$$egin{aligned} \mathcal{L}_{ ext{total}} &= \mathcal{L}_{CE} + \lambda \mathcal{L}_{DL} \ \mathcal{L}(x; \mathbf{w}^T, \mathbf{b}^S_\mathbf{w}) = lpha \mathcal{H}(x; \mathbf{w}^T, \mathbf{b}^S_\mathbf{w}) \end{aligned}$$

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

$$egin{aligned} J(lpha, oldsymbol{b}_{\mathrm{W}}) &= \min_{lpha, \mathbf{b}_{\mathrm{w}}} \|\mathbf{W} - lpha oldsymbol{b}_{\mathrm{W}}\| \ J(lpha, oldsymbol{b}_{\mathrm{w}}) &= \min_{lpha, \mathbf{b}_{\mathrm{w}}} \|\mathbf{z} - Q_a(lpha(oldsymbol{a} \odot \mathbf{b}_{\mathbf{w}}))\| \end{aligned}$$

$$lpha^*, \mathbf{b}_{\mathrm{W}} = \mathop{arg\,\mathrm{min}}_{lpha,\mathbf{b}_{\mathrm{W}}} J(\mathbf{b}_{\mathrm{W}},lpha)$$

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