# A Cascaded Multi-Task Generative Framework for Detecting Aortic Dissection on 3D Non-contrast-enhanced Computed Tomography

Xiangyu Xiong, *Member, IEEE,* Yan Ding, Chuanqi Sun, Zhuoneng Zhang, Xiuhong Guan, Tianjing Zhang, Hao Chen, Hongyan Liu, Zhangbo Cheng, Lei Zhao, Xiaohai Ma, and Guoxi Xie

In the supplementary material, we show that the 3D MTGA (in the original paper) can increase the Jensen-Shannon Divergence (JSD) between AD and non-AD for each NCE-CT volume, thus indirectly improving the AD detection performance.

The idea proposed in the 3D MTGA is to place the 3D classification sub-network ($C$) parallel with the discriminator $D$. $C$ takes the synthesized CE-CT volumes and NCE-CT volumes as input. The classification error is backpropagated through the generator $G$ and the $C$. We show that by placing the $C$ at the output of the $G$ and minimizing the binary cross-entropy as the classification loss, we can increase the JSD between non-AD and AD.

The objective function for the 3D MTGA is given by:

$$\mathcal{L}_{MTGA}(G,D,C) = \mathcal{L}_{cGAN}(G,D) + \mathcal{L}_{L_1}(G) + \mathcal{L}_{bce}(G,C) \quad (1)$$

where $\mathcal{L}_{bce}$ is the binary cross-entropy loss function. From the binary cross-entropy, (1) can be rewritten as:

$$\mathcal{L}_{MTGA}(G,D,C) = \mathcal{L}_{cGAN}(G,D) + \mathcal{L}_{L_1}(G) - \mathbb{E}_{z\sim pZ_1(z)}[log(C(x,G(x,z)_1))] - \mathbb{E}_{z\sim pZ_2(z)}[log(1-C(x,G(x,z)_1))] \quad (2)$$

which is given by:

$$\mathcal{L}_{MTGA}(G,D,C) = \mathcal{L}_{cGAN}(G,D) + \mathcal{L}_{L_1}(G) - \int p_{Z_1}(z)log(C(x,G(x,z)_1))dz - \int p_{Z_2}(z)log(1-C(x,G(x,z)_1))dz. \quad (3)$$

Considering $[x, G(x,z_1)_1] = \tilde{y}_1$ and $[x, G(x,z_2)_1] = \tilde{y}_2$:

$$\mathcal{L}_{MTGA}(G,D,C) = \mathcal{L}_{cGAN}(G,D) + \mathcal{L}_{L_1}(G) - \left\{ \int [p_{\tilde{y}_1}(\tilde{y})log(C(\tilde{y})) + p_{\tilde{y}_2}(\tilde{y})log(1-C(\tilde{y})]d\tilde{y} \right\}. \quad (4)$$

The function $f \leftarrow mlog(f) + nlog(1-f)$ reaches its maximum at $\frac{m}{m+n}$ for any $(m,n) \in \mathbb{R}^2 \setminus \{0,0\}$.

In 3D MTGA, each output of $C$ corresponds to one class $c$, that is, non-AD or AD. For a fixed $G$ and $D$, the optimal output of the $C$ for each class $c$ is

$$C^*_{G,D}(c) = \frac{p_{y_c}(\tilde{y})}{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}, c \in \{1,2\} \quad (5)$$

wherein $C^*_{G,D}(c)$ is the optimal classifier, and $p_{\tilde{y}_1}(\tilde{y})$ and $p_{\tilde{y}_2}(\tilde{y})$ are the distributions of the classifier's inputs for non-AD and AD, respectively.

From (4) and (5), we have

$$\mathcal{L}_{bce}(G,C^*) = -\int p_{\tilde{y}_1}(\tilde{y})log(\frac{p_{\tilde{y}_1}(\tilde{y})}{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})})d\tilde{y} - \int p_{\tilde{y}_2}(\tilde{y})log(\frac{p_{\tilde{y}_2}(\tilde{y})}{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})})d\tilde{y} \quad (6)$$

which is equal to

$$\mathcal{L}_{bce}(G,C^*) = 2log2 - \int p_{\tilde{y}_1}(\tilde{y})log(\frac{p_{\tilde{y}_1}(\tilde{y})}{\frac{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}{2}})d\tilde{y} - \int p_{\tilde{y}_2}(\tilde{y})log(\frac{p_{\tilde{y}_2}(\tilde{y})}{\frac{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}{2}})d\tilde{y} \quad (7)$$

which results in

$$\mathcal{L}_{bce}(G,C^*) = 2log2 - KL(p_{\tilde{y}_1}(\tilde{y})||\frac{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}{2}) - KL(p_{\tilde{y}_2}(\tilde{y})||\frac{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}{2}) \quad (8)$$

where KL is the Kullback–Leibler (KL) divergence. KL is always positive or equal to zero.

Considering the JSD between $p_{\tilde{y}_1}$ and $p_{\tilde{y}_2}$ is given by

$$JSD(p_{\tilde{y}_1}||p_{\tilde{y}_2}) = \frac{1}{2}KL(p_{\tilde{y}_1}(\tilde{y})||\frac{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}{2}) + \frac{1}{2}KL(p_{\tilde{y}_2}(\tilde{y})||\frac{p_{\tilde{y}_1}(\tilde{y}) + p_{\tilde{y}_2}(\tilde{y})}{2}). \quad (9)$$

From (8) and (9), we obtain

$$\mathcal{L}_{bce}(G,C^*) = 2log2 - 2JSD(p_{\tilde{y}_1}||p_{\tilde{y}_2}). \quad (10)$$

Minimizing the binary cross-entropy loss function $\mathcal{L}_{bce}$ increases the JSD between $p_{\tilde{y}_1}$ and $p_{\tilde{y}_2}$ for the $C$.

In the proposed multi-task generative mechanism, $\mathcal{L}_{L_1}$ is used for $G$ to ensure the quality of synthetic CE-CT volumes; $\mathcal{L}_{bce}$ is used for the $C$'s inputs to increase JSD between two classes (non-AD and AD). The $C$'s input comprises original NCE-CT volumes and synthetic CE-CT volumes. $\mathcal{L}_{bce}$ is used to generate features of non-AD or AD in synthetic CE-CT, which indirectly improves the classification task.