$$\mathcal{L}_{\text{total}} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{DL}$$

$$\mathcal{L}(x; \mathbf{w}^T, \mathbf{b}_{\mathbf{w}}^S) = \alpha \mathcal{H}(y, p^T) + \beta \mathcal{H}(y, p^S) + \gamma \mathcal{H}(z^T, p^S)$$

$$J(\alpha, \boldsymbol{b}_{\mathrm{W}}) = \min_{\alpha, \mathbf{b}_{\mathrm{W}}} \| \mathrm{W} - \alpha \boldsymbol{b}_{\mathrm{W}} \|$$

$$J(\alpha, \boldsymbol{b}_{\mathrm{W}}) = \min_{\alpha, \mathbf{b}_{\mathbf{w}}} \| \mathbf{z} - Q_a(\alpha(\boldsymbol{a} \odot \mathbf{b}_{\mathbf{w}})) \|$$

$$\alpha^*, \mathbf{b}_{\mathrm{W}} = arg \min_{\alpha, \mathbf{b}_{\mathrm{W}}} J(\mathbf{b}_{\mathrm{W}}, \alpha)$$

$$\mathrm{ApproxSign}(x) = \begin{cases} -1, & x < -1 \\ 2x + x^2, & -1 \leq x < 0 \\ 2x - x^2, & 0 \leq x < 1 \\ 1, & \text{otherwise} \end{cases}$$

$$\frac{\partial \mathrm{ApproxSign}(x)}{\partial x} = \begin{cases} 2 + 2x, & -1 \leq x < 0 \\ 2 - 2x, & 0 \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$