

# HUST

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



**ĐẠI HỌC**  
**BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

# Autoencoder

ONE LOVE. ONE FUTURE.

# Nội dung chính

- Autoencoder
  - Undercomplete Autoencoder
  - Sparse Autoencoder
  - Denoising Autoencoder
- Variational Autoencoder
  - Evidence Lower Bound
  - Reparameterization Trick
  - Posterior Collapse
- Ứng dụng trong bài toán thực tế

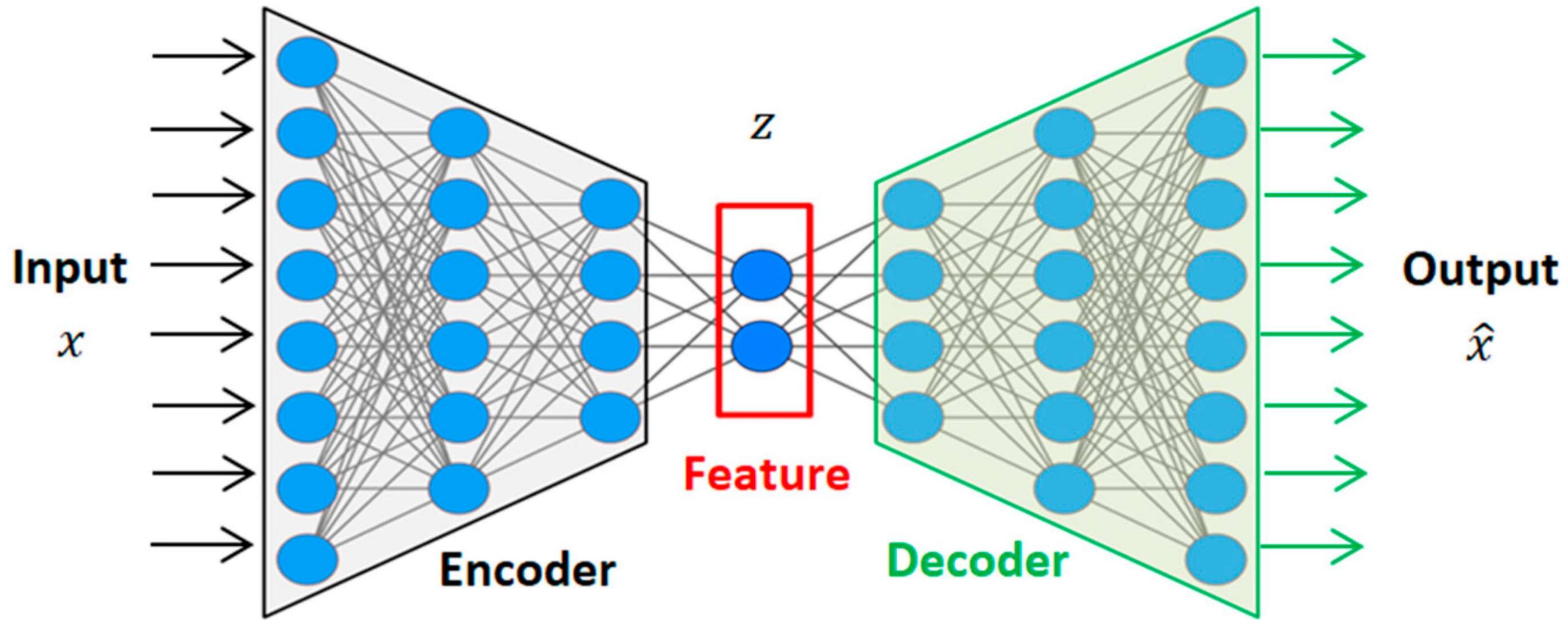


# Autoencoder

- Autoencoder là một mạng neural được huấn luyện để sao chép chính đầu vào của nó sang đầu ra. Mạng gồm hai phần:
  - Encoder: Học hàm mã hóa  $z = f(x)$
  - Decoder: Học hàm giải mã  $x' = g(z)$
- Một autoencoder nếu chỉ đơn giản học cách xây dựng  $g(f(x)) = x$  thì không gì hữu ích cả. Trên thực tế autoencoder được thiết kế sao cho không thể học cách sao chép hoàn hảo với các ràng buộc khác nhau buộc mạng phải tìm hiểu những cách thức mới để tái hiện lại dữ liệu



# Undercomplete Autoencoder



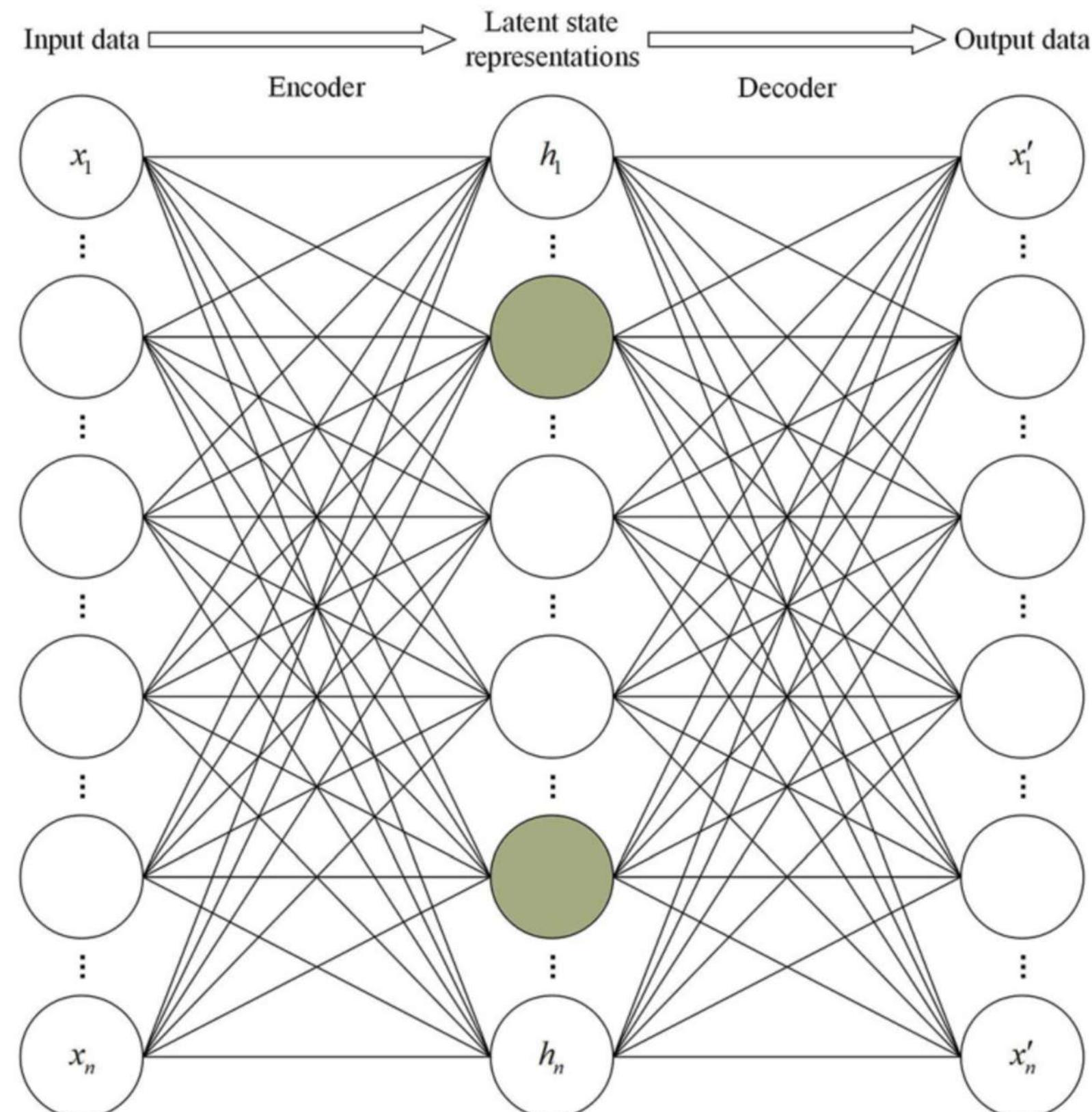
# Regularized Autoencoder

- Trên lý thuyết, nếu encoder và decoder có quá nhiều khả năng biểu diễn, autoencoder có thể học cách thực hiện nhiệm vụ sao chép mà không trích xuất được thông tin hữu ích nào về phân phối của dữ liệu
- Thay vì cố gắng giới hạn khả năng của encoder và decoder bằng cách tạo mạng nhỏ và nông, regularized autoencoder bổ sung thêm ràng buộc vào hàm mất mát, buộc mô hình phải có cách biểu diễn khác ngoài khả năng sao chép đầu vào sang đầu ra



# Sparse Autoencoder

- Sparse Autoencoder đảm bảo mô hình có latent vector nhiều chiều hơn, học được nhiều đặc trưng hơn nhưng vẫn không dính phải identity mapping



# Sparse Autoencoder

- Hàm mất mát của Sparse Autoencoder

$$L = L(x, g(f(x))) + \alpha \Omega(z)$$

- Trong đó hàm Sparsity Penalty:
  - Kullback-Leibler Divergence:

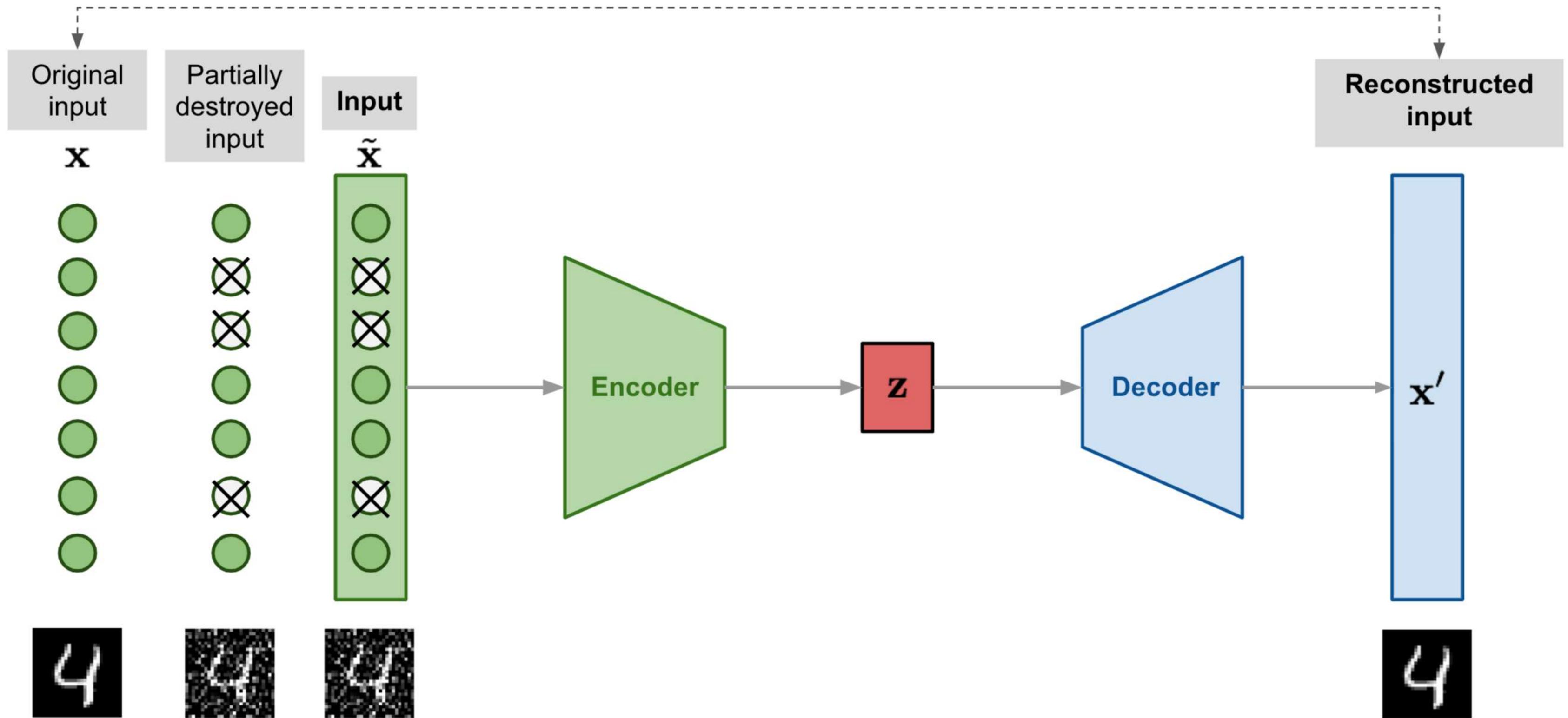
$$\Omega(z) = \sum_{j=1}^k KL(p || p_j) = \sum_{j=1}^k \left( p \log \frac{p}{p_j} + (1-p) \log \frac{1-p}{1-p_j} \right)$$

- L1 Regularization:

$$\Omega(z) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k |z_j^{(i)}|$$



# Denoising Autoencoder

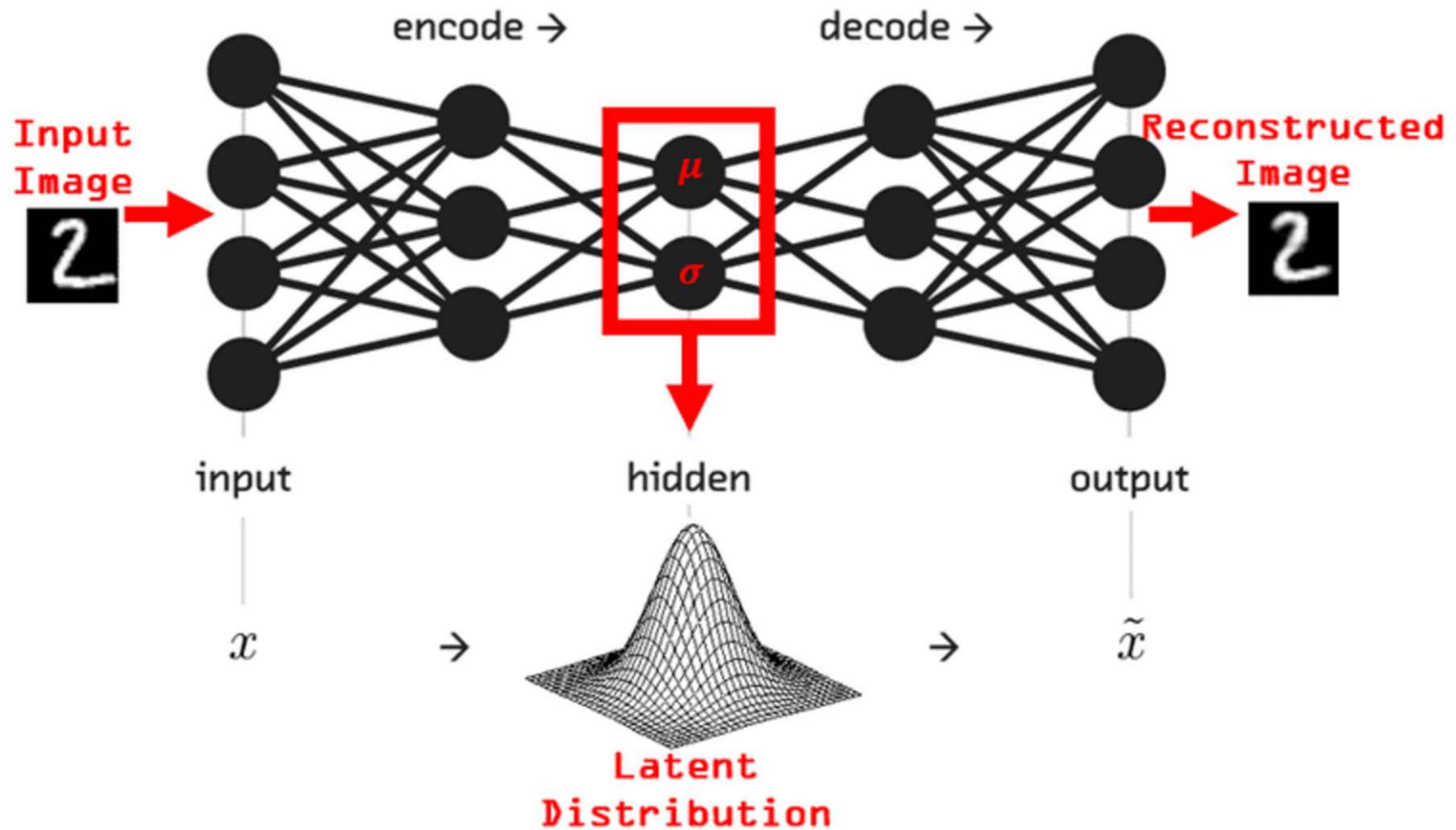


- **Ưu điểm của Autoencoder**
  - Giảm chiều phi tuyến, cho kết quả tốt hơn PCA trong đa số các trường hợp
  - Khả năng phát hiện bất thường tốt do vốn được huấn luyện để tái tạo dữ liệu bình thường
  - Giảm nhiễu dữ liệu
  - Học đặc trưng chuyên biệt cho tác vụ

- Autoencoder có một nhược điểm là do được huấn luyện với mục tiêu duy nhất là tái tạo tốt nhất có thể, mã hóa mỗi dữ liệu thành một latent vector cố định nên latent space thu được vô cùng bất thường:
  - Mất tính liên tục: Các vector gần nhau trong latent space có thể cung cấp dữ liệu được giải mã rất khác nhau
  - Mất tính đầy đủ: Một số điểm của không gian tiềm ẩn cung cấp nội dung vô nghĩa sau khi được giải mã

# Variational Autoencoder

- Variational Autoencoder mã hóa thành một latent distribution bằng cách học trung bình và độ lệch chuẩn
- VAE vừa tái tạo được dữ liệu gốc, vừa tạo ra được dữ liệu mới



# Evidence Lower Bound

- Ngoài Reconstruction Loss, VAE còn sử dụng thêm một hàm gọi là Regularization Loss biểu thị dưới dạng Kulback-Leibler Divergence

$$\begin{aligned} L_{KL} &= KL(N(\mu_x, \sigma_x) || N(0,1)) \\ &= \frac{1}{2} \sum_{i=1}^k (\mu_i(x)^2 + \sigma_i(x)^2 - \log \sigma_i(x)^2 - 1) \end{aligned}$$

- Hàm mất mát của VAE

$$L_{VAE} = L_{RL} + L_{KL}$$



# Reparameterization Trick

- Thay vì trực tiếp lấy ngẫu nhiên từ phân phối xác suất, người ta sử dụng 1 mẹo nhỏ gọi là Reparameterization Trick, vector đầu vào của decoder sẽ được tính bằng công thức

$$z = \mu_x + \varepsilon\sigma_x, \varepsilon \sim N(0, 1)$$



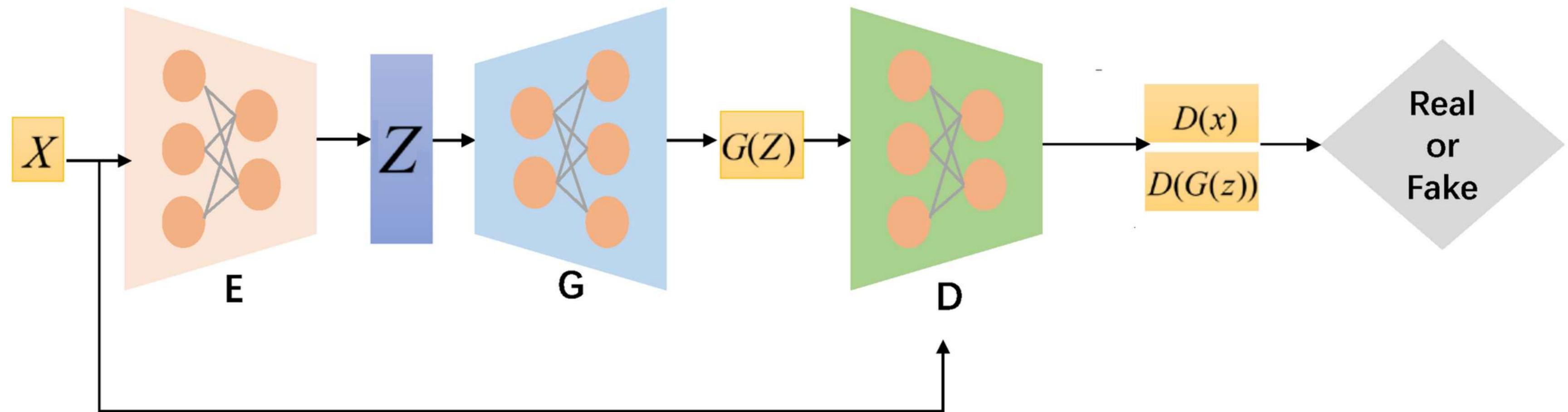
# Posterior Collapse

- Trong tác vụ sinh dữ liệu mới, VAE cũng có thể gặp hiện tượng chỉ sinh ra các mẫu giống nhau hoặc quá đơn giản giống như Mode Collapse ở GAN, hiện tượng này gọi là Posterior Collapse
- Nguyên nhân chủ yếu khi KL Divergence quá mạnh, ép toàn bộ đầu ra về đúng với phân phối chuẩn dẫn đến latent vector không còn thông tin đúng về đầu vào
- Cách khắc phục chính: khi huấn luyện bổ sung trọng số cho KL Divergence, đảm bảo cân bằng với Reconstruction Loss, hoặc ban đầu để bé và tăng dần ở giai đoạn sau



- VAE thường xuyên được so sánh với GAN vì cùng là một kiến trúc mô hình được dùng để sinh dữ liệu giống với dữ liệu huấn luyện, đặc biệt là ảnh
- Khi sinh ảnh, cả hai phương pháp đều có ưu và nhược điểm:
  - GAN tạo ra ảnh rõ nét hơn, nhưng do bản chất đối kháng giữa generator và discriminator, quá trình huấn luyện khó ổn định
  - VAE dễ huấn luyện hơn, nhưng do cách tạo ảnh từ các tính năng trung bình của dữ liệu huấn luyện, ảnh sinh ra thường mờ hơn

# VAE vs GAN



# Ứng dụng trong bài toán thực tế

- Bộ dữ liệu: FFHQ (Flickr-Faces-HQ Dataset)
- Bao gồm 52000 ảnh mặt người đa dạng giới tính, tuổi tác, màu da,... thu thập trên nền tảng Flickr - cộng đồng các nhiếp ảnh gia



# Ứng dụng trong bài toán thực tế

- Với tác vụ tái tạo ảnh, các chỉ số để đánh giá AE, Sparse AE, DAE và VAE là:

- Reconstruction Error (RE) → Càng thấp càng tốt:

$$MSE = \frac{1}{H} \frac{1}{W} \sum_{i=1}^H \sum_{j=1}^W (x_{ij} - x'_{ij})^2$$

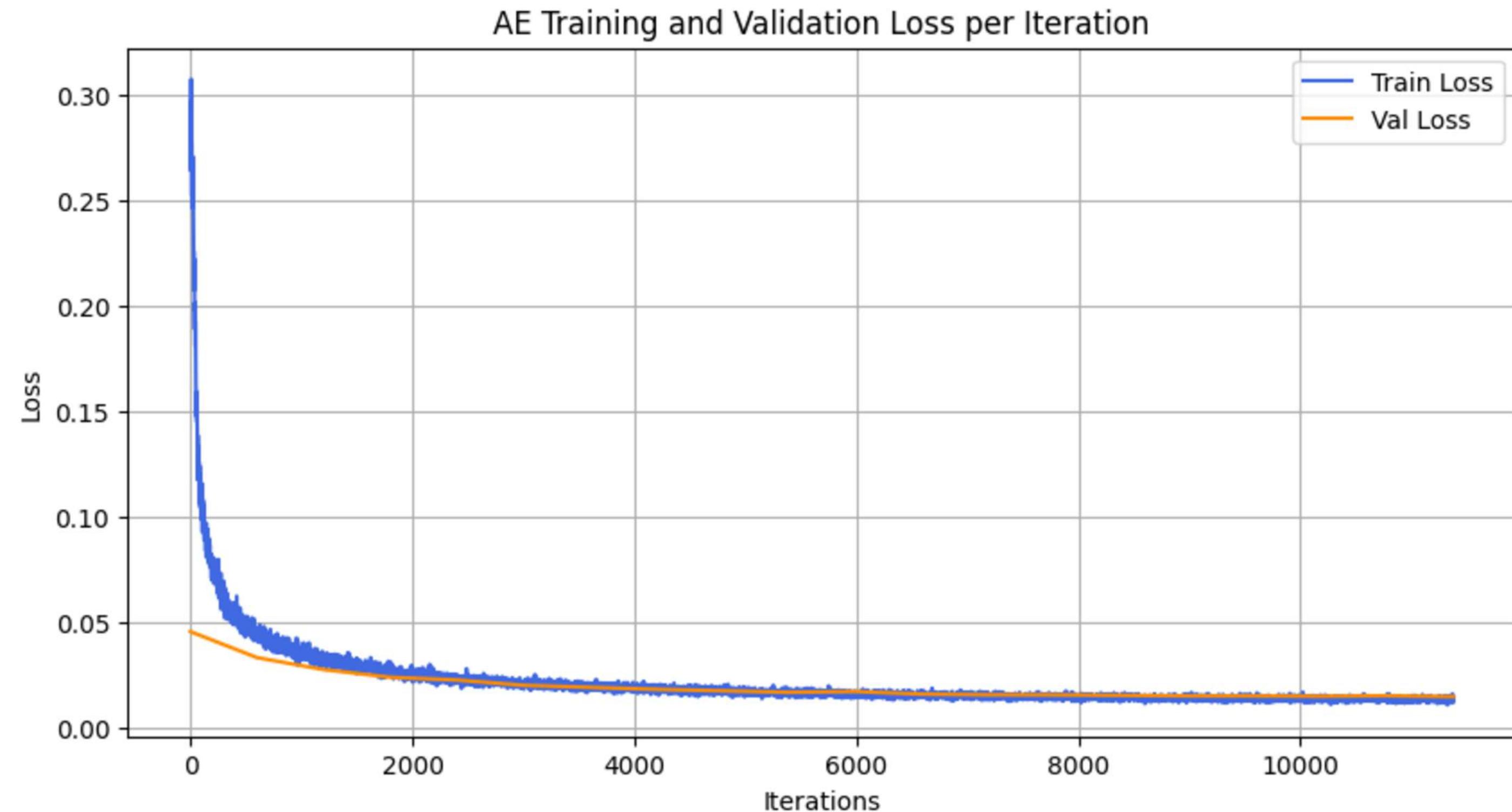
- Peak Signal-to-Noise Ratio (PSNR) → Càng cao càng tốt:

$$PSNR = 20 \log_{10} \left( \frac{MAX}{\sqrt{MSE}} \right)$$

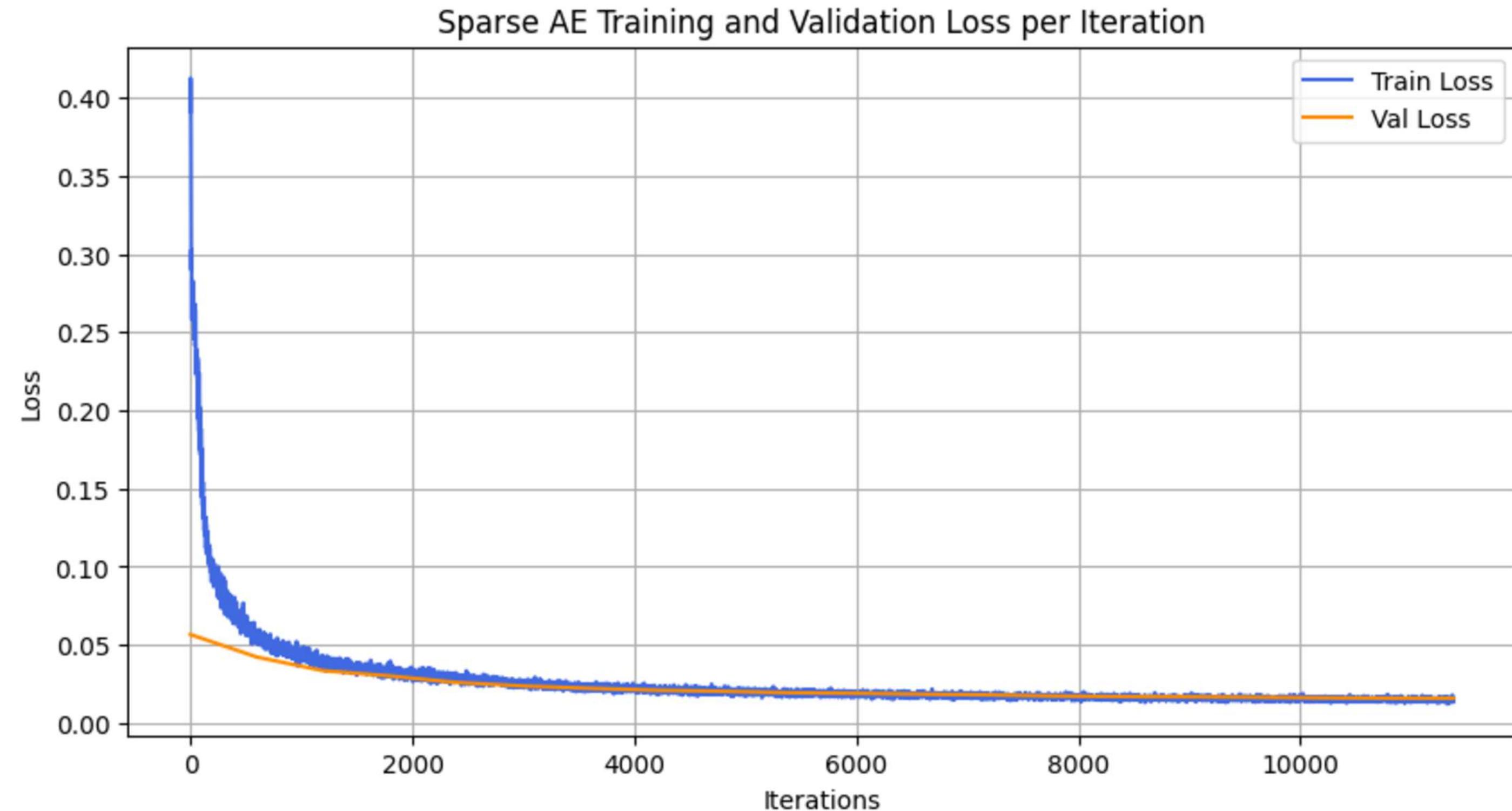
- Structural Similarity Index (SSIM) → Càng gần 1 càng tốt:

$$SSIM(x, y) = \frac{2\mu_x\mu_y + \varepsilon_1}{\mu_x^2 + \mu_y^2 + \varepsilon_1} \frac{2\sigma_x\sigma_y + \varepsilon_2}{\sigma_x^2 + \sigma_y^2 + \varepsilon_2} \frac{\sigma_{xy} + \varepsilon_3}{\sigma_x + \sigma_y + \varepsilon_3}$$

# Ứng dụng trong bài toán thực tế

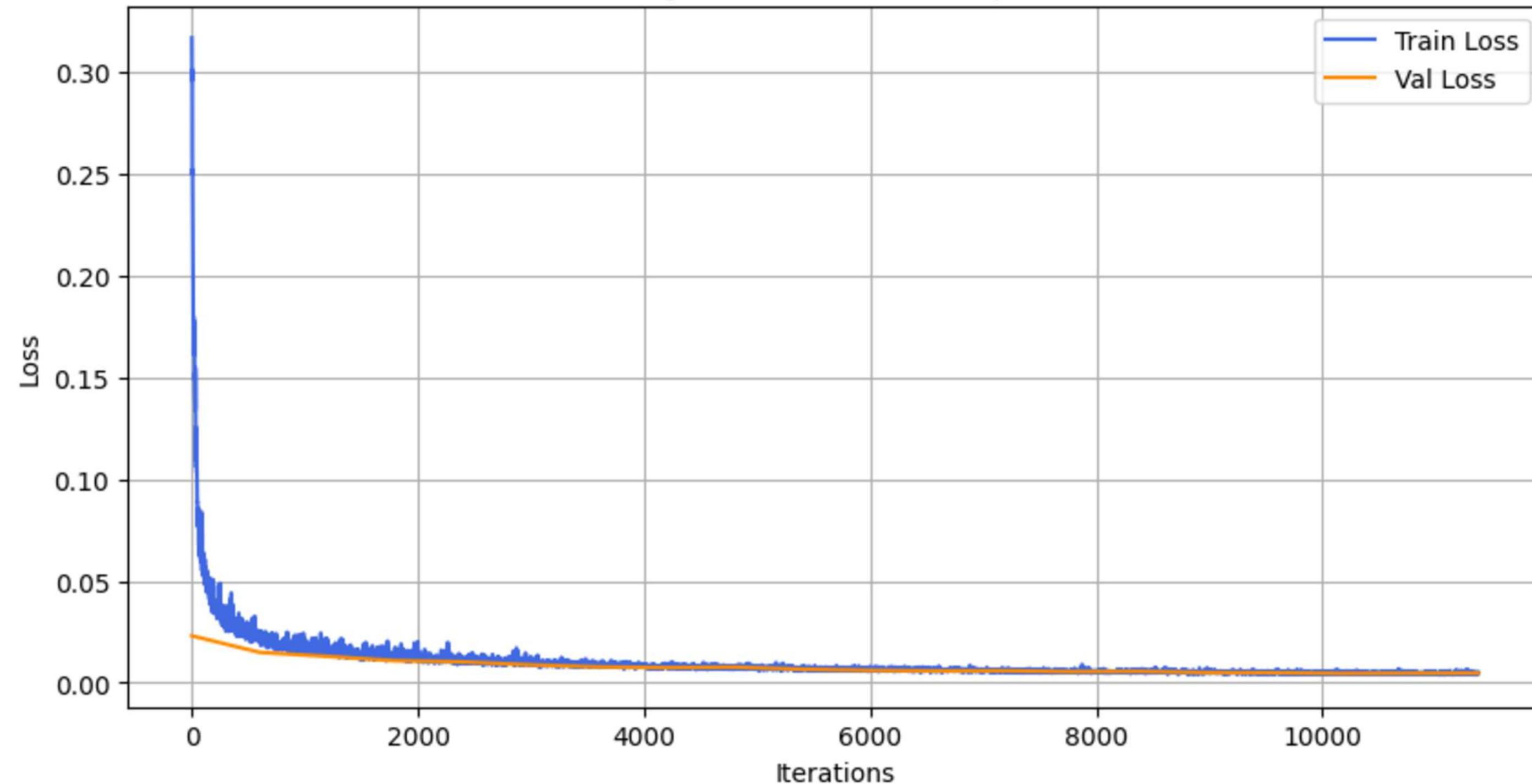


# Ứng dụng trong bài toán thực tế

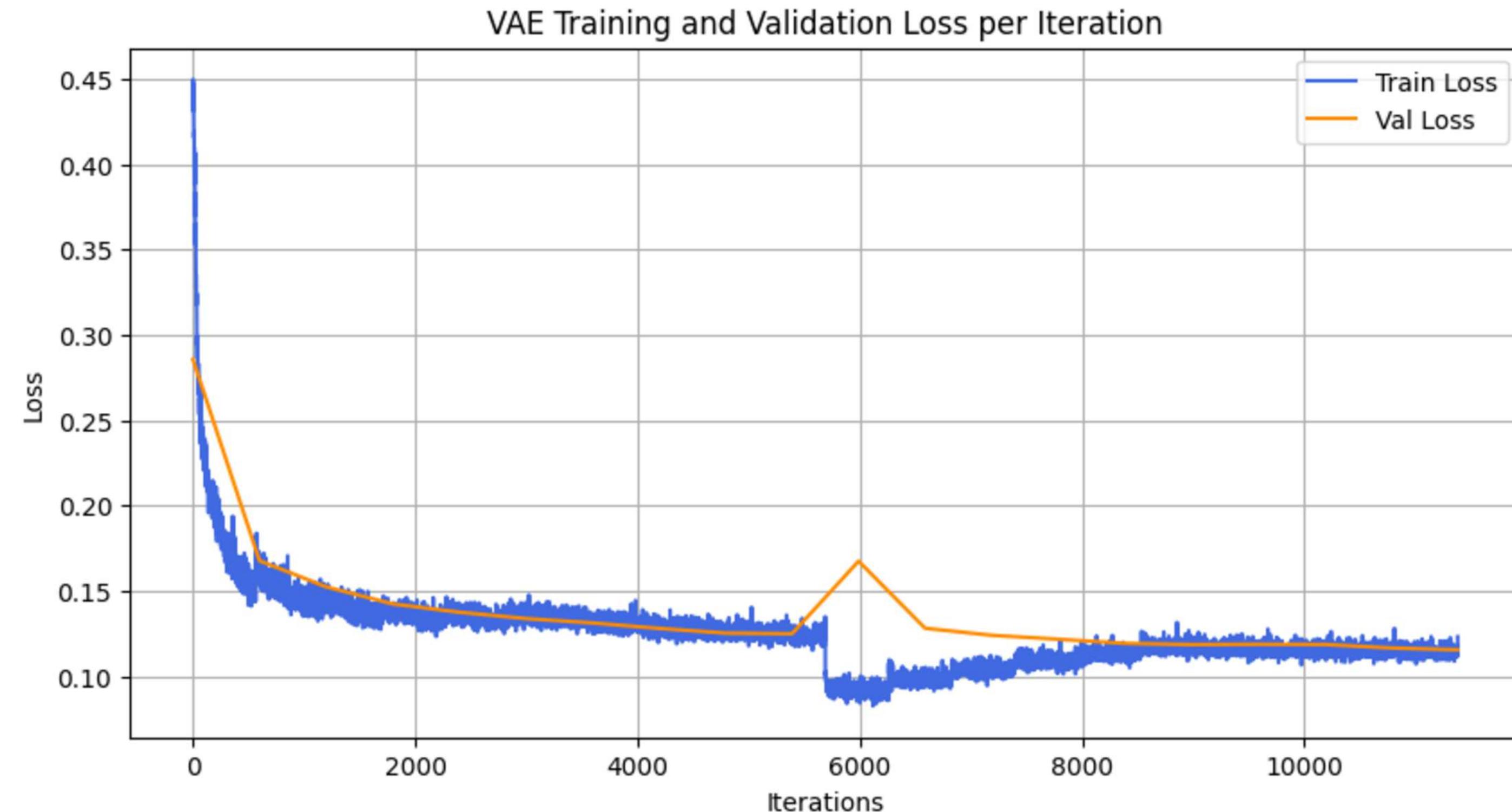


# Ứng dụng trong bài toán thực tế

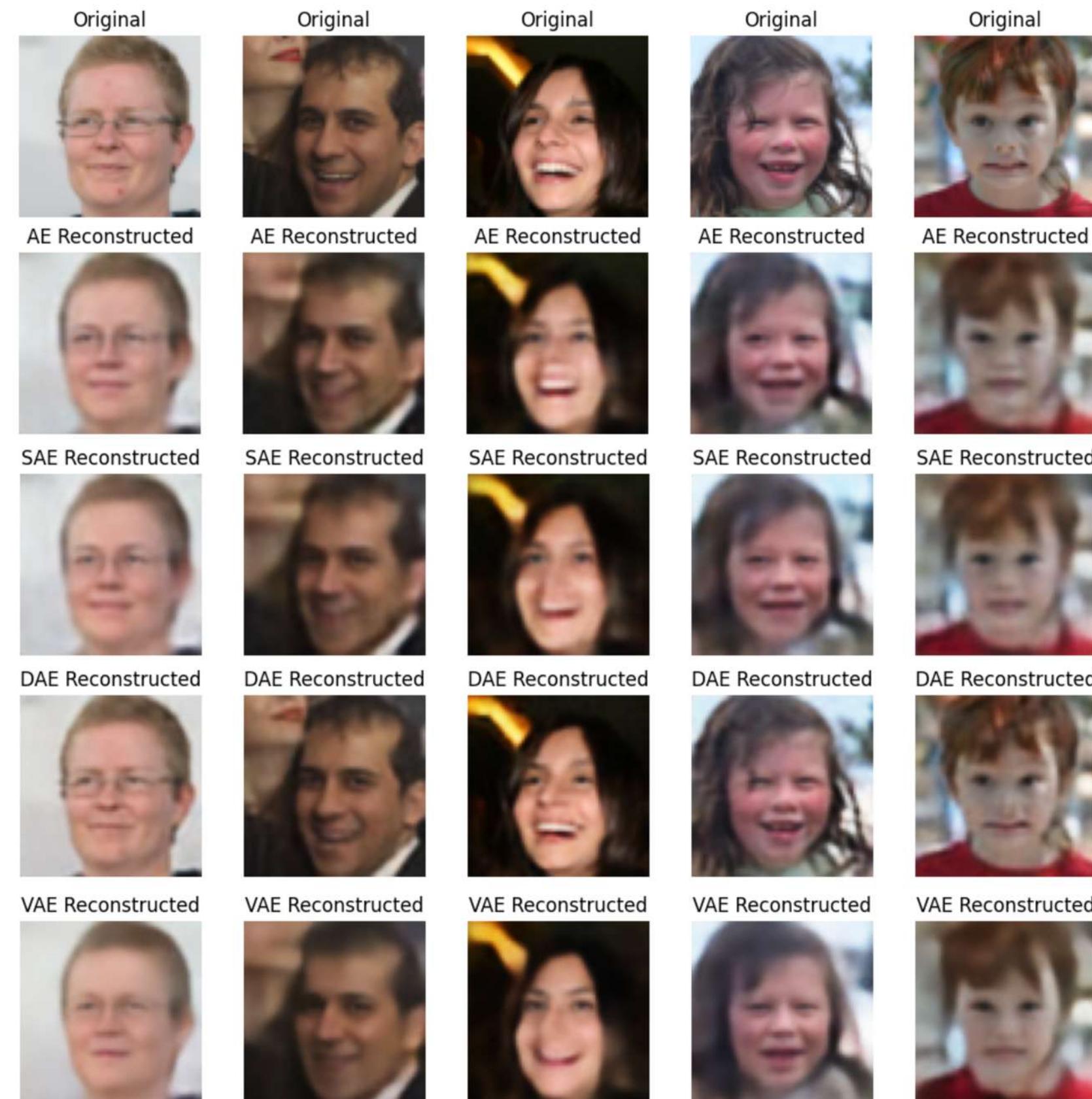
DAE Training and Validation Loss per Iteration



# Ứng dụng trong bài toán thực tế



# Ứng dụng trong bài toán thực tế



# Ứng dụng trong bài toán thực tế

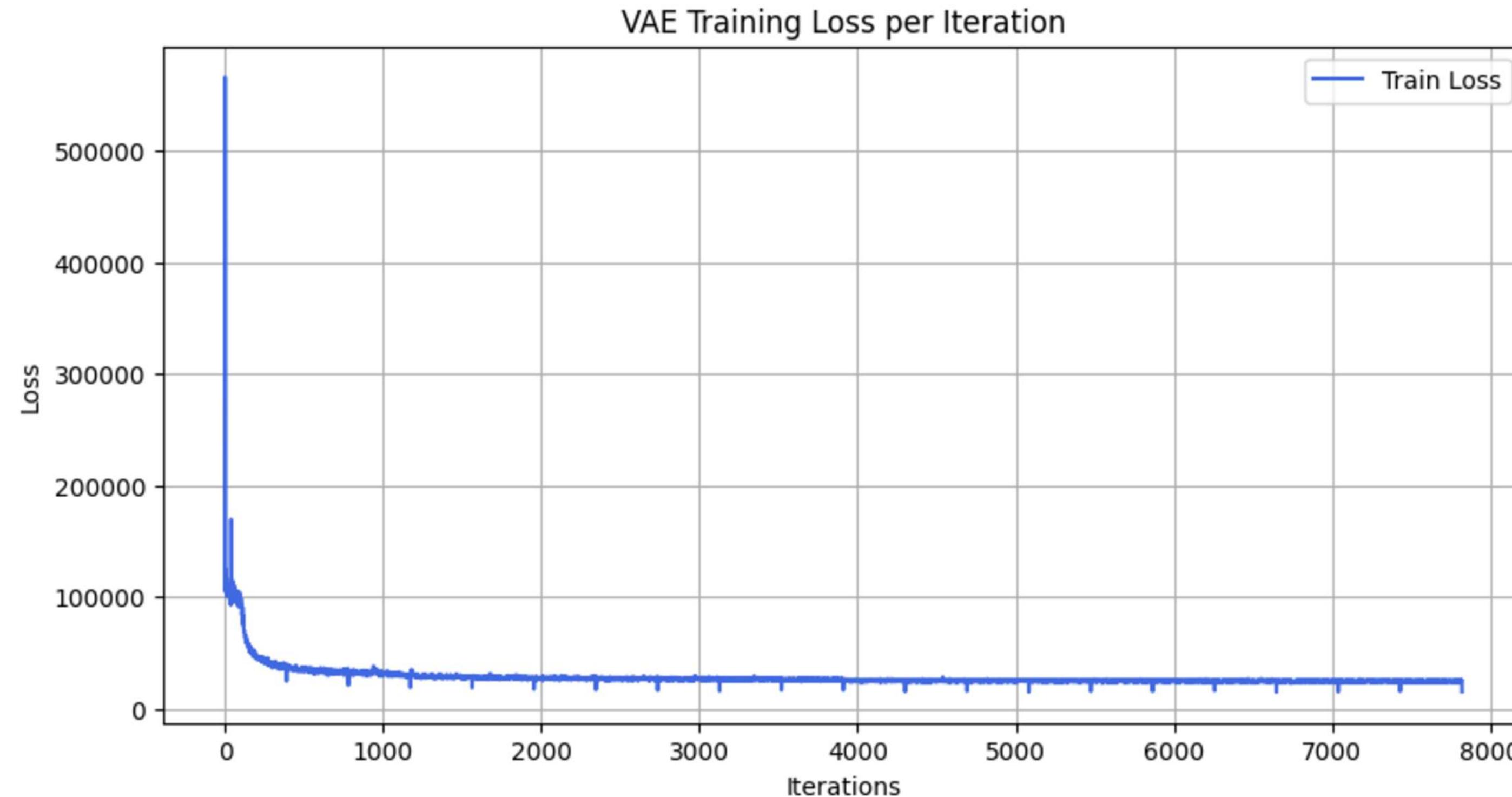
Model	RE	PSNR	SSIM
Autoencoder	0.0036	24.8326	0.8070
Sparse Autoencoder	0.0039	24.4974	0.7951
Denoising Autoencoder	<b>0.0011</b>	<b>29.8509</b>	<b>0.9391</b>
Variational Autoencoder	0.0047	23.6086	0.7775

# Ứng dụng trong bài toán thực tế

- Với tác vụ sinh ảnh, các chỉ số để đánh giá VAE, DCGAN và VAE-GAN là:
  - Precision: Đánh giá chất lượng ảnh tái tạo
  - Recall: Đánh giá độ đa dạng, bao phủ của ảnh tái tạo
  - FID: So sánh phân phối ảnh tái tạo và ảnh thật

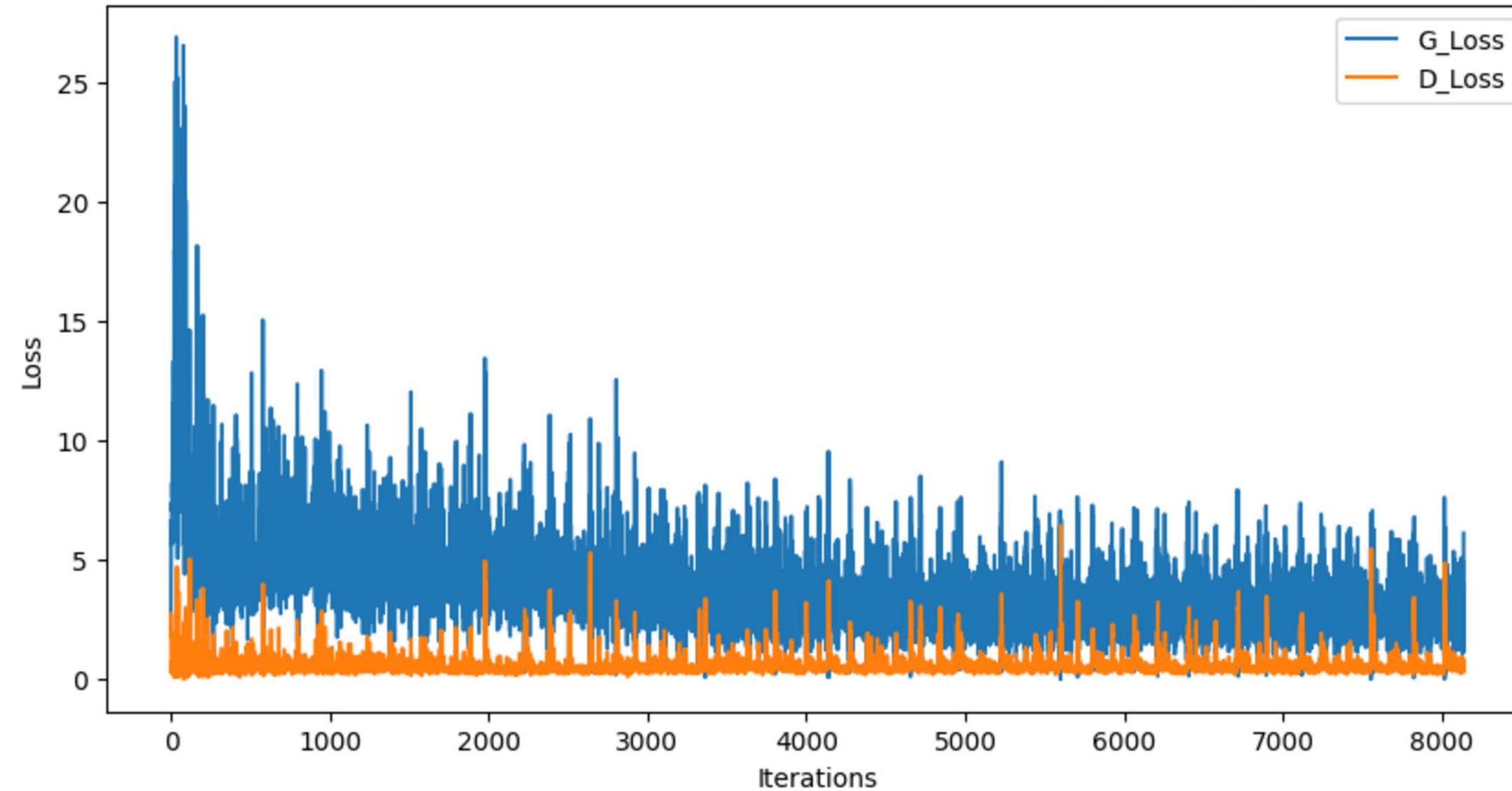


# Ứng dụng trong bài toán thực tế

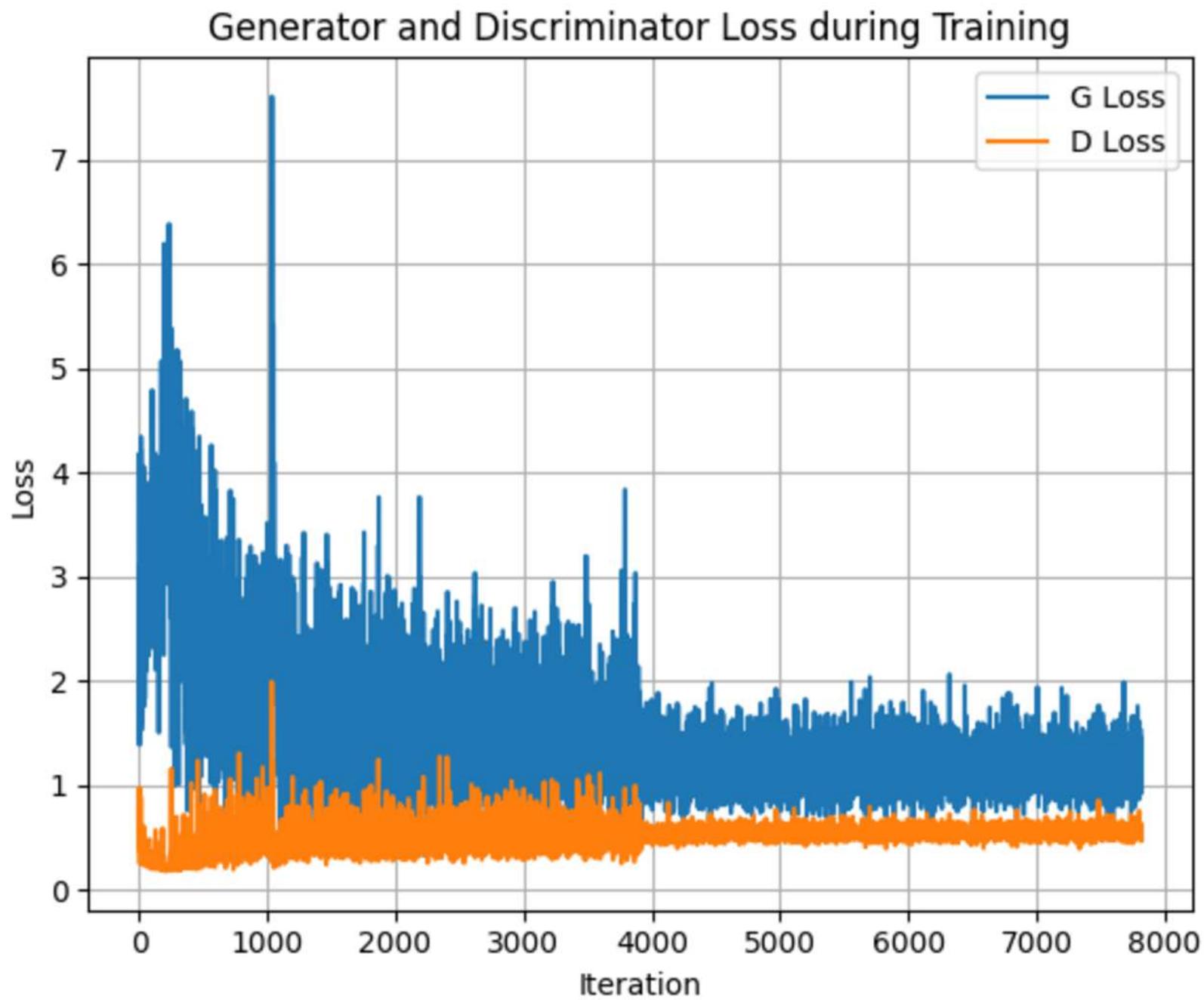
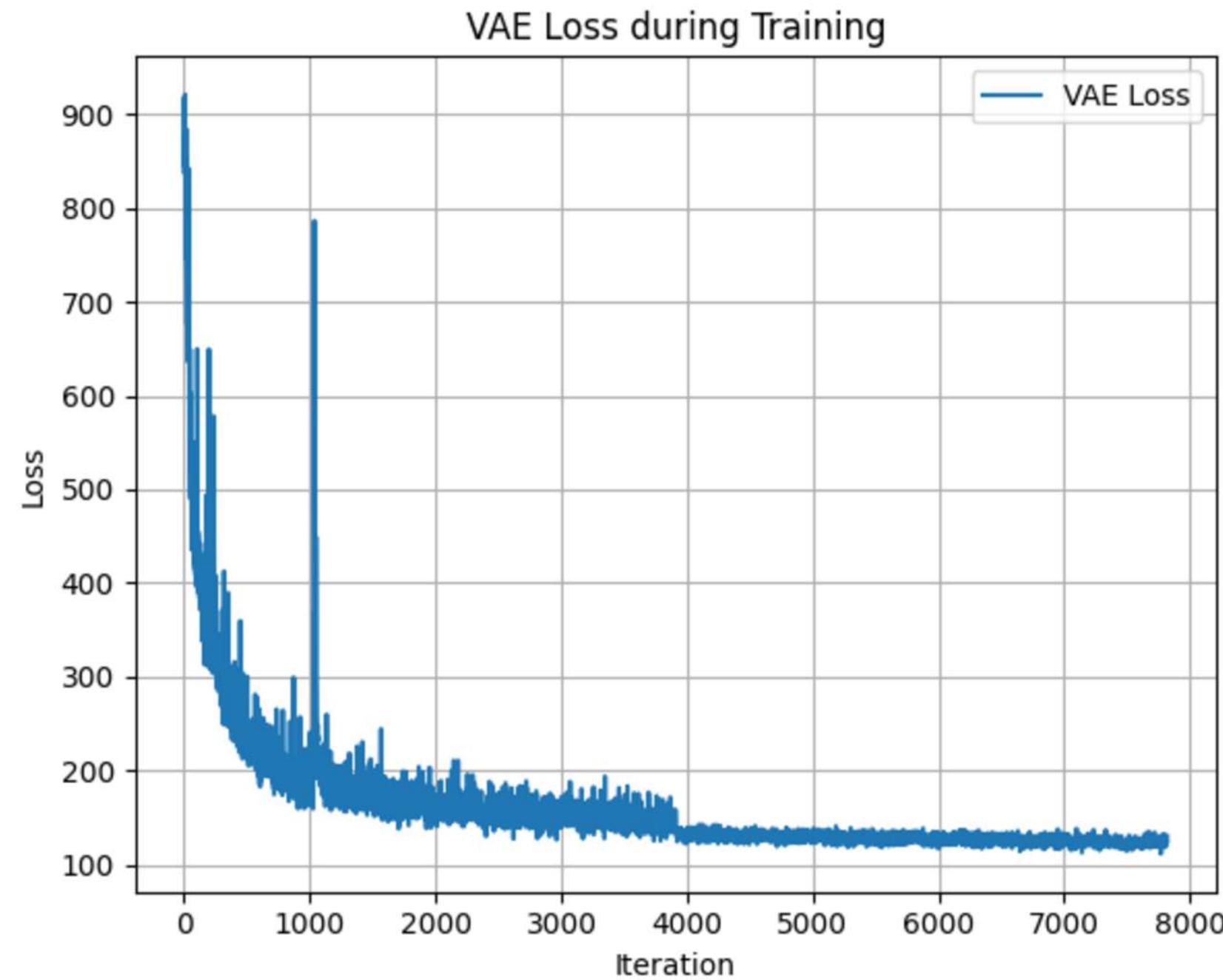


# Ứng dụng trong bài toán thực tế

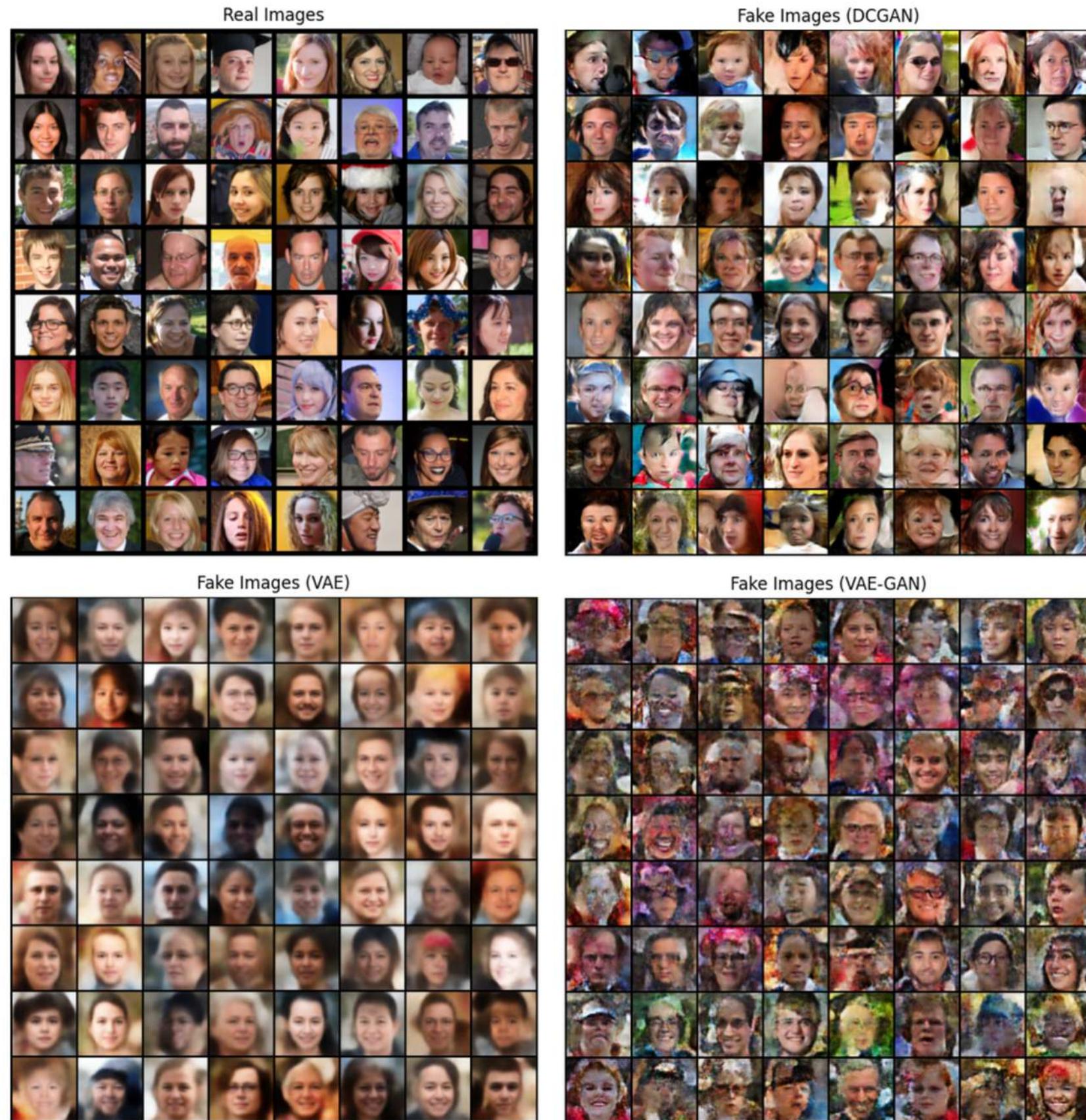
DCGAN Generator and Discriminator Loss During Training



# Ứng dụng trong bài toán thực tế



# Ứng dụng trong bài toán thực tế



# Ứng dụng trong bài toán thực tế

Model	Precision	Recall	FID
VAE	0.0042	0.0011	327.2013
DCGAN	<b>0.0924</b>	<b>0.0192</b>	<b>149.5057</b>
VAE-GAN	0.0040	0.007	303.5775



A large, faint watermark of the HUST logo is visible across the entire background of the slide.

**HUST**

**THANK YOU !**