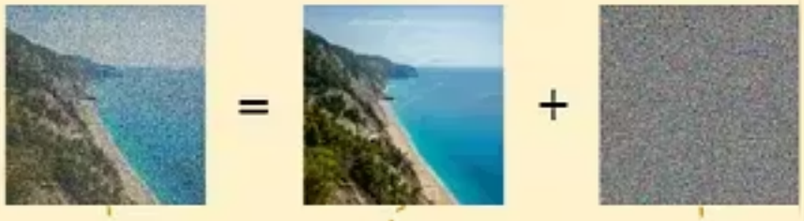
For each training step:

1. Randomly select a time step & encode it

$t = 14$ —encode→ Time step embedding: 0.12 | 0.31 | 0.34 | ⋯ | 0.02

2. Add noise to image



$$x_t = \sqrt{\bar{a}_t}\, x_0 + \sqrt{1 - \bar{a}_t}\, \varepsilon$$

Adjust the amount of noise according to the time step *t*

$$\varepsilon \sim \mathcal{N}(0, 1)$$

$$\alpha_t = 1 - \beta_t$$

$$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$$

3. Train the UNet