

$$\mathcal{L}_{\text{tune}} = \mathbb{E}_{x^{\text{ref}}, t, \epsilon, x'} \left[ \left\| (\epsilon_{\theta}(x_t^{\text{ref}}, t, x', c^{\text{tex}}) - \epsilon) \right\|_2^2 \right]$$