Unsupervised Domain Adaptation for Cardiac Segmentation: Towards Structure Mutual Information Maximization Supplementary Material

Changjie Lu, Shen Zheng, Gaurav Gupta
College of Science and Technology, Wenzhou-Kean University
Wenzhou, China
lucha,zhengsh,ggupta@kean.edu

1. Compact Lower Bound Proof

Proof of Eq.4:
Firstly, We follow the deduction from UDA-VAE [2].

$$\begin{aligned}&\log p_{\theta_S}(x,y)\\ =&\int q_{\phi_S}(z\mid x,y)\log\left[\frac{q_{\phi_S}(z\mid x,y)}{p_{\theta_S}(z\mid x,y)}\cdot\frac{p_{\theta_S}(z)}{q_{\phi_S}(z\mid x,y)}\cdot p_{\theta_S}(x,y\mid z)\right]dz\\ =&D_{KL}\left(q_{\phi_S}(z\mid x,y)\|p_{\theta_S}(z\mid x,y)\right)-\\ &D_{KL}\left(q_{\phi_S}(z\mid x,y)\|p_{\theta_S}(z)\right)+\\ &E_{q_{\phi_S}(z|x,y)}\log\left[p_{\theta_S}(x,y\mid z)\right]\end{aligned}\tag{1}$$

Note that UDA-VAE [2] neglects the term $D_{KL}\left(q_{\phi_S}(z\mid x,y)\|p_{\theta_S}(z\mid x,y)\right)$ as it is greater than 0.
In comparison, we deduce a compact lower bound with the following term.

$$\begin{aligned}&D_{KL}\left(q_{\phi_S}(z\mid x,y)\|p_{\theta_S}(z\mid x,y)\right)\\ =&\int q_{\phi_S}(z\mid x,y)\log\frac{q_{\phi_S}(z\mid x,y)}{p_{\theta_S}(z\mid x,y)}dz\\ =&\int\frac{q_{\phi_S}(x,y,z)}{q_{\phi_S}(x,y)}\log\frac{q_{\phi_S}(x,y,z)}{p_{\theta_S}(x,y,z)}\frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}dz\\ =&\frac{1}{q_{\phi_S}(x,y)}[\int q_{\phi_S}(x,y,z)\log\frac{q_{\phi_S}(x,y,z)}{p_{\theta_S}(x,y,z)}+q_{\phi_S}(x,y,z)\log\frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}dz]\\ =&\frac{1}{q_{\phi_S}(x,y)}\int q_{\phi_S}(x,y,z)\log\frac{q_{\phi_S}(x,y,z)}{p_{\theta_S}(x,y,z)}dz+\log\frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}\\ =&\frac{1}{q_{\phi_S}(x,y)}D_{KL}(q_{\phi_S}(x,y,z)\|p_{\theta_S}(x,y,z))+\log\frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}\\ \geq&D_{KL}(q_{\phi_S}(x,y,z)\|p_{\theta_S}(x,y,z))+\log\frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}\end{aligned}\tag{2}$$

Consider the reconstruction error [1]:

$$\mathcal{R}=\mathbb{E}_{(x,y,z)\sim}q_{\phi_S}(x,y,z)\log\frac{q_{\phi_S}(x,y,z)}{p_{\theta_S}(x,y,z)}-\mathbb{E}_{(x,y,z)\sim q_{\phi_S}(x,y,z)}\log q_{\phi_S}(x,y,z)+\mathbb{E}_{z\sim q_{\phi_S}(z)}\log p_{\theta_S}(z)\tag{3}$$

The second term is the joint entropy $H_q(x,y,z)$.
The third term can be written as:

$$\mathbb{E}_{z\sim q_{\phi_S}(z)}\log p_{\theta_S}(z)=-D_{KL}(q_{\phi S(z)}\|p_{\theta_S})-H_{q_{\phi_S}}(z)\tag{4}$$

With

$$H_{q_{\phi_{S(z)}}}(x,y,z) - H_{q_{\phi_S}}(z) = H_{q_{\phi_S}}(z) - I_{q_{\phi_S}}(x,y,z) \tag{5}$$

where I is mutual information.
The reconstruction error can be written as:

$$\mathcal{R} \leq D_{KL}(q_{\phi_{S(x,y,z)}} \| p_{\theta_{S(x,y,z)}}) - I_{q_{\phi_S}}(x,y,z) + H_{q_{\phi_S}}(z) \tag{6}$$

which is compact when $q_{\phi_{S(z)}}$ matches the prior distribution $p_{\theta_S}(z)$.

$$D_{KL}(q_{\phi_{S(x,y,z)}} \| p_{\theta_{S(x,y,z)}}) \geq \mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) \tag{7}$$

Thus, we obtain the bound,

$$\begin{aligned} &D_{KL}\left(q_{\phi_S}(z \mid x,y) \| p_{\theta_S}(z \mid x,y)\right) \\ &\geq D_{KL}(q_{\phi_S}(x,y,z) \| p_{\theta_S}(x,y,z)) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)} \\ &\geq \mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)} \end{aligned} \tag{8}$$

From, Eq.1 and Eq.8,

$$\begin{aligned} &\log p_{\theta_S}(x,y) \\ \geq&(\mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}) - D_{KL}\left(q_{\phi_S}(z \mid x) \| p_{\theta_S}(z)\right) + E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(x,y \mid z) \\ =&(\mathcal{R} + I_{q_{\phi_S}}(x,y,z) - H_{q_{\phi_S}}(z) + \log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}) - D_{KL}\left(q_{\phi_S}(z \mid x) \| p_{\theta_S}(z)\right) + E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(x \mid y,z) \\ &+ E_{q_{\phi_S}(z|x)} \log p_{\theta_S}(y \mid z) \end{aligned} \tag{9}$$

where R, $\log \frac{p_{\theta_S}(x,y)}{q_{\phi_S}(x,y)}$ and $H_{q_{\phi_S}}(z)$ are constant. The equation holds, as $p_{\theta_S}(x,y \mid z) = p_{\theta_S}(y \mid z) \cdot p_{\theta_S}(x \mid y,z)$. Meanwhile, y_S and z_s are conditionally independent on x_S for distribution q_{ϕ_S}, so that $q_{\phi_S}(z \mid x,y) = q_{\phi_S}(z \mid x)$.
Finally, We get the compact lower bound (plus red terms) than UDA-VAE .
The UDA-VAE++ maximizes the mutual information of $I_{q_{\phi_S}}(x,y,z)$.
Proved.

References

[1] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *International conference on machine learning*, pages 531–540. PMLR, 2018.

[2] Fuping Wu and Xiahai Zhuang. Unsupervised domain adaptation with variational approximation for cardiac segmentation. *IEEE Transactions on Medical Imaging*, 40(12):3555–3567, 2021.