

Uma Introdução ao Aprendizado Não-Supervisionado

Modelagem Generativa Visual e Síntese de Imagens

Hugo Oliveira¹

¹Departamento de Informática (DPI)

Universidade Federal de Viçosa (UFV)

31 de Janeiro, 2024





Mais informações: <https://sites.google.com/view/oliveirahugo>

Agenda

- 1 Modelos de Diffusion
- 2 Outras Aplicações de Modelos Generativos
 - Uma Volta dos Modelos Autorregressivos
 - Tradução de Imagens
 - Vídeo
 - Self-Supervised Learning
 - Open Set Recognition
 - Síntese 3D

Agenda

1 Modelos de Diffusion

2 Outras Aplicações de Modelos Generativos

- Uma Volta dos Modelos Autorregressivos
- Tradução de Imagens
- Vídeo
- Self-Supervised Learning
- Open Set Recognition
- Síntese 3D

Modelagem Generativa no Deep Learning

Tipos de Modelos Generativos Deep

- Redes Generativas Adversariais (*Generative Adversarial Networks – GANs*);
- AutoEncoders Variacionais (*Variational AutoEncoders – VAEs*);
- *Normalizing Flows*;
- *Diffusion*;

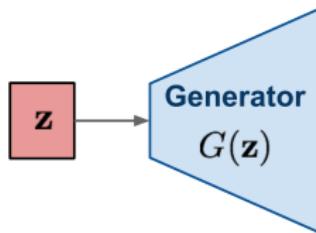
Generative Adversarial Networks (GANs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

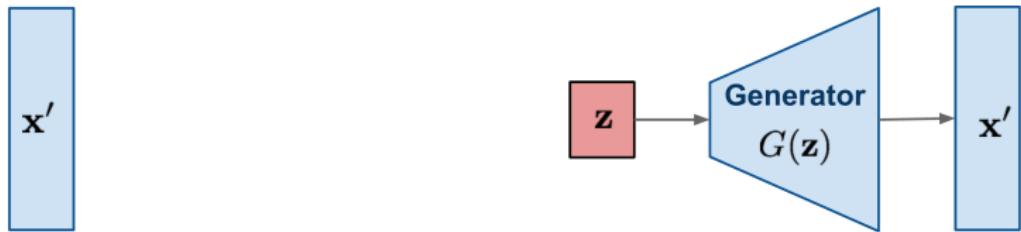
Generative Adversarial Networks (GANs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

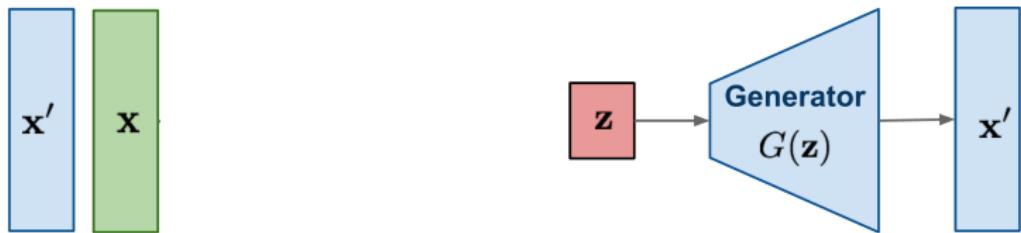
Generative Adversarial Networks (GANs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

Generative Adversarial Networks (GANs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

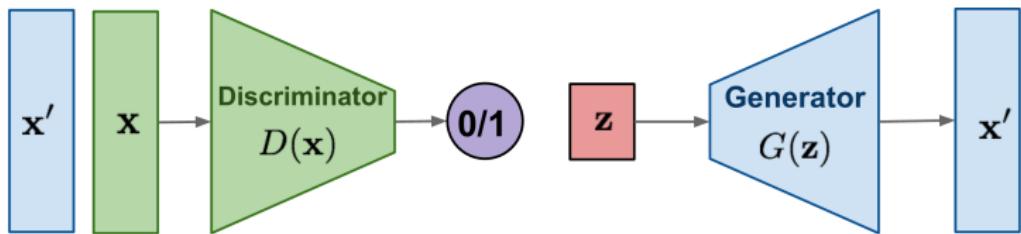
Generative Adversarial Networks (GANs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

Generative Adversarial Networks (GANs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

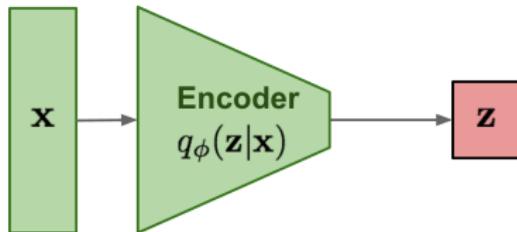
Variational AutoEncoders (VAEs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

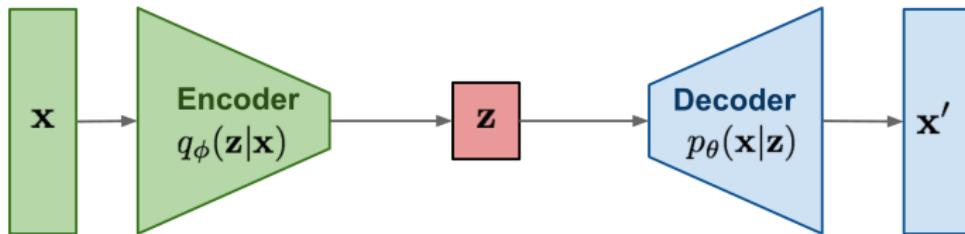
Variational AutoEncoders (VAEs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

Variational AutoEncoders (VAEs)



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

Taxonomia de Modelos Generativos

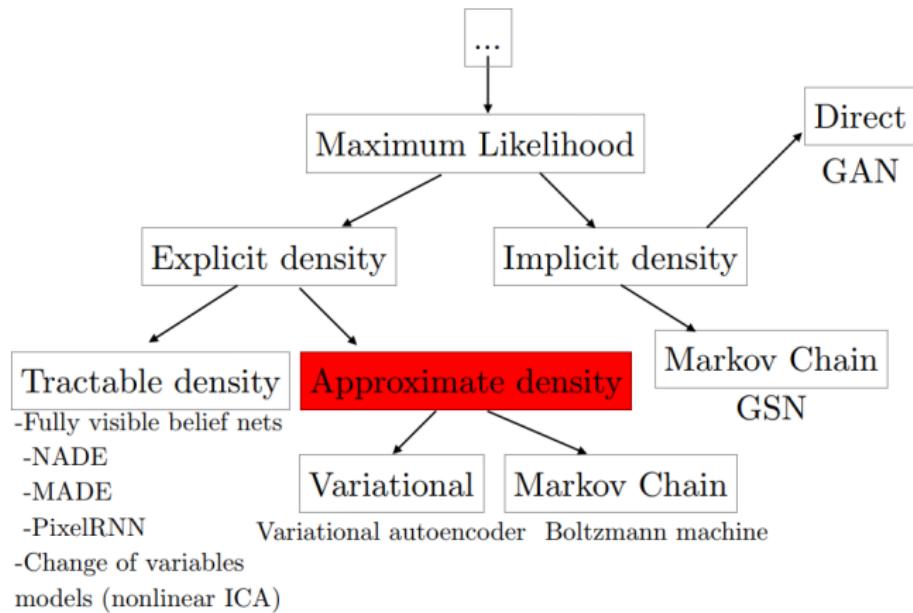
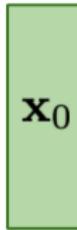


Figura: Uma taxonomia (incompleta) de modelos generativos¹.

¹<https://arxiv.org/abs/1701.00160>

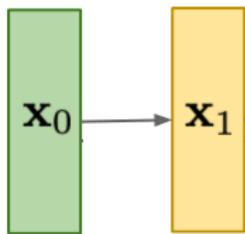
Diffusion



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

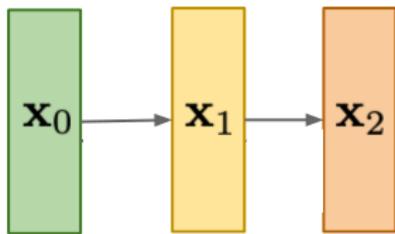
Diffusion



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

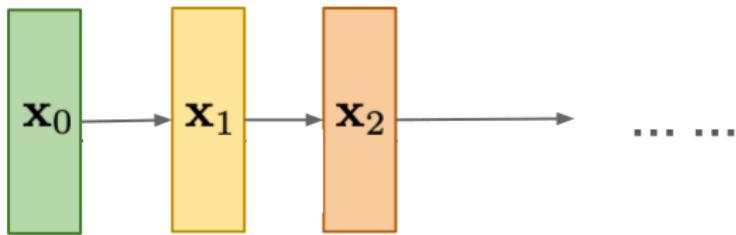
Diffusion



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

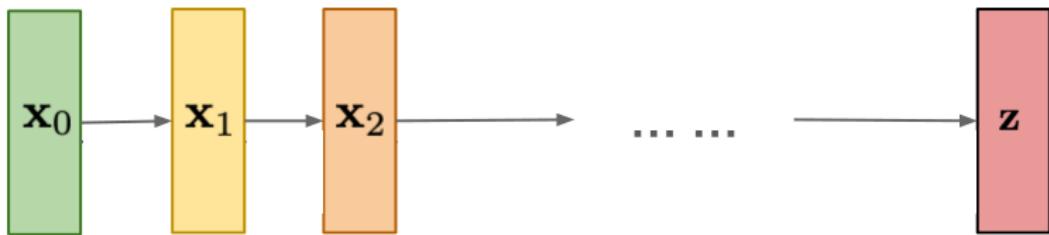
Diffusion



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

Diffusion



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

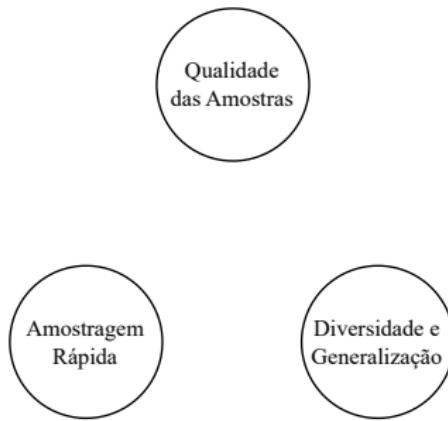
Diffusion



Fonte: Lilian Weng¹

¹<https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>

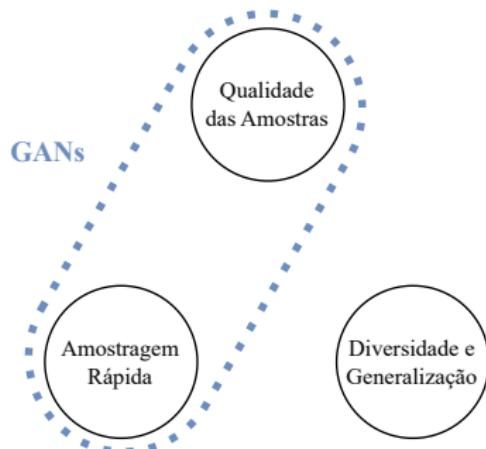
O Trilema da Modelagem Generativa



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

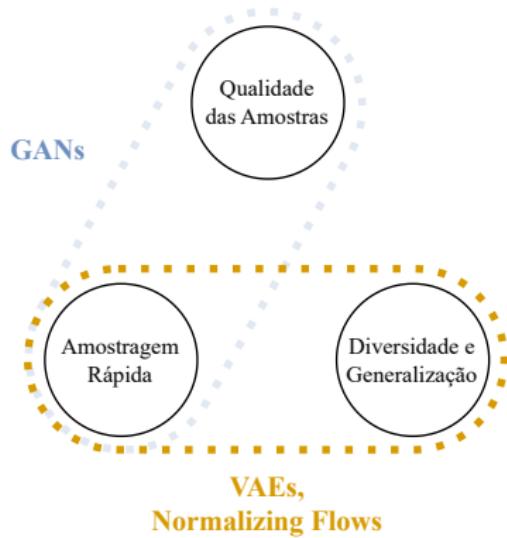
O Trilema da Modelagem Generativa



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

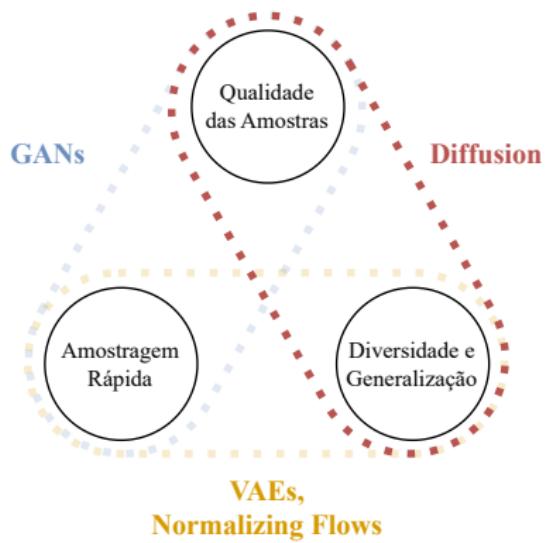
O Trilema da Modelagem Generativa



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

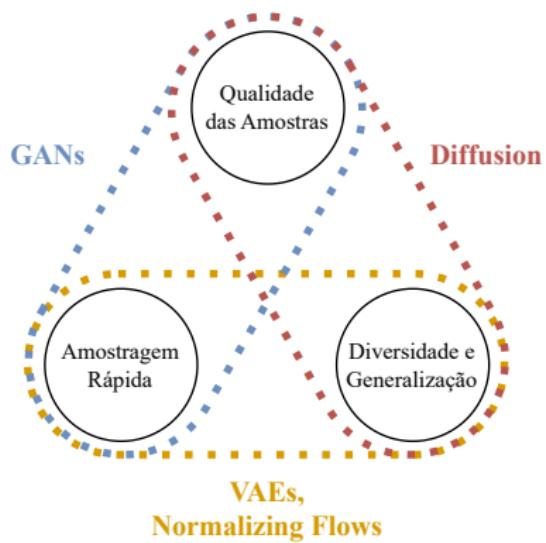
O Trilema da Modelagem Generativa



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

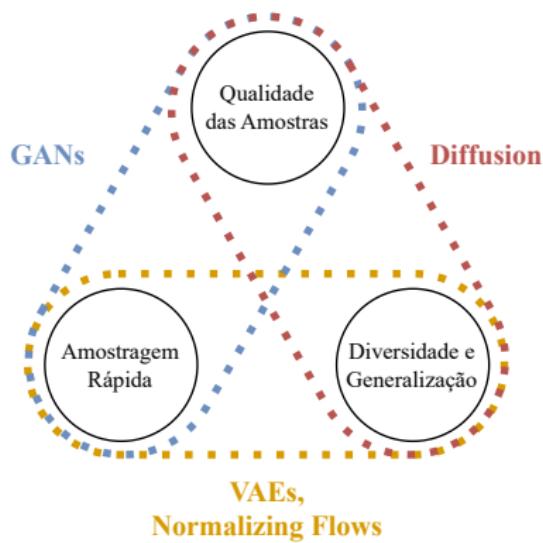
O Trilema da Modelagem Generativa



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

O Trilema da Modelagem Generativa



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

Text-to-Image

- Desde 2020 surgiram múltiplas estratégias de síntese de imagens condicionada a um prompt de texto
 - DALL-E 2¹ da OpenAI
 - Google Imagen²
 - Midjourney³
 - Stable Diffusion⁴

¹<https://openai.com/dall-e-2/>

²<https://Imagen.research.google/>

³<https://midjourney.com/>

⁴<https://stability.ai/blog/stable-diffusion-public-release>

O que é Diffusion?

Imagen



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

O que é Diffusion?

Imagen



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

O que é Diffusion?

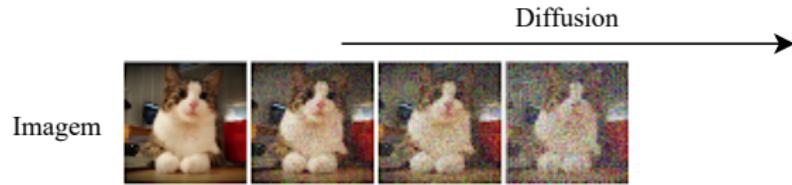
Imagen



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

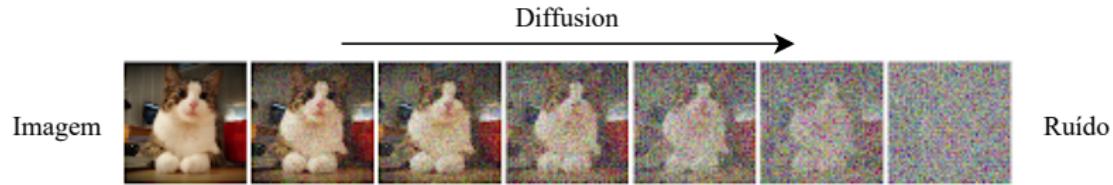
O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

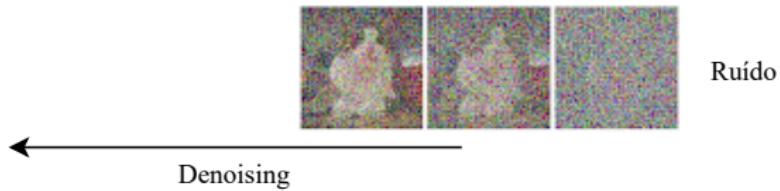
O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

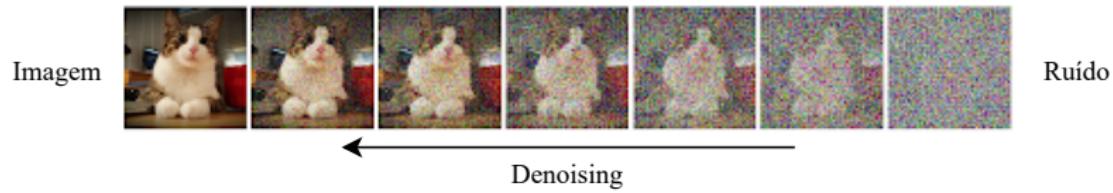
O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

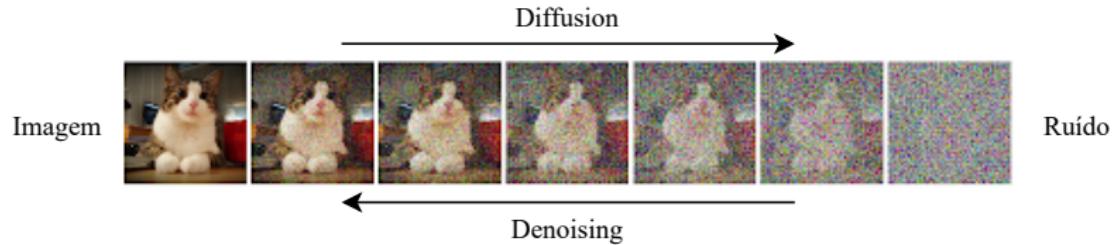
O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

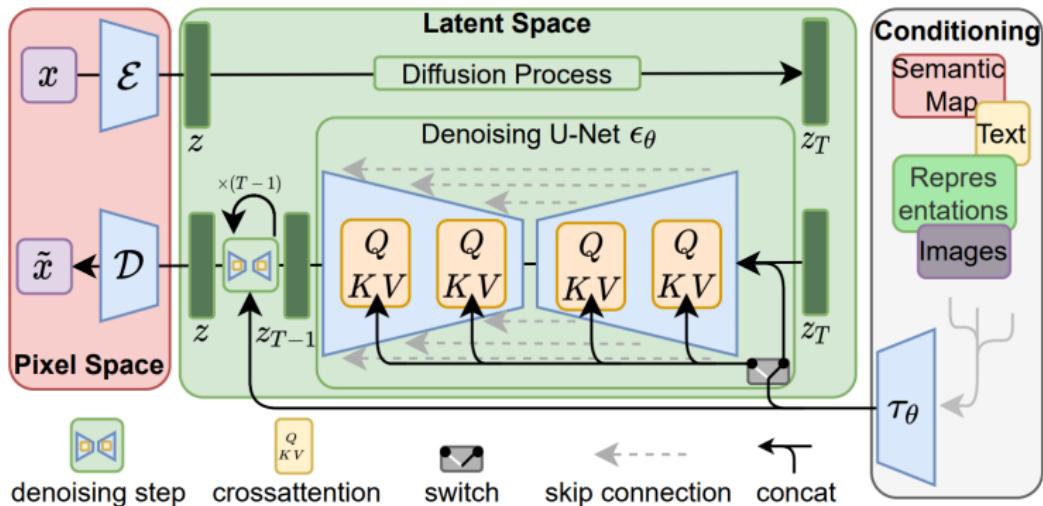
O que é Diffusion?



Fonte: NVIDIA Developer¹

¹<https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/>

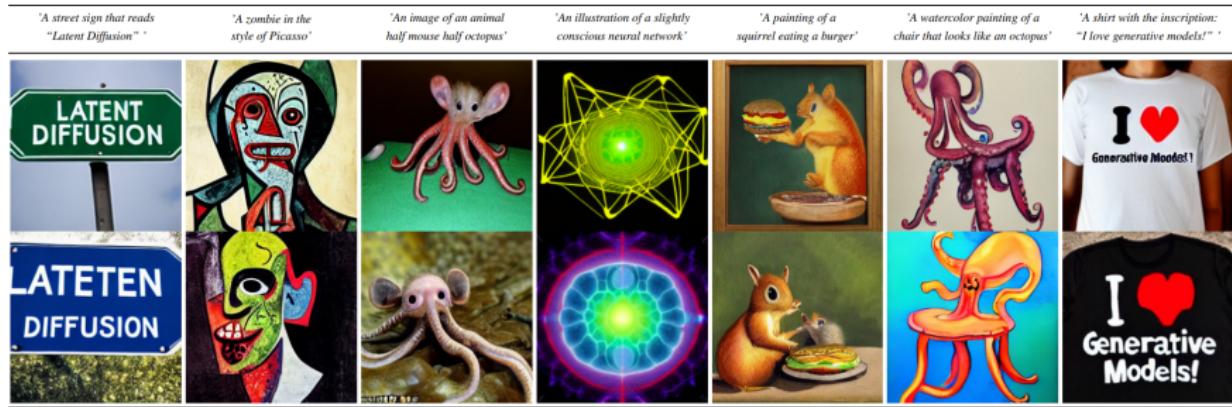
Stable Diffusion



Fonte: Stable Diffusion¹

¹<https://stability.ai/blog/stable-diffusion-public-release>

Stable Diffusion



Fonte: Stable Diffusion¹

¹<https://stability.ai/blog/stable-diffusion-public-release>

Stable Diffusion

Demo - Stable Diffusion

stable_diffusion.ipynb

Diffusion

Limitações

- Custo computacional bastante elevado
- Dependência da qualidade do modelo de texto
- Funcionam bem em conjuntos de dados grandes

Agenda

1 Modelos de Diffusion

2 Outras Aplicações de Modelos Generativos

- Uma Volta dos Modelos Autorregressivos
- Tradução de Imagens
- Vídeo
- Self-Supervised Learning
- Open Set Recognition
- Síntese 3D

Modelos Autorregressivos Modernos

Autorregressão Moderna

Recentemente os modelos Autorregressivos voltaram a ocupar parte do estado-da-arte da modelagem generativa em alguns cenários. Modelos como o Parti¹ do Google utilizam uma estratégia completamente diferente de autorregressão.

¹<https://sites.research.google/parti/>

Modelos Autorregressivos Modernos

Sequence-to-Sequence

O Parti trata o problema de síntese de imagens a partir de prompts de texto como um problema de “tradução” de uma sequência de palavras para uma sequência de tokens de imagens. Tanto o texto quanto as imagens são codificadas e decodificadas por redes Transformer, garantindo acesso ao contexto global.

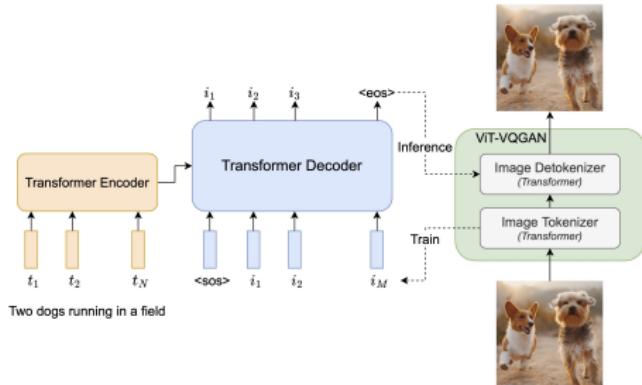
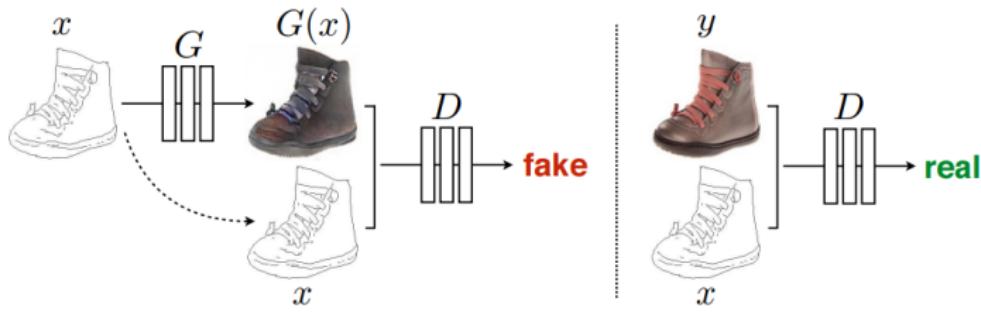


Figura: Arquitetura do Parti¹.

Tradução de Imagens

- Tradução de Imagens Pareadas (Tradução Supervisionada)^{1,2}
 - Imagens **pareadas** nos dois domínios
 - Normalmente loss L1 entre a imagem traduzida e a original do outro domínio
 - Generativa agora é uma arquitetura Encoder-Decoder (U-nets, SegNets, Autoencoders...)
 - Discriminativa agora está **condicionada** à imagem de entrada



¹<https://arxiv.org/abs/1611.07004>

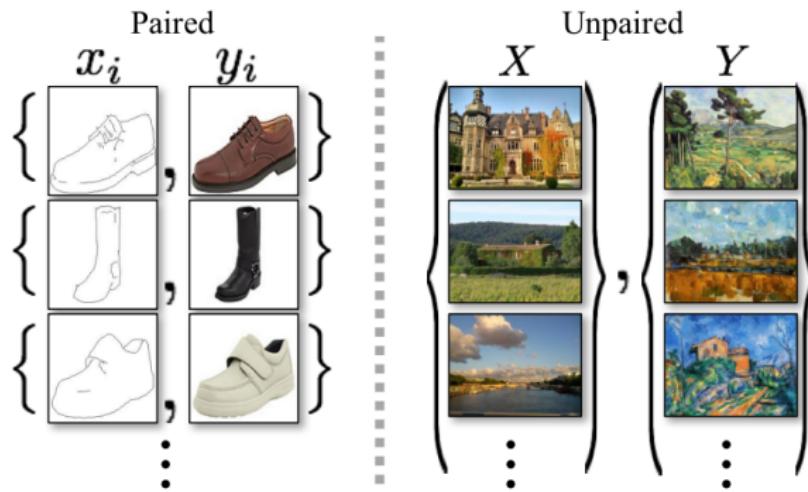
²<https://arxiv.org/abs/1711.11586>

Tradução de Imagens

- Tradução de Imagens Pareadas (Tradução Supervisionada)^{1,2}
 - Imagens **pareadas** nos dois domínios
 - Normalmente loss L1 entre a imagem traduzida e a original do outro domínio
 - Generativa agora é uma arquitetura Encoder-Decoder (U-nets, SegNets, Autoencoders...)
 - Discriminativa agora está **condicionada** à imagem de entrada

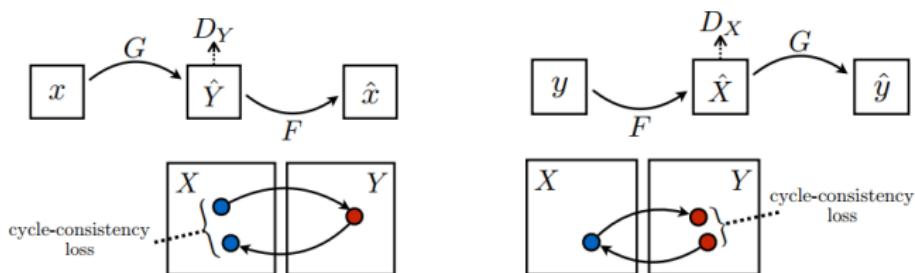
Tradução de Imagens

- O pareamento das imagens limita muito as aplicações práticas dessas redes.



Tradução de Imagens

- Tradução Não-Pareada (Tradução Não-Supervisionada)¹²
 - Imagens não pareadas dos dois domínios
 - Ideia de Cycle Consistency
 - Loss é obtida ao comparar a imagem com ela mesma

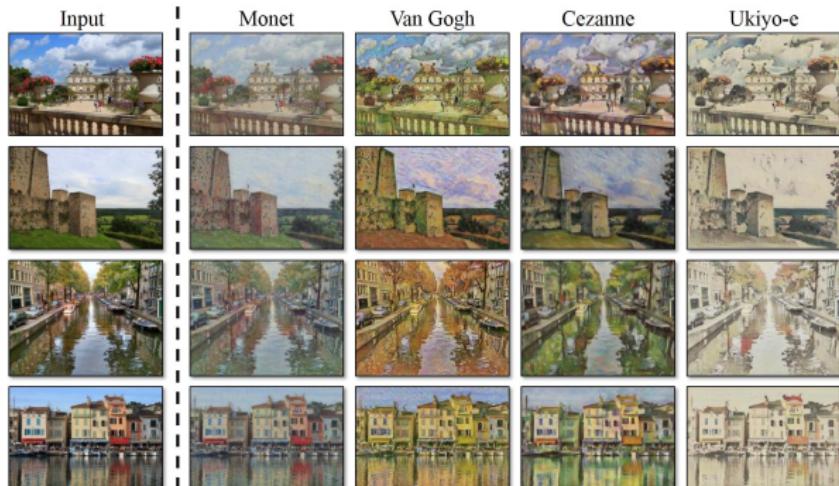


¹<https://arxiv.org/pdf/1703.10593.pdf>

²<https://arxiv.org/abs/1905.01723>

Tradução de Imagens

- Tradução Não-Pareada (Tradução Não-Supervisionada)¹²
 - Imagens não pareadas dos dois domínios
 - Ideia de Cycle Consistency
 - Loss é obtida ao comparar a imagem com ela mesma



¹<https://arxiv.org/pdf/1703.10593.pdf>

²<https://arxiv.org/abs/1905.01723>

Tradução de Imagens

- Tradução Não-Pareada (Tradução Não-Supervisionada)¹²
 - Imagens não pareadas dos dois domínios
 - Ideia de *Cycle Consistency*
 - Loss é obtida ao comparar a imagem com ela mesma

¹<https://arxiv.org/pdf/1703.10593.pdf>

²<https://arxiv.org/abs/1905.01723>

Tradução de Imagens

- Tradução Não-Pareada (Tradução Não-Supervisionada)¹²
 - Imagens não pareadas dos dois domínios
 - Ideia de *Cycle Consistency*
 - Loss é obtida ao comparar a imagem com ela mesma

¹<https://arxiv.org/pdf/1703.10593.pdf>

²<https://arxiv.org/abs/1905.01723>

CycleGAN

Demo (Extra) - CycleGAN

cyclegan.ipynb

Processamento de Vídeo

- Video-to-Video Synthesis¹
 - Samples precisam ter coerência temporal
 - Loss Adversarial que força uma correlação temporal correta entre os frames

¹<https://github.com/NVIDIA/vid2vid>

Processamento de Vídeo

- Stable Video Diffusion¹

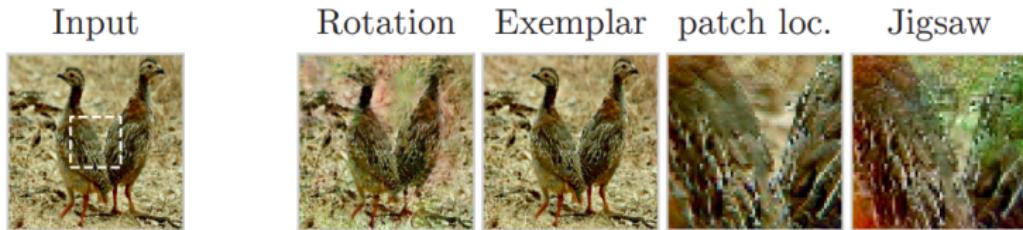
- Assim como GANs e VAEs podem ser adaptados para síntese de vídeo, há modelos de Diffusion que realizam síntese de imagens sequenciais com coerência temporal
- Assim como no caso de imagens estáticas, é possível condicionar a síntese do vídeo via prompts de texto

¹<https://stability.ai/news/stable-video-diffusion-open-ai-video-model>

Aprendendo com Poucos Dados

- **Self-Supervised Learning (SSL)**¹

- Inicialmente SSL foi proposta para texto em tarefas de auto-regressão e para imagens via *data augmentation*
- Rotações, Colorização de Imagens, Localização de Subpatches e Jigsaw podem servir de *pretext tasks* para aprender de forma auto-supervisionada em datasets não rotulados



SSL “tradicional” para imagens. Fonte: Minderer *et al.*².

¹<https://ieeexplore.ieee.org/abstract/document/9086055>

²<https://proceedings.mlr.press/v119/minderer20a/minderer20a.pdf>

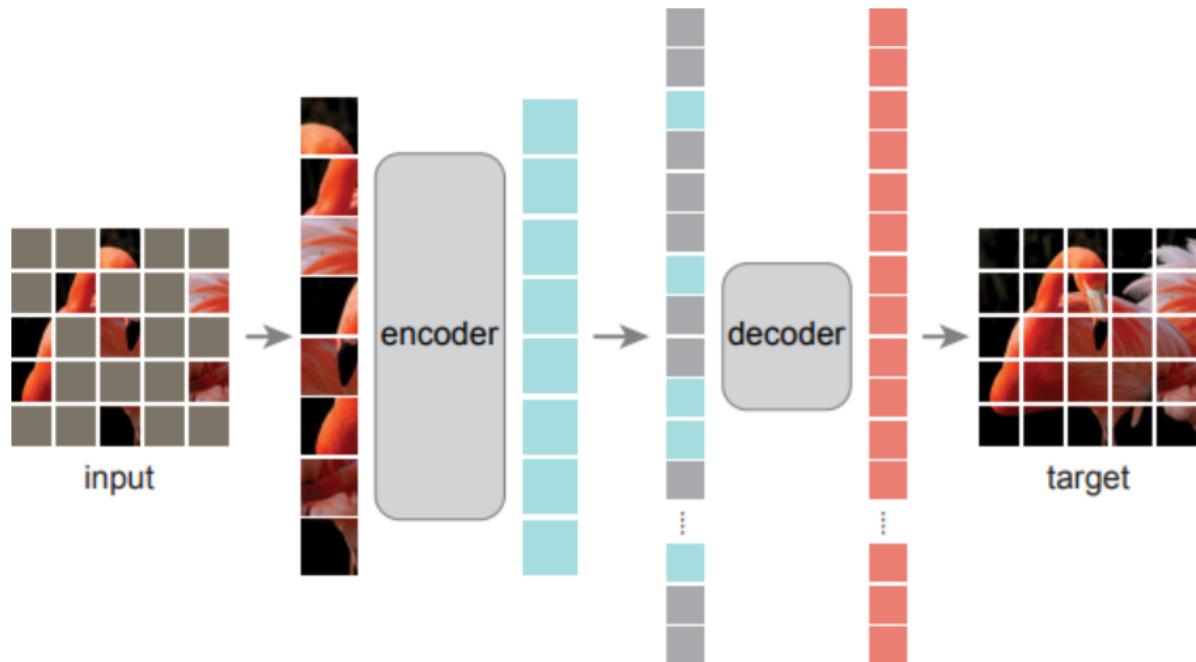
Masked AutoEncoders

- Uma forma simples e muito efetiva de realizar pré-treino não-contrastivo utilizando tanto ViTs¹ quanto CNNs² é por meio de Masked AutoEncoders (MAEs)
- Um MAE simplesmente mascara a maior parte de uma imagem de entrada e tenta reconstruí-la por meio de uma arquitetura Encoder-Decoder
- Apenas a reconstrução dos pixels mascarados é utilizada para o cálculo da loss de um MAE
- Após o pré-treino Self-Supervised, o Encoder pode ser utilizado como backbone e tunado para tarefas Few-Shot

¹<https://arxiv.org/abs/2111.06377>

²<https://arxiv.org/abs/2301.00808>

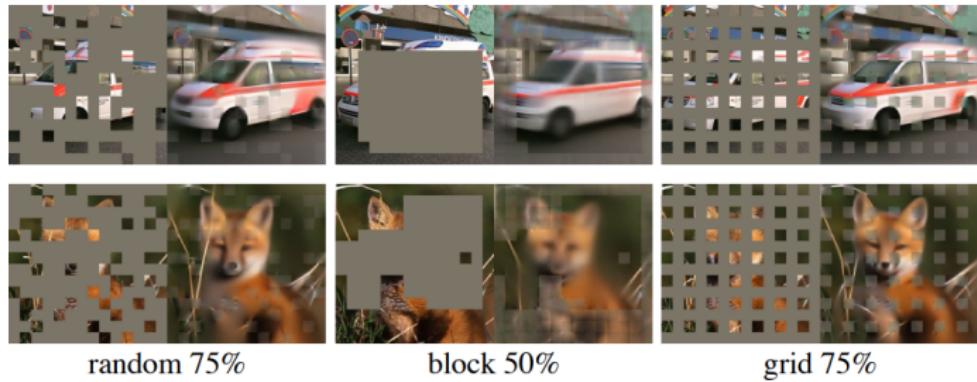
Masked AutoEncoders



Arquitetura de um MAE. Fonte: He *et al.*¹.

¹<https://arxiv.org/abs/2111.06377>

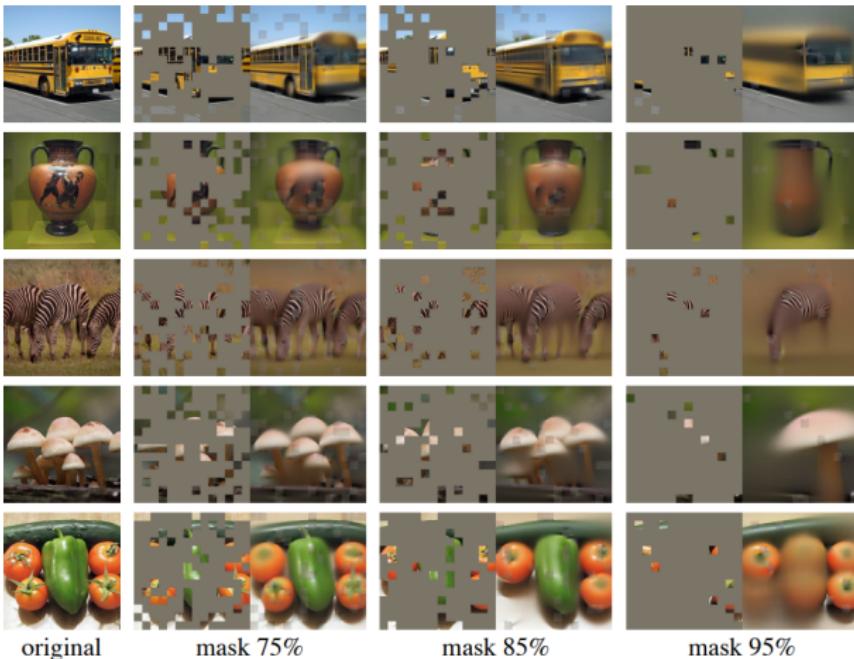
Masked AutoEncoders



Arquitetura de um MAE. Fonte: He *et al.*¹.

¹<https://arxiv.org/abs/2111.06377>

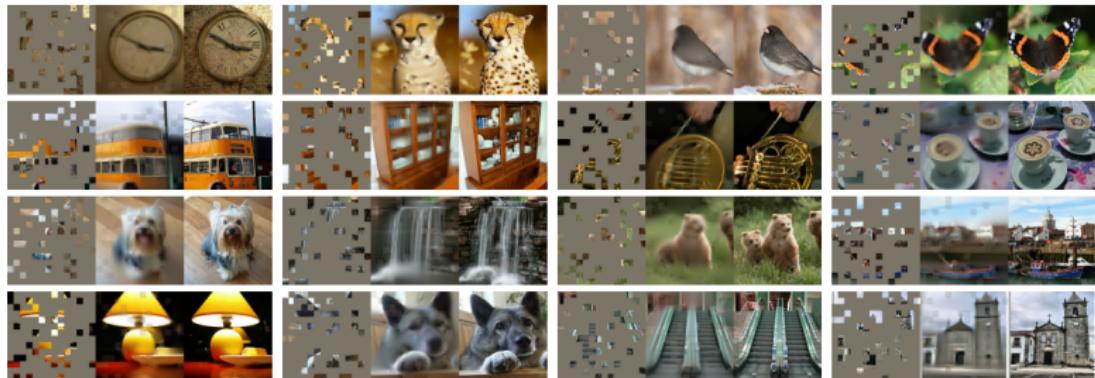
Masked AutoEncoders



Estratégias de Masking de um MAE. Fonte: He et al.¹.

¹<https://arxiv.org/abs/2111.06377>

Masked AutoEncoders



Reconstruções de imagens num MAE. Fonte: He *et al.*¹.

¹<https://arxiv.org/abs/2111.06377>

Masked AutoEncoders

Demo (Extra) - Visualização de um Masked AutoEncoder

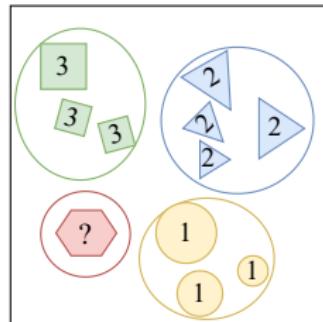
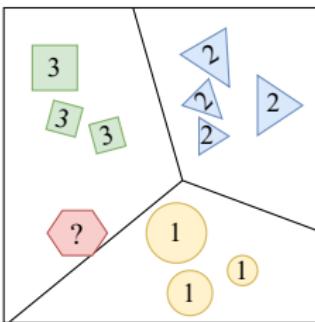
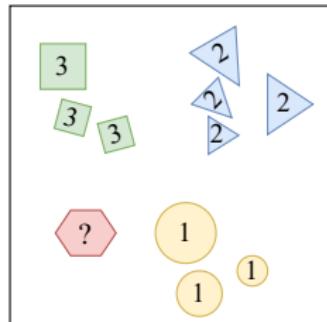
mae_visualize.ipynb

Demo (Extra) - Masked AutoEncoders

mae_vit.ipynb

Open Set Recognition

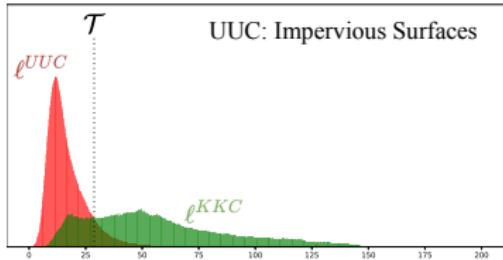
- Problemas do mundo real são menos “comportados” dos que os datasets de visão computacional
- Métodos de aprendizado supervisionado assumem conhecimento total do mundo
- Em problemas reais, há uma variedade de **classes desconhecidas durante o treino** que podem ser vistas durante o teste de uma rede neural



Definição de um problema Open Set. Fonte: Vendramini *et al.*¹.

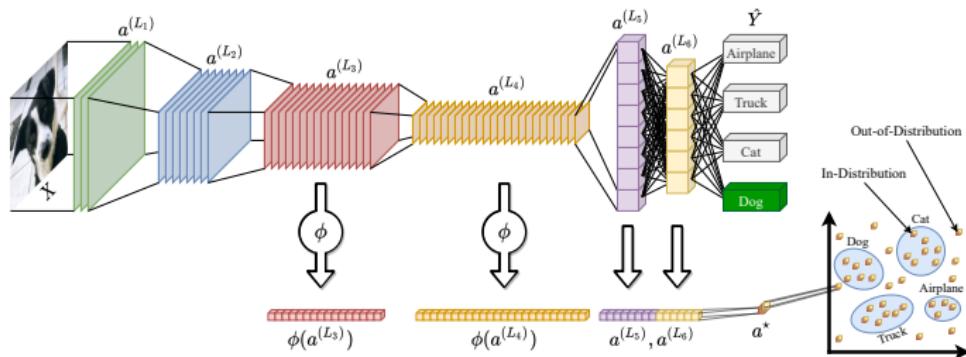
Open Set Recognition

- O GeMOS (Generative Models for Open Set Recognition)¹ faz uso de modelos generativos para delinear as fronteiras entre classes
- Um PCA, por exemplo, usa uma gaussiana multivariada para delinear a distância de uma certa amostra para o centróide da distribuição
- Essa distância pode ser usada para identificar amostras *out-of-distribution* (OOD)



Separação entre as distribuições de classes conhecidas e desconhecidas.

Open Set Recognition



Arquitetura do GeMOS. Fonte: Vendramini et al.¹.

¹<https://arxiv.org/abs/2105.10013>

OSR

Demo (Extra) - GeMOS

gemos_mnist_omniglot.ipynb

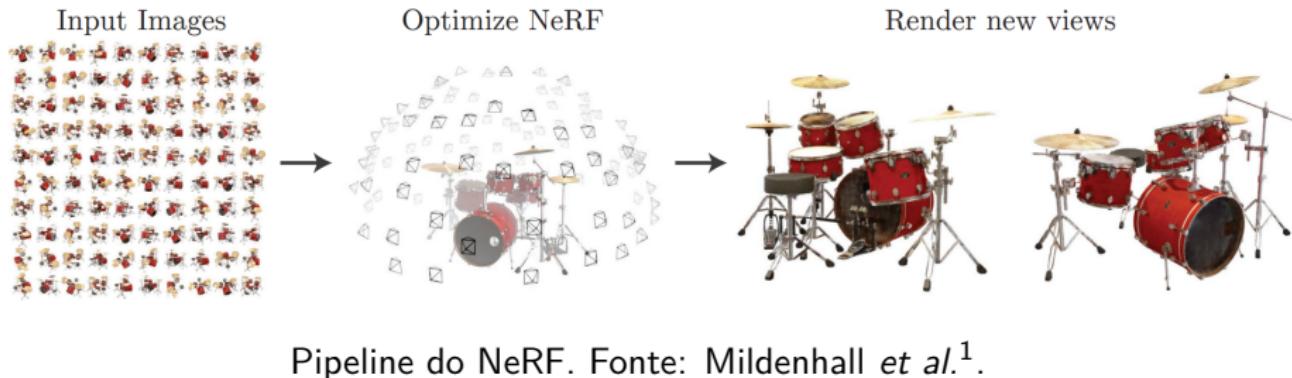
Neural Radiance Fields

- A partir de diferentes *views* esparsamente amostradas de uma mesma cena 3D, é possível treinar uma rede neural para renderizar cenas 3D
- Os primeiros trabalhos a conseguirem resultados expressivos nessa tarefa foram os Neural Radiance Fields (NeRF)

¹<https://dl.acm.org/doi/abs/10.1145/3503250>

Neural Radiance Fields

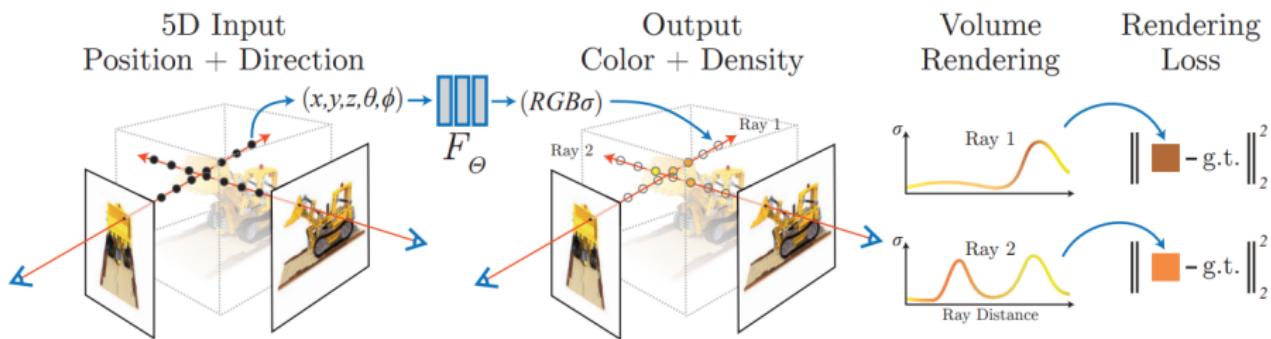
- A partir de diferentes *views* esparsamente amostradas de uma mesma cena 3D, é possível treinar uma rede neural para renderizar cenas 3D
- Os primeiros trabalhos a conseguirem resultados expressivos nessa tarefa foram os Neural Radiance Fields (NeRF)



¹<https://dl.acm.org/doi/abs/10.1145/3503250>

Neural Radiance Fields

- A partir de diferentes *views* esparsamente amostradas de uma mesma cena 3D, é possível treinar uma rede neural para renderizar cenas 3D
- Os primeiros trabalhos a conseguirem resultados expressivos nessa tarefa foram os Neural Radiance Fields (NeRF)

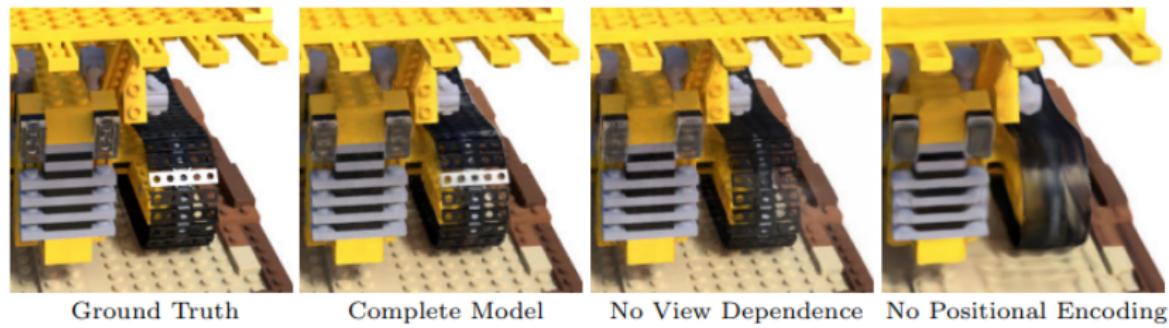


Pipeline do NeRF. Fonte: Mildenhall *et al.*¹.

¹<https://dl.acm.org/doi/abs/10.1145/3503250>

Neural Radiance Fields

- A partir de diferentes *views* esparsamente amostradas de uma mesma cena 3D, é possível treinar uma rede neural para renderizar cenas 3D
- Os primeiros trabalhos a conseguirem resultados expressivos nessa tarefa foram os Neural Radiance Fields (NeRF)



Ablation do NeRF. Fonte: Mildenhall *et al.*¹.

¹<https://dl.acm.org/doi/abs/10.1145/3503250>

Neural Radiance Fields



Resultados do NeRF. Fonte: Mildenhall et al.¹.

Neural Radiance Fields

Demo (Extra) - Neural Radiance Fields

tiny_nerf.ipynb