VAE: VARIATIONAL AUTOENCODER

Presented by Jennifer Zhang

Mentor: Yuhan Qian

AGENDA

1 What is VAE?

2 How to train VAE model?

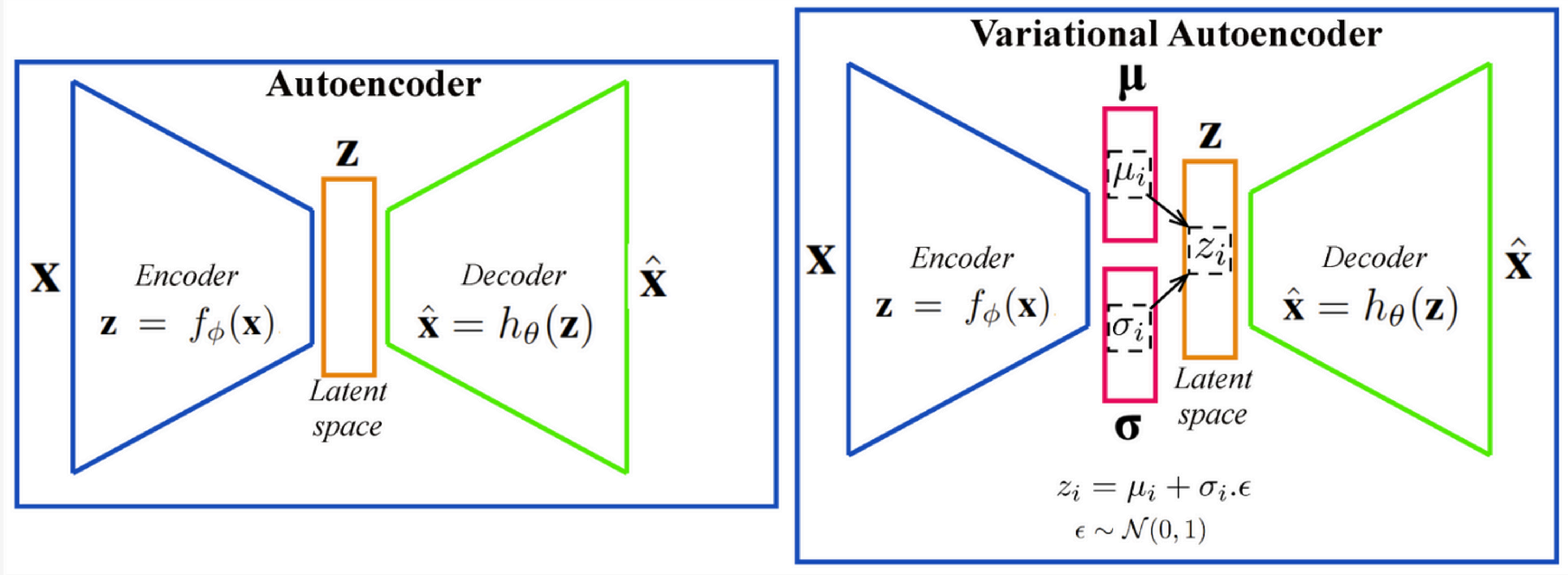
3 Conditional VAE

4 Conclusion

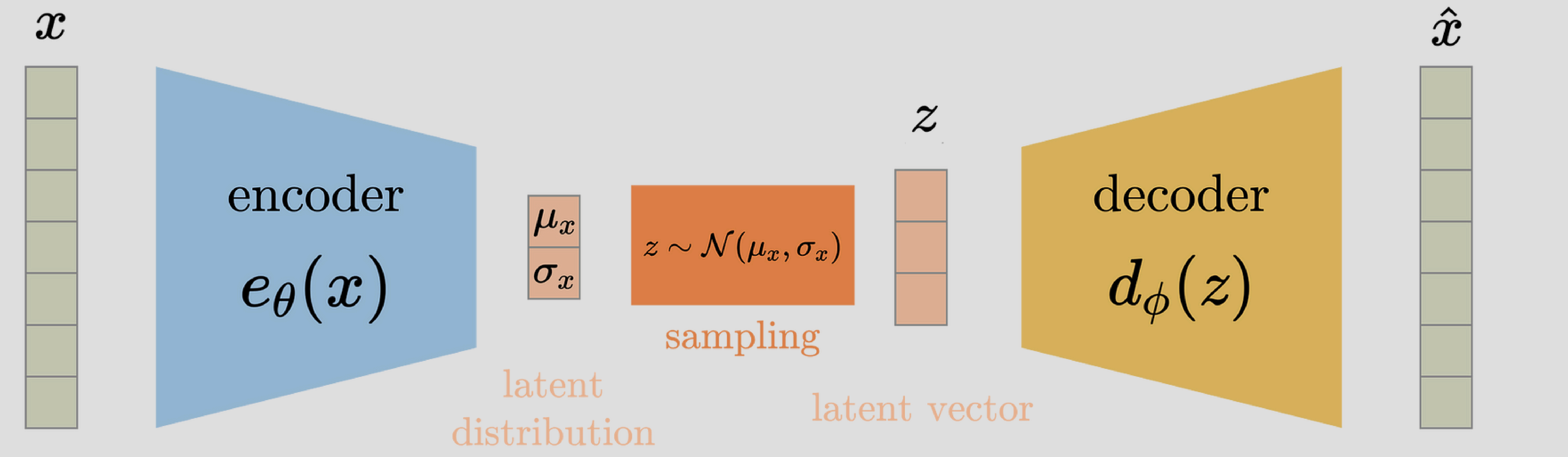
WHY NOT JUST AUTOENCODERS?

- A common Autoencoder outputs a point without probability explanation
- Cannot sample from the latent space to generate new samples.
- VAE learns the distribution (through μ and σ²) and can "generate data".

Figure 1. Schematic architecture of a standard deep autoencoder and a variational deep autoencoder. Both architectures have two parts: an encoder and a decoder.



VAE ARCHITECTURE: ENCODER → LATENT → DECODER



$$\text{reconstruction loss } = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$similarity\ loss = KL\ Divergence = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$loss = reconstruction\ loss + similarity\ loss$$

TRAINING OBJECTIVE = RECONSTRUCTION + REGULARIZATION

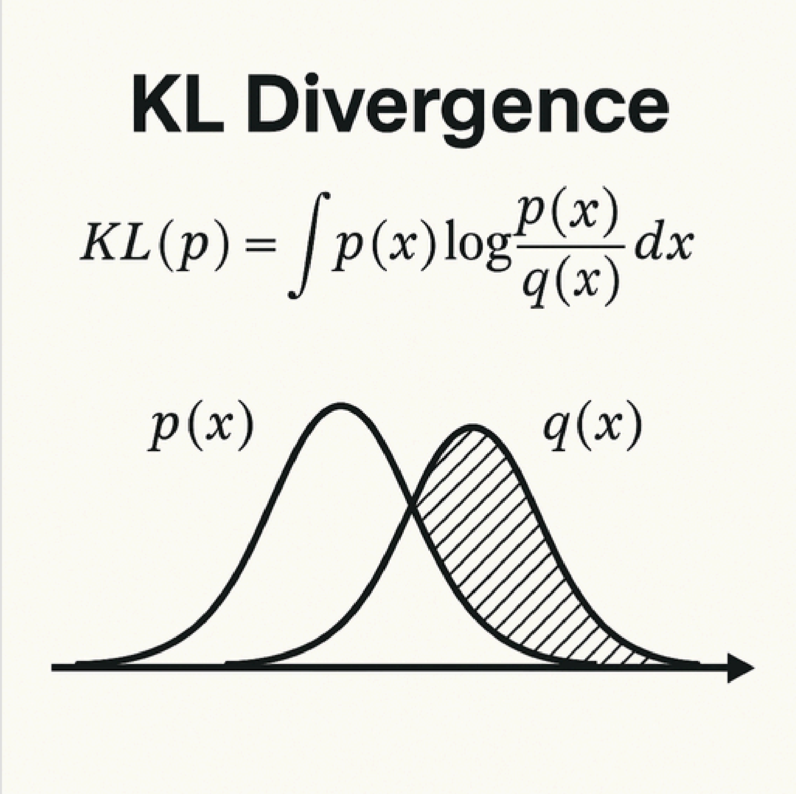
$$\text{reconstruction loss} = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

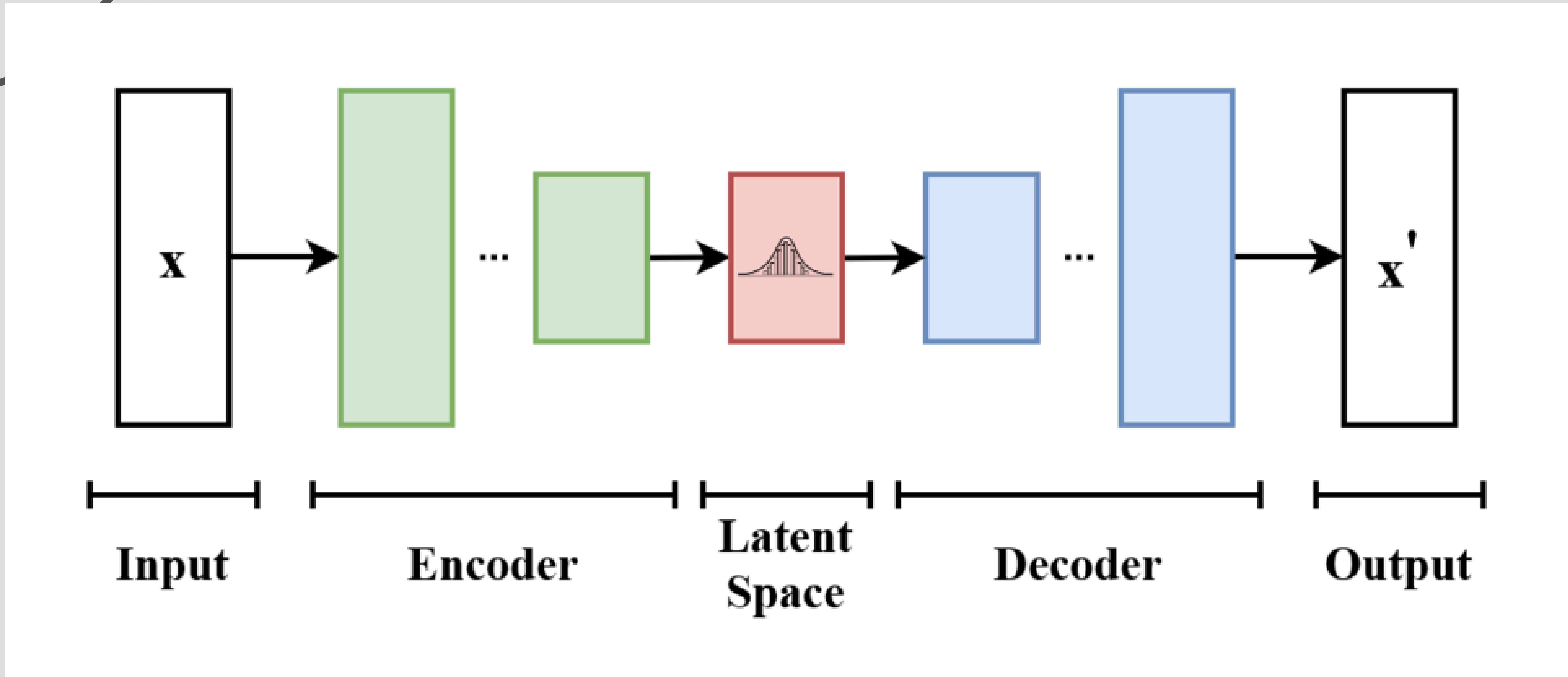
$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$similarity\ loss = KL\ Divergence = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

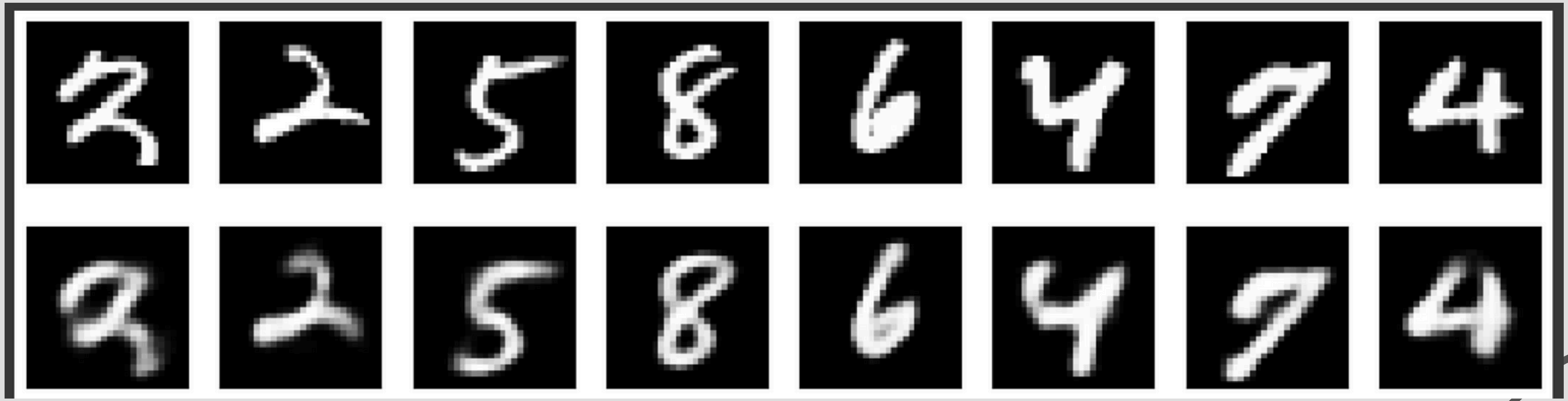
$$loss = reconstruction\ loss + similarity\ loss$$

- Reconstruction: Keep the image restored clearly.
- KL Divergence: Make the distribution of latent variable (z) close to the standard normal

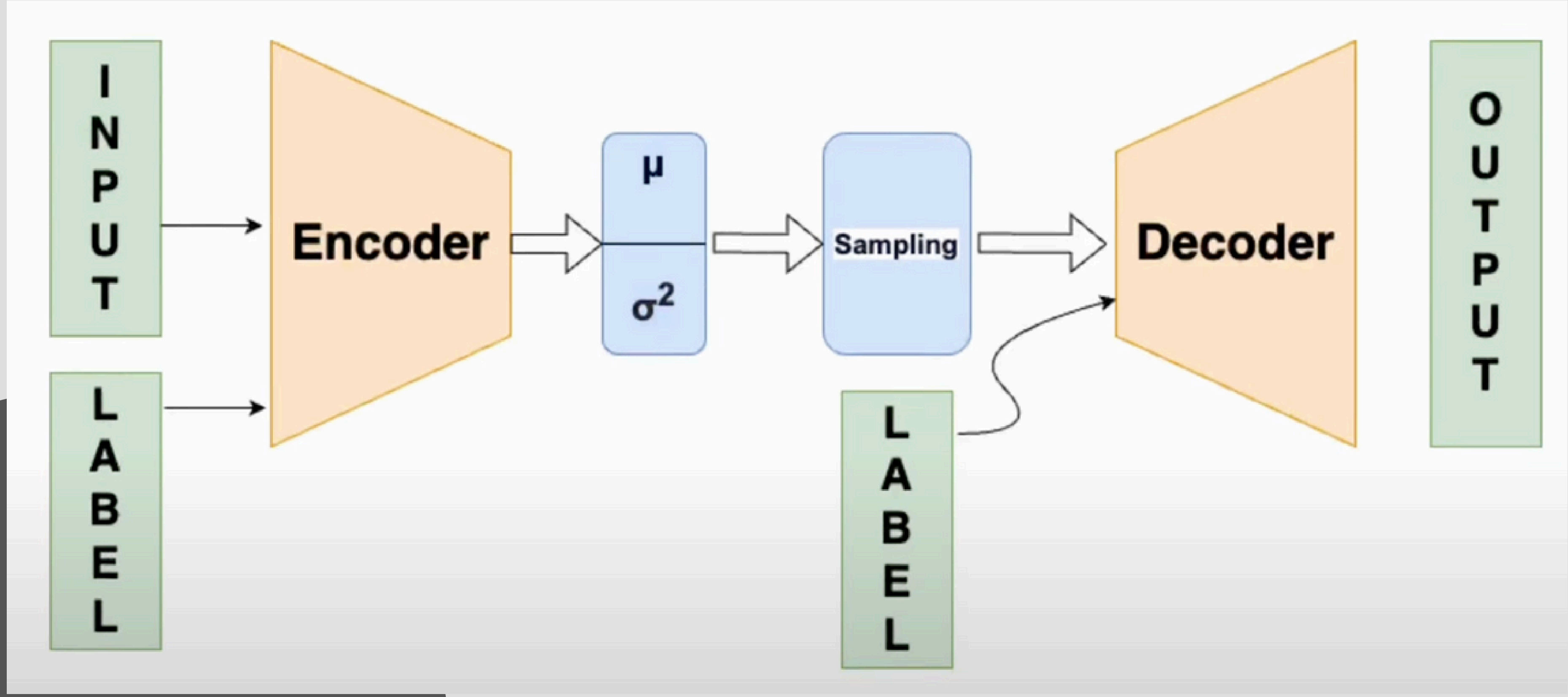




num of my epoches = 10



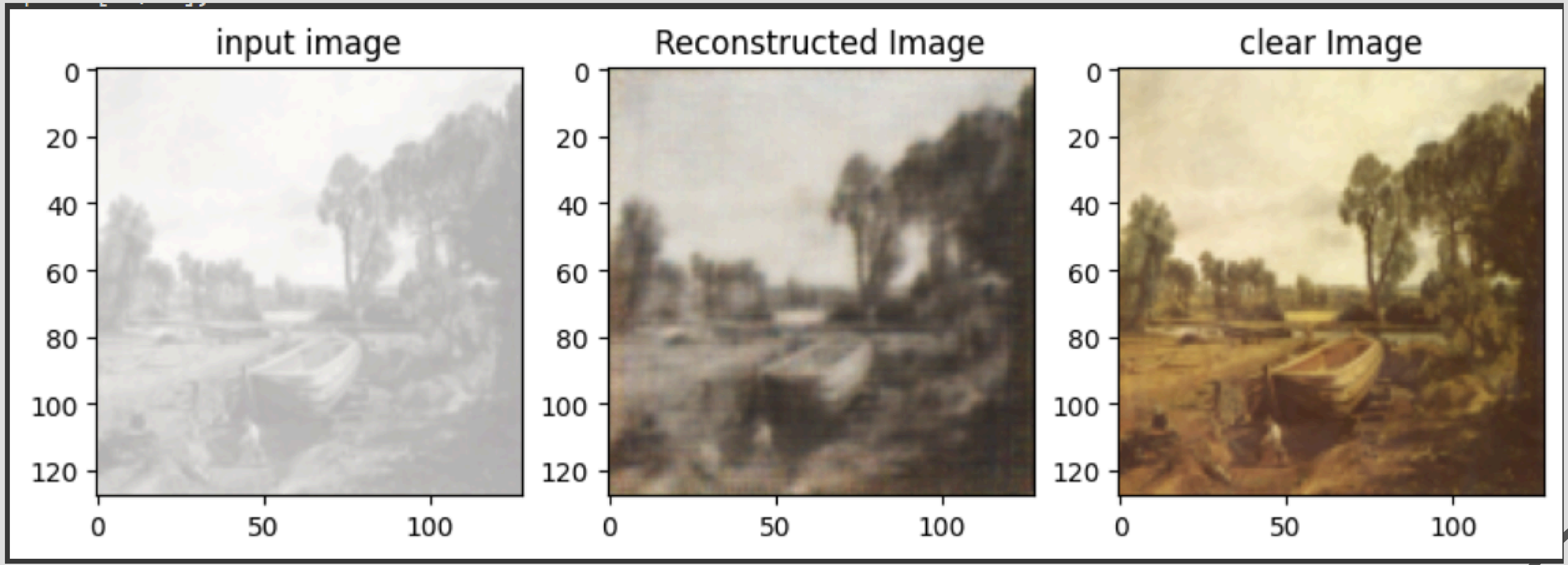
CONDITIONAL VAE



- **Why Conditional VAE?**
- Vanilla VAE has no control over generation (e.g. what kind of numbers are generated)
- CVAE achieves goal-directed generation with condition variables such as labels

WHAT WE LEARNED

- VAEs can compress & generate
- CVAE achieves controllable generation
- Extensible to: image colorization （Conditional VAE, diversity of colors by conditioning on the grayscale image, generating more realistic and varied results.）



CONCLUSION

- VAE extends the traditional autoencoder by adding probabilistic modeling.
- It can generate new data by sampling from a learned latent distribution.
- CVAE improves on this by adding control through condition variables.
- And both models are widely used in today's AI systems.
- In all, these models are very flexible and creative tools in generative AI.

THANK YOU