TRAINING

Goal: minimize the $\boxed{\text{KL divergence}}$ between the true reverse distribution $q(x_{t-1}|x_t, x_0)$ and the model's approximation $p_\theta(x_{t-1}|x_t)$

↳ Reduces to MSE

$$(\varepsilon - \varepsilon_\theta(x_t, t))^2$$

True Noise added during forward process

Model's prediction of the noise

x_t is the noisy image

ask Prof. to explain.

How did we get to MSE? (Struggling w/ ELBO)

- Goal of generative models is to maximize the likelihood of training data.

$$\max_\theta \log p_\theta(x_0)$$, where x_0 is a real image and p_θ is the distribution over all possible images.

- Directly calculating the likelihood is difficult since the model doesn't generate x_0 directly but does so via a chain of latent (hidden) variables.

$$x_T \to x_{T-1} \to \dots \to x_1 \to x_0$$

and so the probability of x_0 is

Marginal Likelihood

$$p_\theta(x_0) = \int p_\theta(x_{0:T})\, dx_{1:T}$$

continuous sum/integral

ignores the sequence as long as final output is the desired

* Imagine a large, high-dim data set. $x_0, \dots x_{0n} \in \mathbb{R}^d$

* Imagine a very large $T = t$.