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