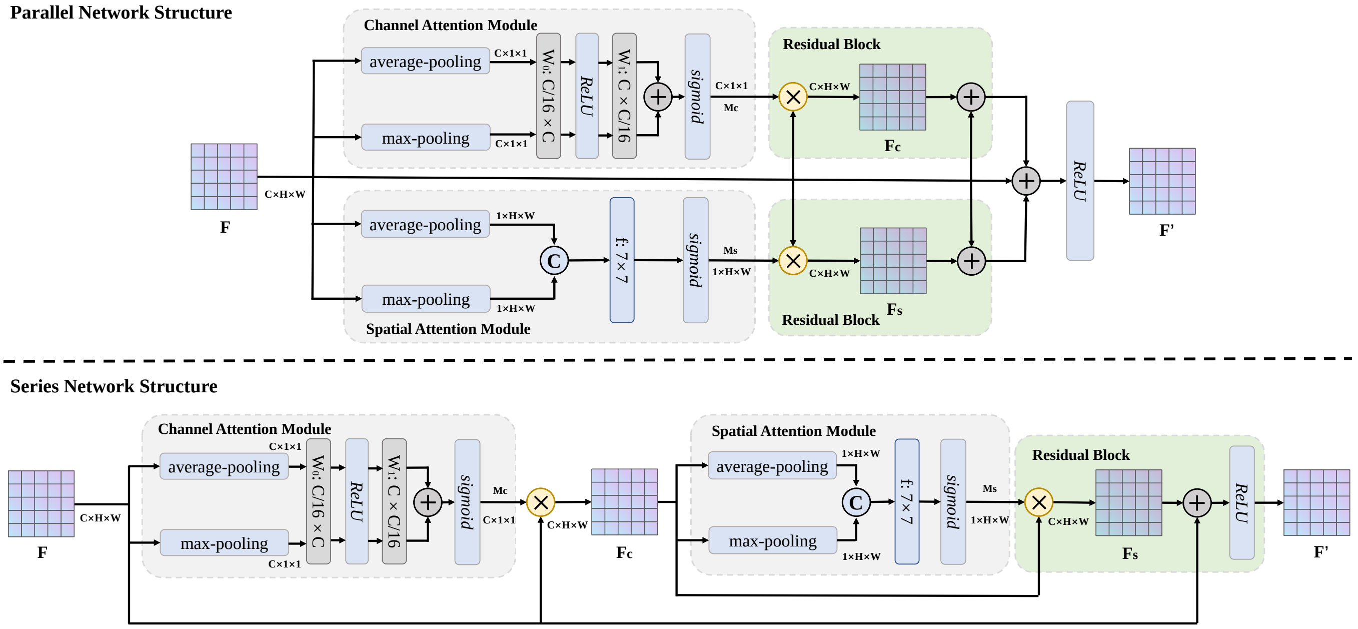

Fig. 3. The comparison of the parallel and series network structure of the attention mechanism in our FGN. We integrate channel and spatial attention modules into our method by adding them as residual blocks in parallel. Of which, the “⊕” symbol denotes element-wise addition, while the “⊗” symbol denotes the Hadamard product. The gray box indicates the convolution layer, while the blue box with a blue border signifies a convolution operation with a filter.