

$$\mathcal{L}_{Cond} = \mathbb{E}_{Z_0, t, \epsilon} [\|\epsilon_\theta(\sqrt{\bar{\alpha}_t}Z_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t, c^{\text{tex}}, c^{\text{img}}) - \epsilon\|_2^2],$$