Supplementary Material for Parameter Updates

First, let us assume the parameters of source and target regressor as $\theta_{reg}^{s}, \theta_{reg}^{t}$. For each domain, we can simply update these parameters using proper learning rate μ and regularization term γ as follows:

$$\theta_{reg}^{*d} = \theta_{reg}^{d} - \mu\gamma \frac{\partial \mathcal{L}_{reg}^{d}}{\partial \theta_{reg}^{d}} \tag{1}$$

Using chain rules, we can define the losses for the update of two encoding networks θ_{enc} as follows:

$$\theta_{enc}^{*d} = \theta_{enc}^{d} - \mu(\beta \frac{\partial \mathcal{L}_{enc}^{d}}{\partial \theta_{enc}^{d}} + \gamma \frac{\partial \mathcal{L}_{reg}^{d}}{\partial \theta_{reg}^{d}} \frac{\partial \theta_{reg}^{d}}{\partial \theta_{enc}^{d}}) \tag{2}$$

, where $\mathcal{L}_{enc}^{d} = ||E^{d} - I^{d}||_{F}^{2}$.

Using domain loss $\mathcal{L}_{dom} = a(\mathcal{L}_{com}^{s,dom} + \mathcal{L}_{spe}^{s,dom}) + (1-a)(\mathcal{L}_{com}^{t,dom} + \mathcal{L}_{spe}^{t,dom})$, the parameter of domain discriminator θ_{dom} can be updated as follows:

$$\theta_{dom}^{*} = \theta_{dom} - \mu\alpha \frac{\partial \mathcal{L}_{dom}}{\partial \theta_{dom}} \tag{3}$$

Since domain discriminator utilizes four types of inputs; one from source FE, the other from target FE, and another two features from common FE, we should consider them individually for the update of three FEs. Here, $a\mathcal{L}_{spe}^{s,dom}$ updates source FE and $a\mathcal{L}_{spe}^{t,dom}$ is used for target FE. Likewise, passing through gradient reversal layer (GRL), a negative constant is applied for $a\mathcal{L}_{com}^{s,dom} + (1-a)\mathcal{L}_{com}^{t,dom}$ that updates common FE.

Consequently, we can define the update functions for three FEs $\theta_{src,fe}$, $\theta_{com,fe}$, $\theta_{trg,fe}$ integrating three types of back-propagated losses:

$$\begin{aligned}
\theta_{src,fe}^{*} = \theta_{src,fe} - \mu \Bigg\{ & \alpha \frac{a\partial \mathcal{L}_{src}^{s,dom}}{\partial \theta_{dom}} \frac{\partial \theta_{dom}}{\partial \theta_{src,fe}} + \\
& \beta(\frac{\partial \mathcal{L}_{enc}^{s}}{\partial \theta_{enc}^{s}} \frac{\partial \theta_{enc}^{s}}{\partial \theta_{src,fe}} + \frac{\partial \mathcal{L}_{reg}^{s}}{\partial \theta_{reg}^{s}} \frac{\partial \theta_{reg}^{s}}{\partial \theta_{enc}^{s}} \frac{\partial \theta_{enc}^{s}}{\partial \theta_{src,fe}}) \Bigg\}, \\
\theta_{trg,fe}^{*} = \theta_{trg,fe} - \mu \Bigg\{ & \alpha \frac{(1-a)\partial \mathcal{L}_{trg}^{t,dom}}{\partial \theta_{dom}} \frac{\partial \theta_{dom}}{\partial \theta_{trg,fe}} + \\
& \beta(\frac{\partial \mathcal{L}_{enc}^{t}}{\partial \theta_{enc}^{t}} \frac{\partial \theta_{enc}^{t}}{\partial \theta_{trg,fe}} + \frac{\partial \mathcal{L}_{reg}^{t}}{\partial \theta_{reg}^{t}} \frac{\partial \theta_{reg}^{t}}{\partial \theta_{enc}^{t}} \frac{\partial \theta_{enc}^{t}}{\partial \theta_{trg,fe}}) \Bigg\} \\
\theta_{com,fe}^{*} = \theta_{com,fe} - \mu \Bigg\{ & \alpha \frac{\partial(-a\mathcal{L}_{com}^{s,dom} - (1-a)\mathcal{L}_{com}^{t,dom})}{\partial \theta_{dom}} \frac{\partial \theta_{dom}}{\partial \theta_{com,fe}} \\
& + \beta \left(\frac{\mathcal{L}_{enc}^{s}}{\partial \theta_{enc}^{s}} \frac{\partial \theta_{enc}^{s}}{\partial \theta_{com,fe}} + \frac{\mathcal{L}_{enc}^{t}}{\partial \theta_{enc}^{t}} \frac{\partial \theta_{enc}^{t}}{\partial \theta_{com,fe}} \right) \\
& + \frac{\partial \mathcal{L}_{reg}^{s}}{\partial \theta_{reg}^{s}} \frac{\partial \theta_{reg}^{s}}{\partial \theta_{enc}^{s}} \frac{\partial \theta_{enc}^{s}}{\partial \theta_{com,fe}} + \frac{\partial \mathcal{L}_{reg}^{t}}{\partial \theta_{reg}^{t}} \frac{\partial \theta_{reg}^{t}}{\partial \theta_{enc}^{t}} \frac{\partial \theta_{enc}^{t}}{\partial \theta_{com,fe}} \Bigg\},
\end{aligned} \tag{4}$$