

Siddhant Mishra-Sharma (MIT/AI FI) Summer School

162



38

The denotating objective

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



x-prediction; MLE

ϵ -prediction; MLE

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





ϵ -prediction; "simple"

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

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

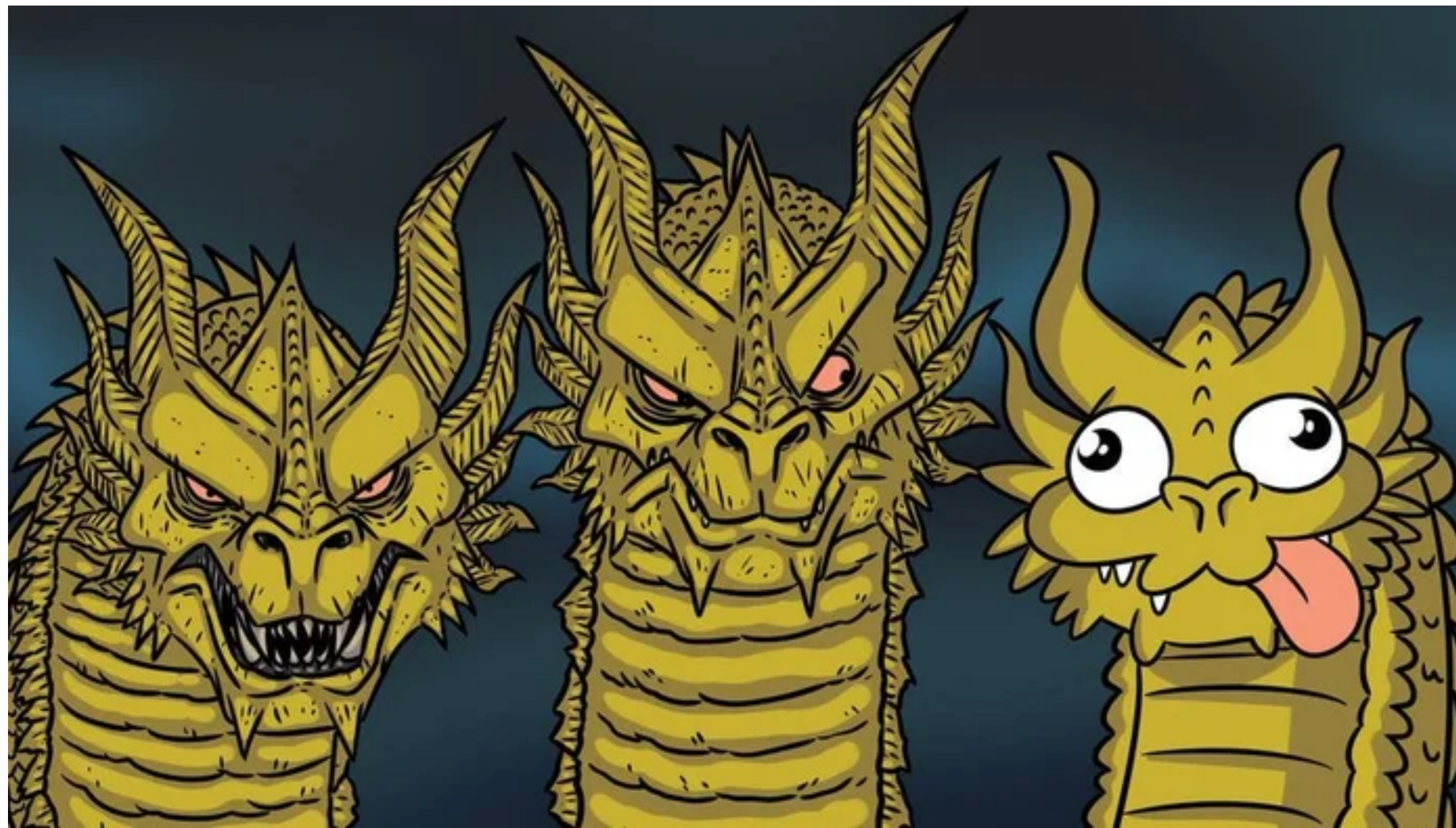
The denoising objectives

x -prediction; MLE

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

ϵ -prediction; MLE

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



ϵ -prediction; “simple”

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

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

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

Additional weighting
(w_{ML}^{-1} for ϵ -prediction)

Weighting for ELBO/
ML objective