Adding Additional Control to One-Step Diffusion with Joint Distribution Matching

Supplementary Material

A. Proof of Lemma 3.1

Since we assume the condition c is discrete, its entropy $\mathcal{H}(c)$ and conditional entropy $\mathcal{H}(c|\mathbf{x}_t)$ would be non-negative. Combine $\mathcal{H}(x_t, c) = \mathcal{H}(x_t) + \mathcal{H}(c|x_t)$, we have:

$$\begin{aligned} \mathcal{H}(x_t, c) &= -\mathbb{E}_{p_\theta(\mathbf{x}_t, c)} \log p_\theta(\mathbf{x}_t, c) \\ &\geq \mathcal{H}(x_t) = -\mathbb{E}_{p_\theta(\mathbf{x}_t)} \log p_\theta(\mathbf{x}_t). \end{aligned} \tag{13}$$

By substituting Eq. (13) into the integral joint KL divergence $\mathbb{E}_t \lambda_t \mathrm{KL}(p_\theta(\mathbf{x}_t, c)||p(\mathbf{x}_t, c))$, we have:

$$\begin{aligned} &\mathbb{E}_t \lambda_t \mathrm{KL}(p_\theta(\mathbf{x}_t, c)||p(\mathbf{x}_t, c)) \\ &= -\lambda_t \mathbb{E}_{p_\theta(\mathbf{x}_t|c)p(c),t} \log p(c|\mathbf{x}_t) p_\phi(\mathbf{x}_t) + \lambda_t \mathbb{E}_{p_\theta(\mathbf{x}_t, c)} \log p_\theta(\mathbf{x}_t, c) \\ &\leq -\lambda_t \mathbb{E}_{p_\theta(\mathbf{x}_t|c)p(c),t} \log p(c|\mathbf{x}_t) p_\phi(\mathbf{x}_t) + \lambda_t \mathbb{E}_{p_\theta(\mathbf{x}_t)} \log p_\theta(\mathbf{x}_t) \\ &= \lambda_t \mathbb{E}_{p_\theta(\mathbf{x}_t|c)p(c),t} [-\log p(c|\mathbf{x}_t) p_\phi(\mathbf{x}_t) + \log p_\theta(\mathbf{x}_t)] \end{aligned} \tag{14}$$

This completes the proof.

B. Details of conditions

- Canny: a canny edge detector [3] is employed to generate canny edges;
- Hed: a holistically-nested edge detection model is utilized for the purpose;
- Depthmap: we employ the Midas [24] for depth estimation;
- Super-resolution: we use the nearest kernel to downscale the images by a factor of 8 as the condition.