Physical-priors-guided Aortic Dissection Detection using Non-Contrast-Enhanced CT images (Supplementary)

Zhengyao Ding[1], Yujian Hu[2], Hongkun Zhang[2], Fei Wu[1], Shifeng Yang[4], Xiaolong Du[5], Yilang Xiang[2], Tian Li[3], Xuesen Chu[6(✉)], and Zhengxing Huang[1]

[1] Zhejiang University, China {zhengyao.ding,wufei,zhengxinghuang}@zju.edu.cn
[2] The First Affiliated Hospital of Zhejiang University School of Medicine, China {3170103999,1198050,21618130}@zju.edu.cn
[3] The Hong Kong Polytechnic University tianli@polyu.edu.hk
[4] Department of Radiology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, China ysfirst@126.com
[5] Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, China dd0341@163.com
[6] China Ship Scientific Research Center chuxs@cssrc.com.cn

1 The way physical information is combined with Unet

Our method was inspired by the way U-Net combines with timestep t in the diffusion model. In our approach, at each layer of the U-Net, we assume the shape of the image data is [batch, channel, depth, height, width], and the shape of the physical information data is [batch, 3]. We reshape the image data to [batch, channel, depth $\times$ height $\times$ width] and the physical information parameters to [batch, 1, 3]. Then we perform cross-attention where Q is the physical information data, and K and V are the image data. This enhances the model's ability to distinguish based on specific physical information.

2 Model Optimization Algorithm

Algorithm 1 Detailed Model Optimization Flow

Require: x: Input NCE-CT image; t: Output from transformer encoder; y_1: Real CE-CT image; y_2: Real segmentation image; $\tilde{y}_1$: Predicted CE-CT image; $\tilde{y}_2$: Predicted segmentation image; α: Learning rate; β_1, β_2: Adam optimizer parameters; λ_C: Classifier loss weight; λ_{L_1}: L_1 loss weight; λ_P: Perceptual loss weight; λseg: Dice loss weight; M: Batch size; ω: discriminators parameters; θ: classifier parameters; ζ: generator parameters.

for N training iterations **do**
    for M batches **do**
        Sample noise: $z \sim p_z$.
        Sample real images with class:$x, y_1, y_2, c \sim p_{data}$.
        $\tilde{y}_1, \tilde{y}_2 \leftarrow G_\theta(x, z), \tilde{y}_1$ is the synthesized CE-CT, $\tilde{y}_2$ is the predicted segment.
        $\mathcal{L}_D^{(i)} = logD_\omega(x, y_1) + log(1 - D_\omega(x, \tilde{y}_1))$.
        $\mathcal{L}_C^{(i)} = BinaryCrossEntropy(C_\zeta(t), c)$.
        $\mathcal{L}_G^{(i)} = log(1 - D_\omega(x, \tilde{y}_1)) + \lambda_{L_1}\|y - \tilde{y}_1\|_1 + \lambda_P \mathcal{L}^i_{Perceptual} + \lambda_C \mathcal{L}^i_C + \lambda_{seg}\mathcal{L}^i_{seg}(y_2, \tilde{y}_2)$.
    end for
    $\mathcal{L}_D = \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}_D^{(i)}$
    $\mathcal{L}_C = \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}_C^{(i)}$
    $\mathcal{L}_G = \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}_G^{(i)}$
    $\omega \leftarrow Adam(\nabla_\omega \mathcal{L}_D, \alpha, \beta_1, \beta_2)$
    $\zeta \leftarrow Adam(\nabla_\zeta \mathcal{L}_C, \alpha, \beta_1, \beta_2)$
    $\theta \leftarrow Adam(\nabla_\theta \mathcal{L}_G, \alpha, \beta_1, \beta_2)$
end for