

Relatório Técnico Final

Conditional Variational Autoencoder (CVAE) aplicado ao dataset
MNIST

Aluno: André Moreira Pimentel

Curso: Bootcamp Cientista de Dados

Instituição: XPE

1. Introdução

O aprendizado profundo (Deep Learning) trouxe avanços expressivos em geração de dados sintéticos, com destaque para os Variational Autoencoders (VAEs). A variação condicional (Conditional Variational Autoencoder – CVAE) estende a formulação original ao condicionar a geração em rótulos, permitindo maior controle sobre o processo. Neste trabalho, desenvolveu-se e avaliou-se um CVAE aplicado ao dataset MNIST, buscando compreender seus limites e potencial em reconstrução e geração de dígitos manuscritos.

2. Fundamentação Teórica

2.1 Variational Autoencoders

Os VAEs combinam autoencoders com princípios de inferência variacional. Diferem de autoencoders clássicos por impor distribuições latentes probabilísticas, o que garante amostragem contínua e suavidade no espaço latente.

2.2 Conditional VAEs

O CVAE introduz rótulos como condicionantes no encoder e decoder. Isso aumenta a capacidade de geração direcionada (ex.: dígito "6" específico) e melhora a estruturação do espaço latente.

2.3 Função de Perda

A função de perda de um CVAE combina:

- Erro de reconstrução (Binary Cross-Entropy ou MSE)
- Kullback–Leibler Divergence (KLD)

3. Metodologia

3.1 Ambiente

Implementação em Python 3.10. Frameworks: PyTorch, Torchvision, NumPy, Pandas, Matplotlib, MLflow. Gerenciamento: Poetry e Typer CLI.

3.2 Dataset

MNIST: 70.000 imagens de dígitos manuscritos (28×28 pixels, tons de cinza). Classes: 10 dígitos (0–9).

3.3 Arquitetura

Encoder e decoder condicionados a rótulos. Espaço latente de dimensão 100. Treinamento supervisionado com classes como condicionantes.

3.4 Hiperparâmetros

```
dataset:
batch_size : 128
test_batch_size: 128
image_size: 28
num_classes: 10

model:
latent_dim : 100
seed: 1

train:
lr: 0.001
momentum: 0.9
epochs: 10
```

4. Resultados

n_images	bce_per_image	mse_per_image	ssim	psnr
10000	67.668683	6.45037363	0.915203	21.497921

As métricas obtidas indicam:

- **BCE**: 67.66 – boa reconstrução, mas longe da perfeição.
- **MSE**: 6.45 – consistente com BCE, penaliza mais erros grandes.
- **SSIM**: 0.91 – preservação estrutural excelente (>0.9).
- **PSNR**: 21.49 dB – aceitável, mas ideal seria >30 dB.

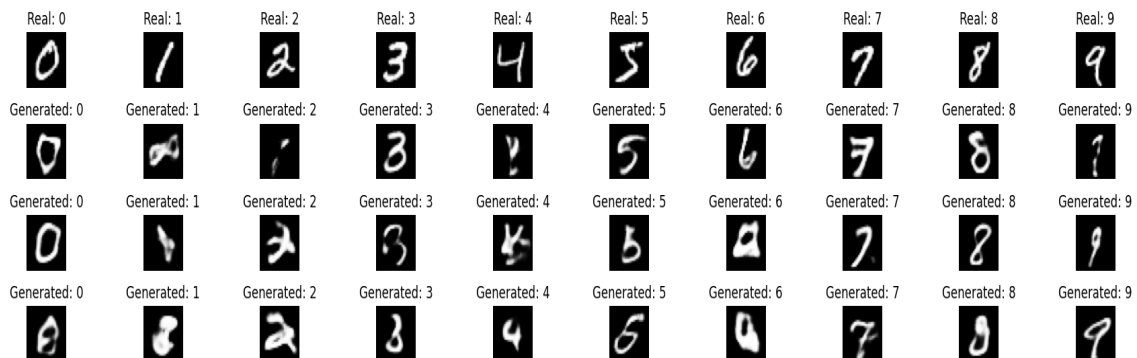


Figura 1 – Comparação de dígitos reais e gerados pelo CVAE.

5. Discussão

Apesar de métricas favoráveis, o modelo apresenta limitações:

- Blurring residual em alguns dígitos.
- Latent_dim fixo pode restringir representações.
- Treinamento curto (10 epochs).

Recomendações: β -VAE, embeddings de rótulo, data augmentation, maior latent_dim, treinamento mais longo, pipeline CI/CD com MLflow.

6. Conclusão

O CVAE treinado no MNIST demonstrou capacidade de reconstrução com alta similaridade estrutural ($SSIM=0,91$), mas limitações em nitidez e fidelidade ($PSNR \approx 21$ dB). Ajustes na arquitetura, hiperparâmetros e pipeline experimental podem aumentar a qualidade dos resultados.

7. Referências

- Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. arXiv:1312.6114.
- Sohn, K., Lee, H., & Yan, X. (2015). Learning Structured Output Representation using Deep Conditional Generative Models. NeurIPS.
- Zhang, R., et al. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. CVPR.
- Repositório GitHub do projeto: github.com/moreira-and/cvae-mnist