

Aprendizado Auto-Supervisionado de Espaços Latentes com Poucas Amostras: Impacto da Aumentação de Dados por Difusão e Validação via Agrupamento

Prof. Alexandre Xavier Falcão

MO433/MC934
10 de novembro de 2025

Enunciado

As técnicas de aprendizado auto-supervisionado estudadas anteriormente — SimCLR, BYOL, DINO e DINO-MultiCrop — envolvem aumento de dados como parte essencial do projeto de um extrator de características de imagem (*encoder*). Métodos tipicamente utilizados, como jittering, podem alterar significativamente a aparência da imagem e as características principais relacionadas à classe.

Neste projeto, exploraremos alternativas à aumento tradicional, incluindo modelos de difusão [1] e técnicas que dispensam aumento explícito, como CNN-JEPA [2]. Também avaliaremos o uso de modelos de difusão para melhorar diretamente a representação latente de um *autoencoder*, como proposto no artigo DGAE [3].

O objetivo é reavaliar o agrupamento de imagens da base Corel com diferentes abordagens de aprendizado auto-supervisionado e aumento de dados, comparando os resultados com métricas de agrupamento utilizadas em projeto anterior.

Tarefas

1. Estudo sobre modelos de difusão estável (Stable Diffusion) [1], especialmente em cenários com poucas amostras, e suas aplicações no aprendizado auto-supervisionado de espaços latentes (*encoders*).
2. Criar metadados para a base Corel e adaptar os códigos 6 e 7 da aula de difusão estável para a geração de imagens da base Corel com modelo pré-treinado. Verificar se é necessário ou não adaptar um modelo por classe.
3. Com base no estudo do item 1, melhorar os códigos 4 e 5 da aula de difusão estável para a geração de imagens da base Corel com modelo treinado do zero. Verificar se é necessário ou não treinar um modelo por classe.

4. Implementar a abordagem do artigo DGAE [3], utilizando o modelo pré-treinado do item 2 (ou outro de sua escolha) para treinar um *autoencoder* convolucional. O objetivo aqui não é a augmentação de dados, mas sim o aprendizado de características latentes para fins de agrupamento.
5. Avaliar o agrupamento das imagens da base Corel (com e sem augmentação por difusão) utilizando as seguintes técnicas de extração de características:
 - Uma técnica auto-supervisionada entre SimCLR, BYOL, DINO ou DINO-MultiCrop.
 - CNN-JEPA [2].
 - DGAE [3].

Estrutura do Relatório

1. Introdução

2. Motivação

- Revisão das técnicas auto-supervisionadas.
- Difusão como augmentação e como guia latente.

3. Metodologia

- Descrição da base Corel.
- Implementações dos itens 2, 3 e 4.
- Técnicas de extração de características (implementação do item 5):
 - Uma técnica auto-supervisionada entre SimCLR, BYOL, DINO ou DINO-MultiCrop.
 - CNN-JEPA [2].
 - DGAE [3].

4. Resultados

- Imagens geradas.
- Qualidade das representações.
- Métricas de agrupamento.

5. Discussão

- Comparação entre abordagens.
- Impacto da augmentação por difusão.
- Eficiência do DGAE vs. CNN-JEPA.

6. Conclusão

- Principais achados.
- Limitações e sugestões futuras.

Referências

Referências

- [1] Chieh-Hsin Lai, Yang Song, Dongjun Kim, Yuki Mitsufuji, and Stefano Ermon. The Principles of Diffusion Models: From Origins to Advances. *arXiv preprint arXiv:2510.21890*, 2025.
- [2] András Kalapos and Bálint Gyires-Tóth. *CNN-JEPA: Self-Supervised Pretraining Convolutional Neural Networks Using Joint Embedding Predictive Architecture*. arXiv preprint arXiv:2408.07514v2, 2025.
- [3] Dongxu Liu, Yuang Peng, Haomiao Tang, Yuwei Chen, Chunrui Han, Zheng Ge, Daxin Jiang, Mingxue Liao. *DGAE: Diffusion-Guided Autoencoder for Efficient Latent Representation Learning*. arXiv preprint arXiv:2506.09644, 2025.