Siddharth Mishra-Sharma (MIT/IAIFI) | IAIFI Summer School





The denoising objectives

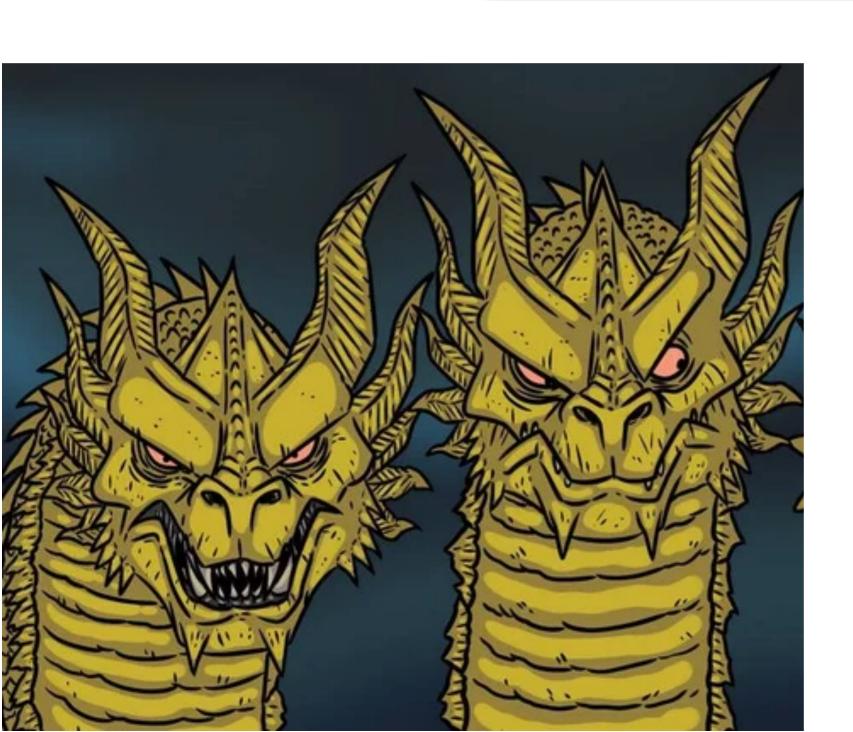
 $\frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1} \left(1 - \alpha_t\right)^2}{\left(1 - \bar{\alpha}_t\right)^2} \left[\left\| \hat{x}_{\theta} \left(z_t, t\right) - x \right\|^2 \right]$



x-prediction; MLE

ϵ -prediction; MLE

$$\frac{1}{2\sigma_q^2(t)} \frac{\left(1-\alpha_t\right)^2}{\left(1-\bar{\alpha}_t\right)\alpha_t} \left[\left\| \epsilon - \hat{\epsilon}_\theta\left(x_t, t\right) \right\|^2 \right]$$





 ϵ -prediction; "simple"

$$\| \epsilon - \hat{\epsilon}_{\theta} (x_t, t) \|^2$$

Typical objective for training image diffusion models: SOTA on many tasks!

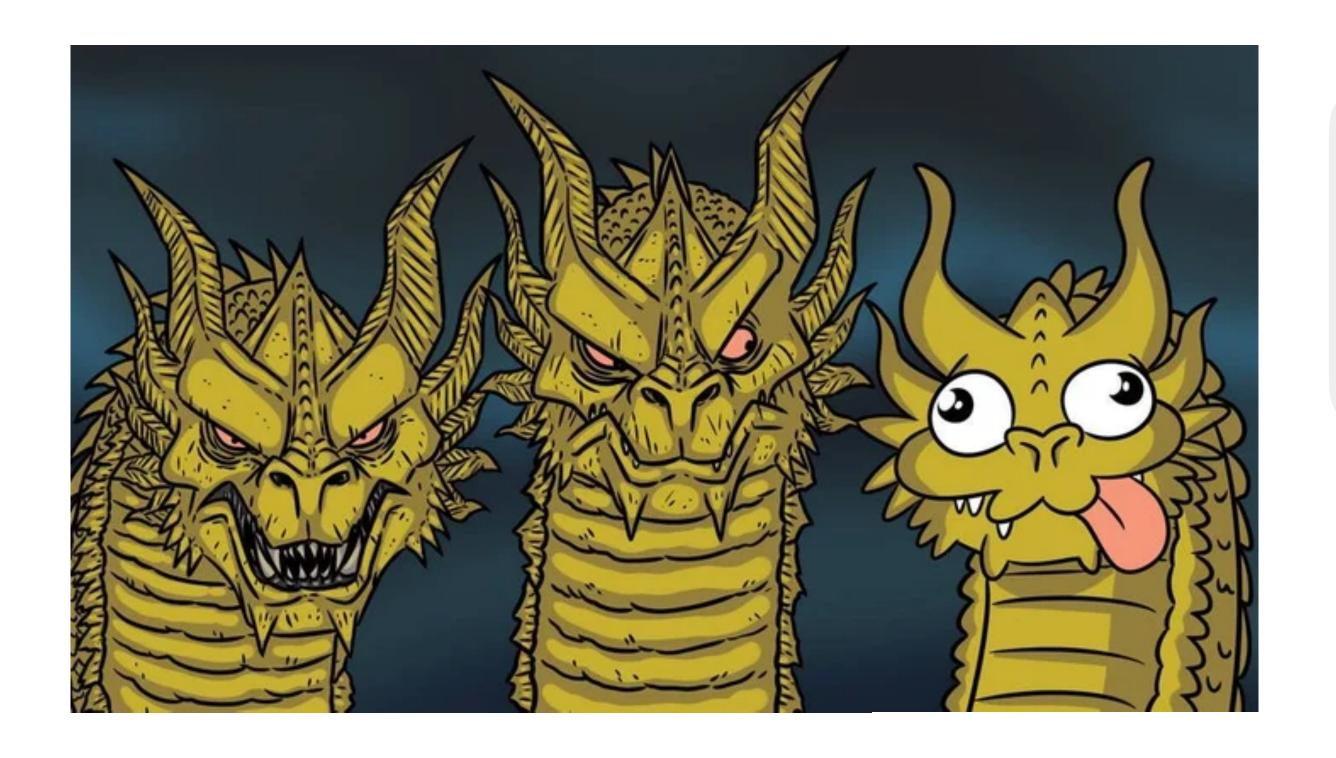
The denoising objectives

x-prediction; MLE

$$\frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1} \left(1 - \alpha_t\right)^2}{\left(1 - \bar{\alpha}_t\right)^2} \left[\left\| \hat{x}_{\theta} \left(z_t, t\right) - x \right\|^2 \right]$$

 ϵ -prediction; MLE

$$\frac{1}{2\sigma_q^2(t)} \frac{\left(1 - \alpha_t\right)^2}{\left(1 - \bar{\alpha}_t\right)\alpha_t} \left[\left\| \epsilon - \hat{\epsilon}_\theta\left(x_t, t\right) \right\|^2 \right]$$



 ϵ -prediction; "simple"

$$\| \epsilon - \hat{\epsilon}_{\theta} (x_t, t) \|^2$$

Typical objective for training image diffusion models: SOTA on many tasks!

Simple objectives as a weighted sum of ELBOs

Kingma et al (2023) showed that common objectives can be written as a weighted sum (across different noise levels) of ELBOs

$$L_{w}(x) = \left\langle w(t) \cdot w_{\text{ML}}(t) \mid \epsilon - \hat{\epsilon}_{\theta} \left(x_{t}, t \right) \mid \right|^{2} \right\rangle$$

Additional weighting $(w_{\text{ML}}^{-1} \text{ for } \epsilon\text{-prediction})$

Weighting for ELBO/
ML objective