Adam Training Loss 10⁰ $L_{train}(t)$ 10^{-1} 10⁰ 10^{1} 10^{2}