

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Tese de Doutorado

Área: Computação Inteligente

Título: Explorando A Multimodalidade em Modelos Generativos baseados em Difusão

Orientador – Carmelo José Albanez Bastos Filho (carmelo.filho@upe.br)

Descrição – Os Modelos de Difusão (MDs) têm se destacado como o estado da arte em Inteligência Artificial (IA) para problemas generativos, superando em alguns aspectos abordagens tradicionais como GANs e VAEs. Operando em um paradigma encoder-decoder, esses modelos mapeiam a entrada em um espaço latente e a reconstróem iterativamente, simulando um processo de difusão progressiva e reversa. Ho et al. (2020) [1] introduziram o conceito de Modelos Probabilísticos de Difusão com Denoising (DDPMs), estabelecendo as bases teóricas para muitos trabalhos subsequentes.

Embora inicialmente propostos para remoção de ruído em imagens, os modelos de difusão demonstraram grande versatilidade e estão sendo aplicados em diversos campos. Dhariwal & Nichol (2021) [2] mostraram que esses modelos podem superar GANs na síntese de imagens, alcançando resultados superiores em termos de qualidade e diversidade. Além disso, Rombach et al. (2022) [3] propuseram uma abordagem para gerar imagens de alta resolução usando modelos de difusão latentes, permitindo a criação de imagens detalhadas e realistas.

Esta pesquisa propõe uma revisão sistemática da literatura sobre modelos de difusão, abrangendo desde os trabalhos pioneiros de Ho et al. (2020) [1] até os avanços mais recentes, como a técnica de classifier-free diffusion guidance proposta por Ho & Salimans (2021) [4]. Além de comparar diferentes tipos de modelos de difusão, como DDPMs e modelos baseados em Equações Diferenciais Estocásticas (SDEs). Esta pesquisa tem como objetivo explorar a multimodalidade em modelos de difusão, investigando e propondo técnicas para integrar informações de áudio e vídeo, como proposto por Chen et al. (2024) [5]. O objetivo é expandir as capacidades e aplicações dos modelos de difusão, impulsionando o campo da IA generativa em direção a novas possibilidades criativas e aplicações inovadoras.

Referências Bibliográficas

- [1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33.
- [2] Dhariwal, P., & Nichol, A. (2021). Diffusion models beat GANs on image synthesis. *arXiv preprint arXiv:2105.05233*.
- [3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695).
- [4] Ho, J., & Salimans, T. (2021). Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.
- [5] Chen, C., Ding, H., Sisman, B., Xu, Y., Xie, O., Yao, B., ... & Zeng, B. (2024). Diffusion models for multi-modal generative modeling. In *International Conference on Learning Representations (ICLR)*.