

Modelagem Generativa Profunda: Uma Revisão Comparativa de VAEs, GANs, Fluxos Normalizantes, Modelos Baseados em Energia e Autoregressivos

Deep Generative Modelling: A Comparative Review of VAEs, GANs,
Normalizing Flows, Energy-Based and Autoregressive Models

André Aparecido da Silva, Luiza Ruivo Marinho Santos, Tiago José da Silva



Artigo publicado pela revista científica *IEEE Transactions on Pattern Analysis and Machine Intelligence* da *IEEE Computer Society* em 28 de março de 2022.



SAM BOND-TAYLOR

Estudante de PhD no
Departamento de Ciência da
Computação na Universidade
de Durham, no Reino Unido.

YANG LONG

Professor assistente no Departamento de
Ciência da Computação na Universidade de
Durham. Trabalha no conselho de pesquisas
médicas no Reino Unido, com o objetivo de
conceber soluções de IA escaláveis para
grandes aplicações na área da saúde.



OS AUTORES

ADAM LEACH

Estudante de PhD no
Departamento de Ciência da
Computação na Universidade
de Durham, no Reino Unido.



CHRIS G. WILLCOCKS

Professor assistente no
Departamento de Ciência da
Computação na Universidade
de Durham, no Reino Unido.



Modelos generativos profundos são uma classe de técnicas que treinam redes neurais profundas para modelar a distribuição de treinamento amostras.

A pesquisa se fragmentou em várias abordagens interconectadas, cada uma das quais faz compensações, incluindo tempo de execução, diversidade e restrições arquitetônicas.

Em particular, este resumo / síntese cobre modelos baseados em energia, autoencoders variacionais, redes adversárias generativas, modelos autorregressivos, fluxos normalizadores, além de inúmeras abordagens híbridas.

Essas técnicas são comparadas e contrastadas, explicando as premissas por trás de cada uma e como elas estão inter-relacionadas, enquanto revisam as técnicas atuais avanços e implementações de última geração.

QUAL O OBJETIVO DO ARTIGO?

Comparar algumas técnicas e abordagens utilizadas para criação e modelos Generativos Profundos (Deep Generative Modelling).

Modelos generativos profundos são uma classe de técnicas que treinam redes neurais profundas para modelar a distribuição de treinamento amostras. A pesquisa se fragmentou em várias abordagens interconectadas, cada uma das quais faz compensações, incluindo tempo de execução, diversidade e restrições arquitetônicas.

Em particular, este compêndio cobre modelos baseados em energia, autoencoders variacionais, redes adversárias generativas, modelos autorregressivos, fluxos normalizadores, além de inúmeras abordagens híbridas. Essas técnicas são comparadas e contrastadas, explicando as premissas por trás de cada uma e como elas estão inter-relacionadas, enquanto revisam as técnicas atuais avanços e implementações de última geração



01

GENERATIVE MODELS

QUAL O OBJETIVO DOS MODELOS GENERATIVOS PROFUNDOS (DEEP GENERATIVE MODELLING)?

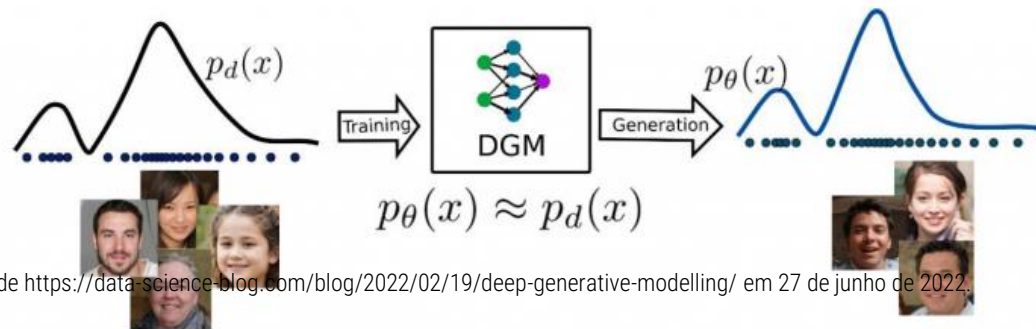
Aprender a gerar novos exemplos a partir de um conjunto de dados analisado e fazendo uso de redes neurais profundas.

Deep generative models (DGM) are neural networks with many hidden layers trained to approximate complicated, high-dimensional probability distributions using a large number of samples. When trained successfully, we can use the DGMs to estimate the likelihood of each observation and to create new samples from the underlying distribution.

Ruthotto and Haber (2021)

Modelos generativos profundos (dgm) são redes neurais com camadas ocultas treinadas para aproximar a distribuição de probabilidade de alta dimensão complicada usando um grande número de amostras. Quando treinados com sucesso, podemos usar o dgm para estimar a probabilidade de cada observação e criar novas amostras a partir da distribuição subjacente

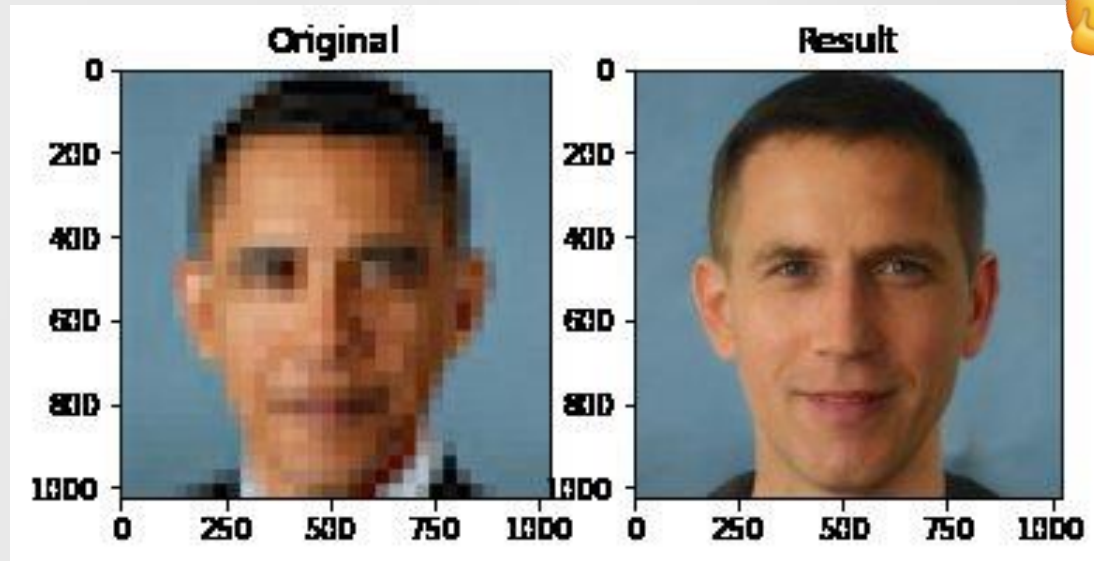
As DGMs treinadas geram novas amostras com a mesma distribuição dos dados treinados. Assim, a probabilidade de x na distribuição resultante é aproximada à probabilidade de x na distribuição treinada.



Extraído de <https://data-science-blog.com/blog/2022/02/19/deep-generative-modelling/> em 27 de junho de 2022.

APLICAÇÕES DE DGMS

Super resolução de imagens



Extraído de <https://zentralwerkstatt.org/blog/pulse> em 27 de Junho de 2022.

APLICAÇÕES DE DGMS

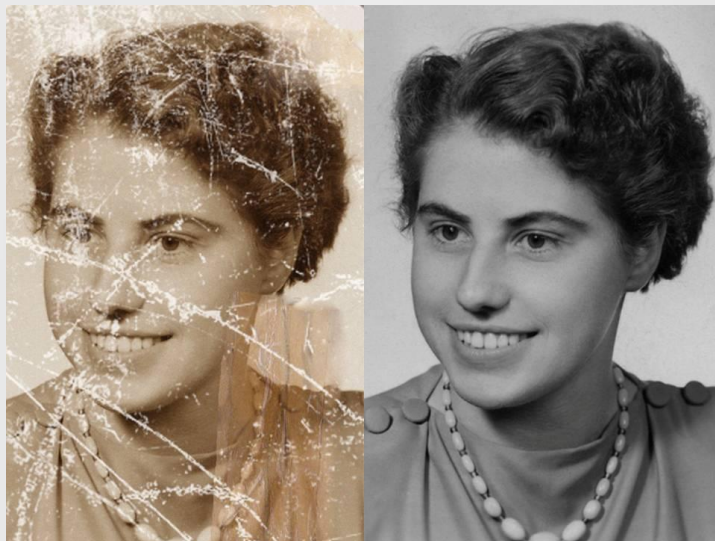
Impainting



Extraído de <https://medium.com/jamieai/image-inpainting-with-deep-learning-dd8555e56a32> em 27 de Junho de 2022.

APLICAÇÕES DE DGMS

Impainting



Extraído de <https://medium.com/jamieai/image-inpainting-with-deep-learning-dd8555e56a32> em 27 de Junho de 2022.

APLICAÇÕES DE DGMS

Síntese de drogas

Research article | [Open Access](#) | [Published: 23 February 2021](#)

Using GANs with adaptive training data to search for new molecules

[Andrew E. Blanchard](#), [Christopher Stanley](#) & [Debsindhu Bhowmik](#) 

[Journal of Cheminformatics](#) **13**, Article number: 14 (2021) | [Cite this article](#)

4456 Accesses | **2** Citations | **7** Altmetric | [Metrics](#)

Extraído de <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-021-00494-3> em 27 de Junho de 2022.

APLICAÇÕES DE DGMS

Outras aplicações:



SÍNTESE DE VOZ E MÚSICA

**SUMARIZAÇÃO E
TRADUÇÃO DE TEXTOS**



SÍNTESE DE VÍDEOS

E outras aplicações diversas...

APLICAÇÕES DE DGMS

Outras aplicações:



SÍNTESE DE VOZ E MÚSICA

**SUMARIZAÇÃO E
TRADUÇÃO DE TEXTOS**



SÍNTESE DE VÍDEOS

E outras aplicações diversas...

APLICAÇÕES DE DGMS

Outras aplicações:



SÍNTESE DE VOZ E MÚSICA

**SUMARIZAÇÃO E
TRADUÇÃO DE TEXTOS**



SÍNTESE DE VÍDEOS

E outras aplicações diversas...



02

TÉCNICAS DO ARTIGO

QUAIS TÉCNICAS FORAM AVALIADAS PELO ARTIGO?

- Energy-based models;
- Variational autoencoders;
- GANs;
- Autoregressive models;
- Normalizing Flows.

COMPARAÇÃO ENTRE MODELOS DE DIVERSOS ARTIGOS AGRUPADOS PARA REVIEW

Method	Train Speed	Sample Speed	Num. Params.	Resolution Scaling	Free-form Jacobian	Exact Density	FID	NLL (in BPD)
Generative Adversarial Networks								
DCGAN [182]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	37.11	-
ProGAN [114]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	15.52	-
BigGAN [19]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	14.73	-
StyleGAN2 + ADA [115]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	2.42	-
Energy Based Models								
IGEBM [46]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	37.9	-
Denosing Diffusion [87]	★★★★★	★★★★★	★★★★★	★★★★★	✓	(✓)	3.17	≤ 3.75
DDPM++ Continuous [206]	★★★★★	★★★★★	★★★★★	★★★★★	✓	(✓)	2.20	-
Flow Contrastive (EBM) [55]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	37.30	≈ 3.27
VAEBM [247]	★★★★★	★★★★★	★★★★★	★★★★★	✓	✗	12.19	-
Variational Autoencoders								
Convolutional VAE [123]	★★★★★	★★★★★	★★★★★	★★★★★	✓	(✓)	106.37	≤ 4.54
Variational Lossy AE [29]	★★★★★	★★★★★	★★★★★	★★★★★	✗	(✓)	-	≤ 2.95
VQ-VAE [184], [235]	★★★★★	★★★★★	★★★★★	★★★★★	✗	(✓)	-	≤ 4.67
VD-VAE [31]	★★★★★	★★★★★	★★★★★	★★★★★	✓	(✓)	-	≤ 2.87
Autoregressive Models								
PixelRNN [234]	★★★★★	★★★★★	★★★★★	★★★★★	✗	✓	-	3.00
Gated PixelCNN [233]	★★★★★	★★★★★	★★★★★	★★★★★	✗	✓	65.93	3.03
PixelIQN [173]	★★★★★	★★★★★	★★★★★	★★★★★	✗	✓	49.46	-
Sparse Trans. + DistAug [32], [110]	★★★★★	★★★★★	★★★★★	★★★★★	✗	✓	14.74	2.66
Normalizing Flows								
RealNVP [43]	★★★★★	★★★★★	★★★★★	★★★★★	✗	✓	-	3.49
GLOW [124]	★★★★★	★★★★★	★★★★★	★★★★★	✗	✓	45.99	3.35
FFJORD [62]	★★★★★	★★★★★	★★★★★	★★★★★	✓	(✓)	-	3.40
Residual Flow [26]	★★★★★	★★★★★	★★★★★	★★★★★	✓	(✓)	46.37	3.28

Method	Train Speed	Sample Speed	Num. Params.	Resolution Scaling	Free-form Jacobian	Exact Density	FID	NLL (in BPD)
Generative Adversarial Networks								
DCGAN [182]	*****	*****	*****	*****	✓	✗	37.11	-
ProGAN [114]	*****	*****	*****	*****	✓	✗	15.52	-
BigGAN [19]	*****	*****	*****	*****	✓	✗	14.73	-
StyleGAN2 + ADA [115]	*****	*****	*****	*****	✓	✗	2.42	-
Energy Based Models								
IGEBM [46]	*****	*****	*****	*****	✓	✗	37.9	-
Denoising Diffusion [87]	*****	*****	*****	*****	✓	(✓)	3.17	≤ 3.75
DDPM++ Continuous [206]	*****	*****	*****	*****	✓	(✓)	2.20	-
Flow Contrastive (EBM) [55]	*****	*****	*****	*****	✓	✗	37.30	≈ 3.27
VAEBM [247]	*****	*****	*****	*****	✓	✗	12.19	-
Variational Autoencoders								
Convolutional VAE [123]	*****	*****	*****	*****	✓	(✓)	106.37	≤ 4.54
Variational Lossy AE [29]	*****	*****	*****	*****	✗	(✓)	-	≤ 2.95
VQ-VAE [184], [235]	*****	*****	*****	*****	✗	(✓)	-	≤ 4.67
VD-VAE [31]	*****	*****	*****	*****	✓	(✓)	-	≤ 2.87
Autoregressive Models								
PixelRNN [234]	*****	*****	*****	*****	✗	✓	-	3.00
Gated PixelCNN [233]	*****	*****	*****	*****	✗	✓	65.93	3.03
PixelIQN [173]	*****	*****	*****	*****	✗	✓	49.46	-
Sparse Trans. + DistAug [32], [110]	*****	*****	*****	*****	✗	✓	14.74	2.66
Normalizing Flows								
RealNVP [43]	*****	*****	*****	*****	✗	✓	-	3.49
GLOW [124]	*****	*****	*****	*****	✗	✓	45.99	3.35
FFJORD [62]	*****	*****	*****	*****	✓	(✓)	-	3.40
Residual Flow [26]	*****	*****	*****	*****	✓	(✓)	46.37	3.28

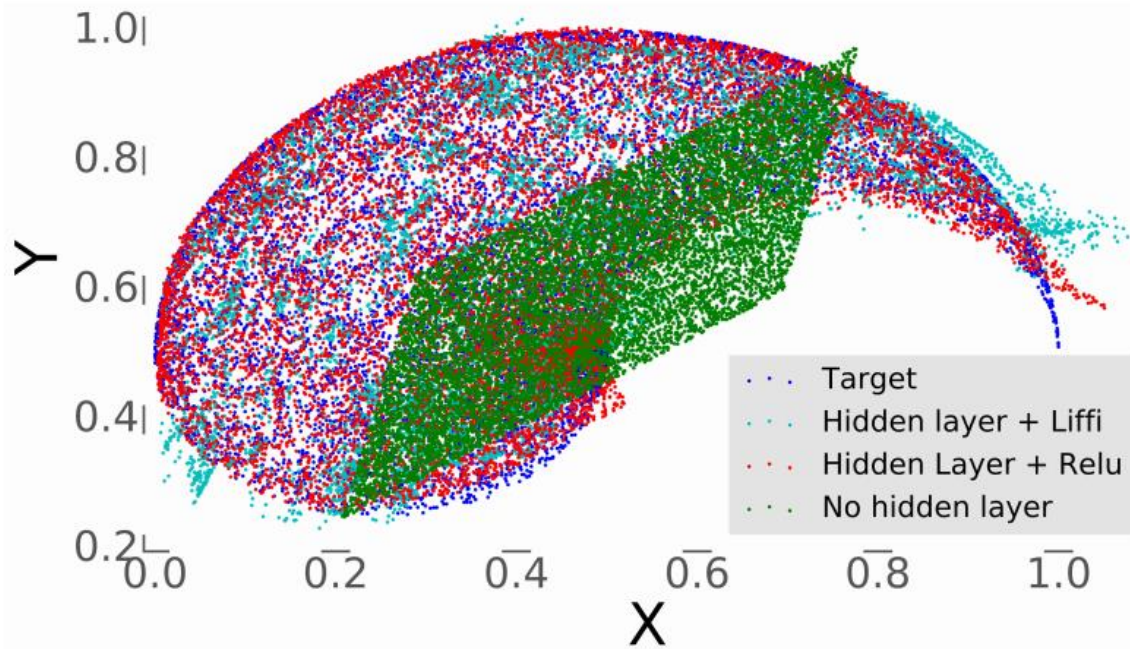


03

ENERGY BASED MODELS

ENERGY BASED MODELS

Um tipo de abordagem que teve origem na física estatística. A função utilizada no modelo - função energética - captura a probabilidade estatística do dado de entrada e gera novos dados extraindo amostras dos modelos aprendidos.



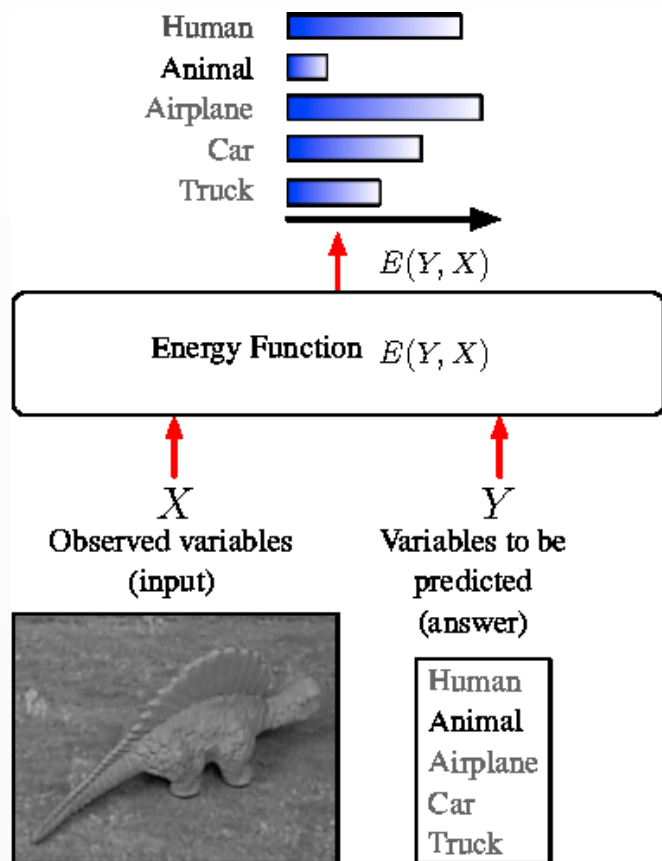
O aprendizado baseado em energia é uma estrutura unificada para todas essas abordagens probabilísticas e não probabilísticas. Especialmente para treinamento não probabilístico de modelos gráficos ou outros modelos estruturados.

O que são modelos baseados em energia no aprendizado de máquina?

Modelos baseados em energia (EBMs) descobrem dependências de dados aplicando uma medida de compatibilidade (energia escalar) a cada configuração das variáveis.

Um modelo mede a compatibilidade entre as variáveis X observadas e as variáveis para ser previsto Y usando uma função de energia $E(Y, X)$.

Por exemplo, X poderia ser os pixels de uma imagem, e Y uma etiqueta discreta descrevendo o objeto na imagem. Dado X , o modelo produz a resposta Y que minimiza a **energia E** .

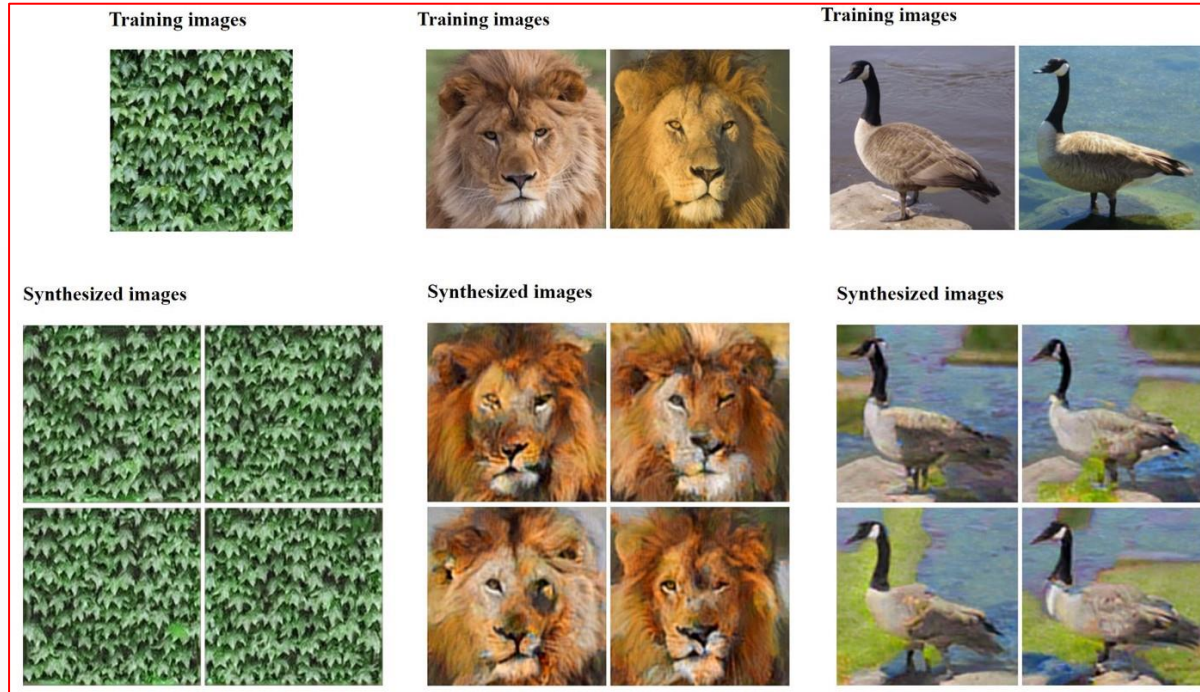


EXEMPLO

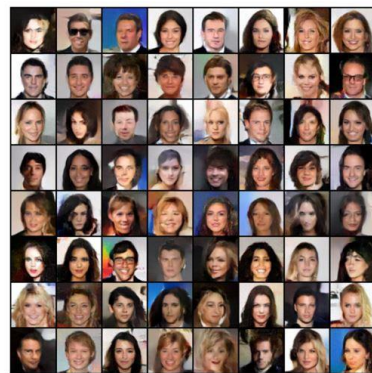
Rede Neural Generativa Baseada em Energia

Um bom exemplo do uso dos modelos baseados em energias é utilizá-lo juntamente com rede neural profunda e dinâmica de Langevin, no qual, juntos, proporcionam uma forma eficiente e poderosa de sintetizar dados de alta dimensão com alta qualidade.

Geração de imagens pela dinâmica Langevin a partir de modelos baseados em energia com ConvNets como funções energéticas potenciais em 2016 (pelo grupo Song-Chun Zhu's)



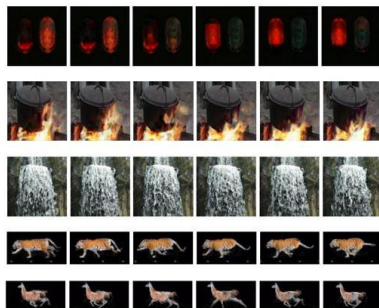
Geração de imagens, vídeos, formas volumétricas 3D e nuvens de pontos pela dinâmica Langevin das Redes Neurais Generativas baseadas em energia aprendidas



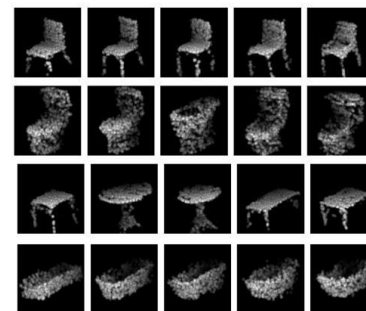
(a) synthesized images



(c) synthesized volumetric shapes



(b) synthesized videos



(d) synthesized point clouds



04

VARIATIONAL AUTOENCODERS – VAES

VAE - O QUE É?

Um VAE é um **autoencoder** cuja distribuição de codificações é regularizada durante o treinamento, a fim de garantir que seu espaço latente tenha boas propriedades, o que nos permite gerar novos dados.

VAMOS DAR ALGUNS PASSOS PARA TRÁS?

VAE

AUTOENCODERS

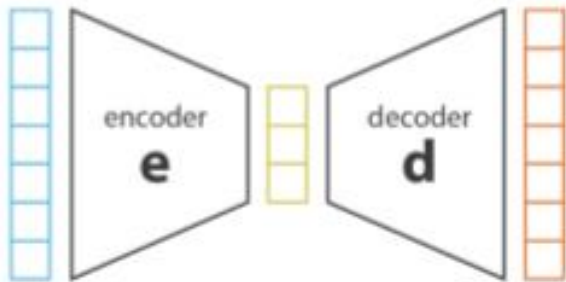
ENCODERS



ENCODERS x DECODERS

ENCODERS são estruturas criadas que permitem redução de dimensionalidade através de seleção ou extração de features explicativas.

DECODERS são estruturas que fazem o caminho inverso.



x

$e(x)$

$d(e(x))$

initial data
in space R^n

encoded data
in latent space R^m (with $m < n$)

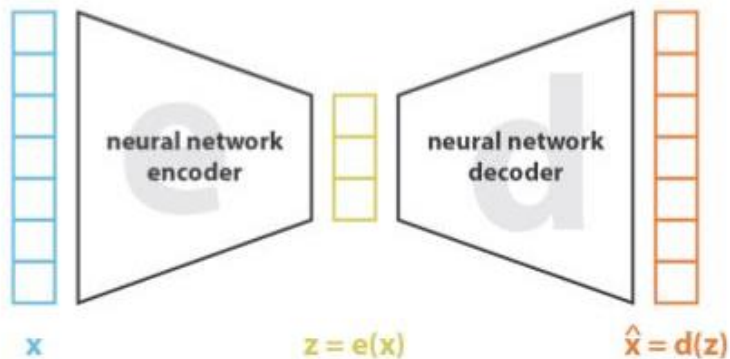
encoded-decoded data
back in the initial space R^n

$x = d(e(x)) \rightarrow$ **lossless encoding**
no information is lost
when reducing the
number of dimensions

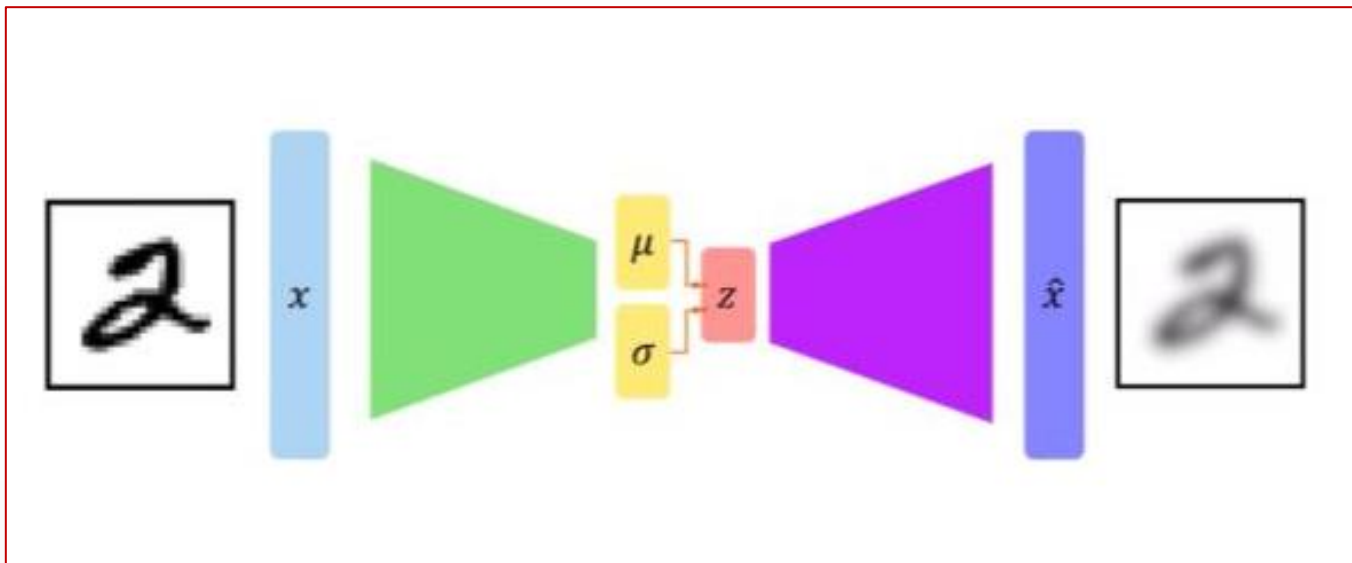
$x \neq d(e(x)) \rightarrow$ **lossy encoding**
some information is lost
when reducing the
number of dimensions and
can't be recovered later

AUTOENCODERS

São redes neurais posicionadas como encoders e decoders que retropropaga a diferença entre x e $d(e(x))$ de forma a se auto otimizar.



$$\text{loss} = ||x - \hat{x}||^2 = ||x - d(z)||^2 = ||x - d(e(x))||^2$$

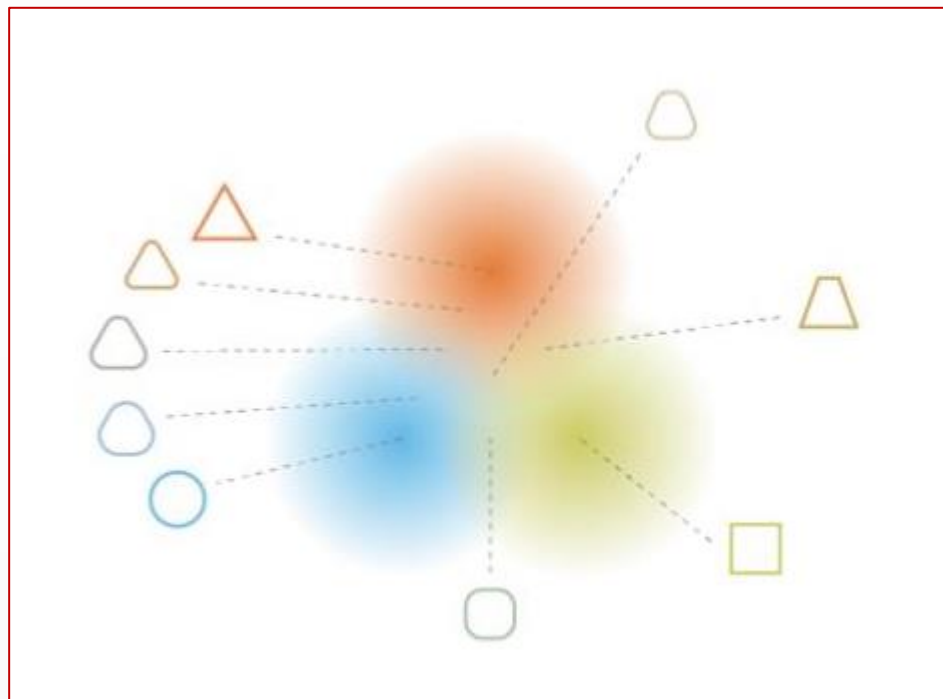


Extraído de <https://www.deeplearningbook.com.br/variational-autoencoders-vaes-definicao-reducao-de-dimensionalidade-espaco-latente-e-regularizacao/> em 28 de Junho de 2022

VAEs

VAEs são autoencoders cujo treino é regularizado para evitar o overfitting e garantir que seu espaço latente tenha boas propriedades generativas.







05

AUTOREGRESSIVE MODELS

CARACTERÍSTICAS

Generativa: Este é um tipo de aprendizado não supervisionado onde o objetivo é **aprender** sobre um conjunto de dados de treinamento; e, em seguida, gerar novas amostras.

CARACTERÍSTICAS

Probabilística: Eles se concentram na geração de imagens modelando explicitamente uma distribuição de probabilidade e, em seguida, gerando amostras a partir dela.

CARACTERÍSTICAS

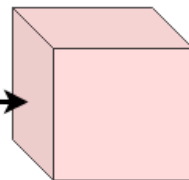
Autoregressiva: O modelo aprende a distribuição de probabilidade dos dados de treinamento explicitamente.

Training

Training Image



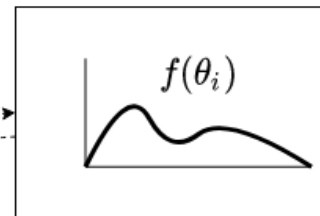
$\in [0, 255]$



Logits $\in [-\infty, \infty]$

Loss θ_i

Optimize parameters of probability distribution



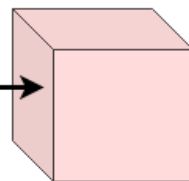
Probability distribution for RGB pixel values

Sampling

Generated Image



$\in [0, 255]$



θ_i

Sampling Function

Repeat 32x32 times, each time update only a single pixel in the image

REDES UTILIZADAS

Comum a utilização de:

- Redes Neurais Recorrentes;
- Redes Neurais Convolucionais.

REDES UTILIZADAS

Comum a utilização de:

- Redes Neurais Recorrentes;
- Redes Neurais Convolucionais.



06

GENERATIVE ADVERSARIAL NETWORKS

Tradução Literal: REDES ADVERSARIAIS GENERATIVAS

Generative Adversarial Networks - GANs

Possui 2 partes:

GERADOR

Aprendem a gerar amostras.

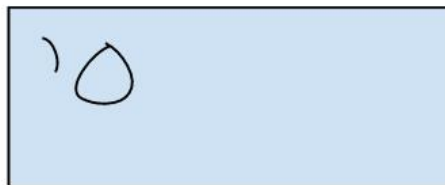
DISCRIMINADOR

Aprende a avaliar se uma amostra é fruto da base real ou do gerador.

Generated Data

Discriminator

Real Data

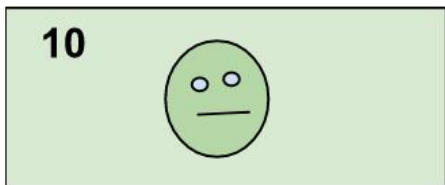


FAKE

REAL



As training progresses, the generator gets closer to producing output that can fool the discriminator:



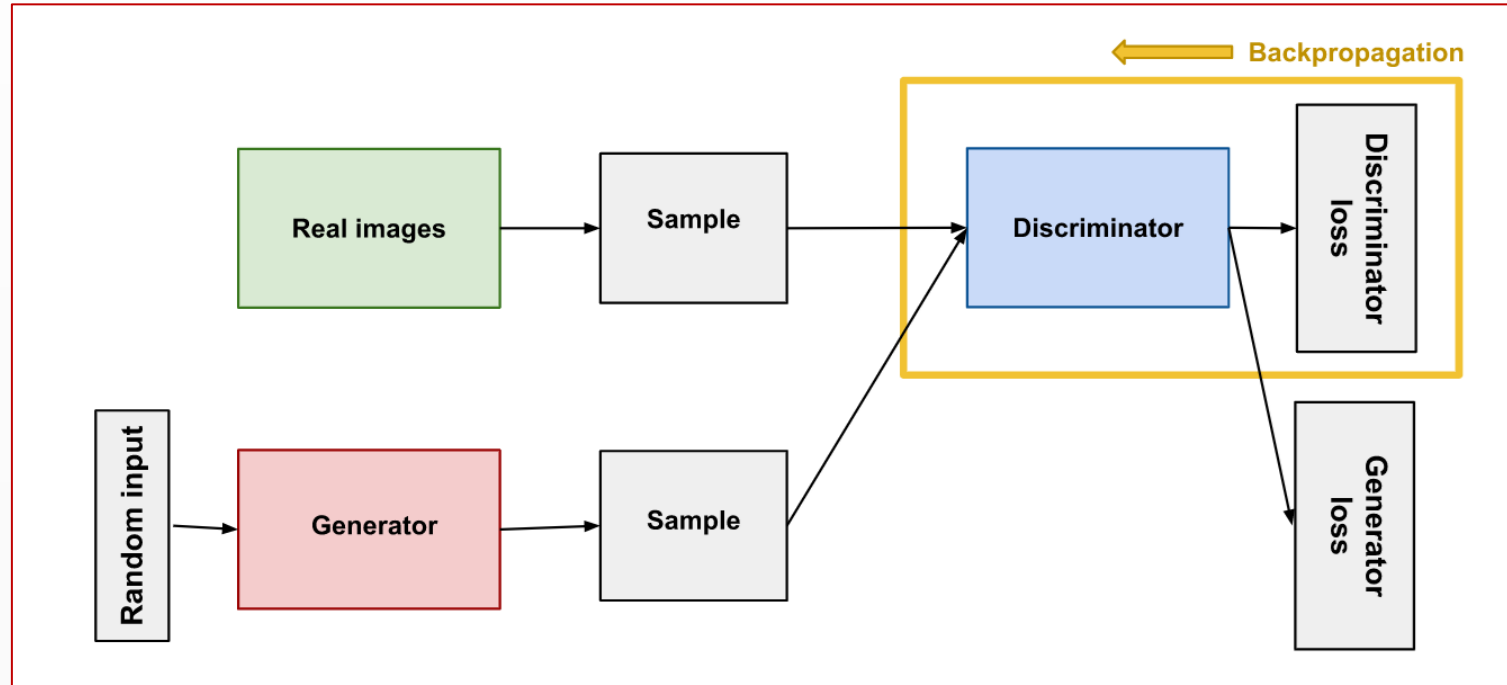
FAKE

REAL



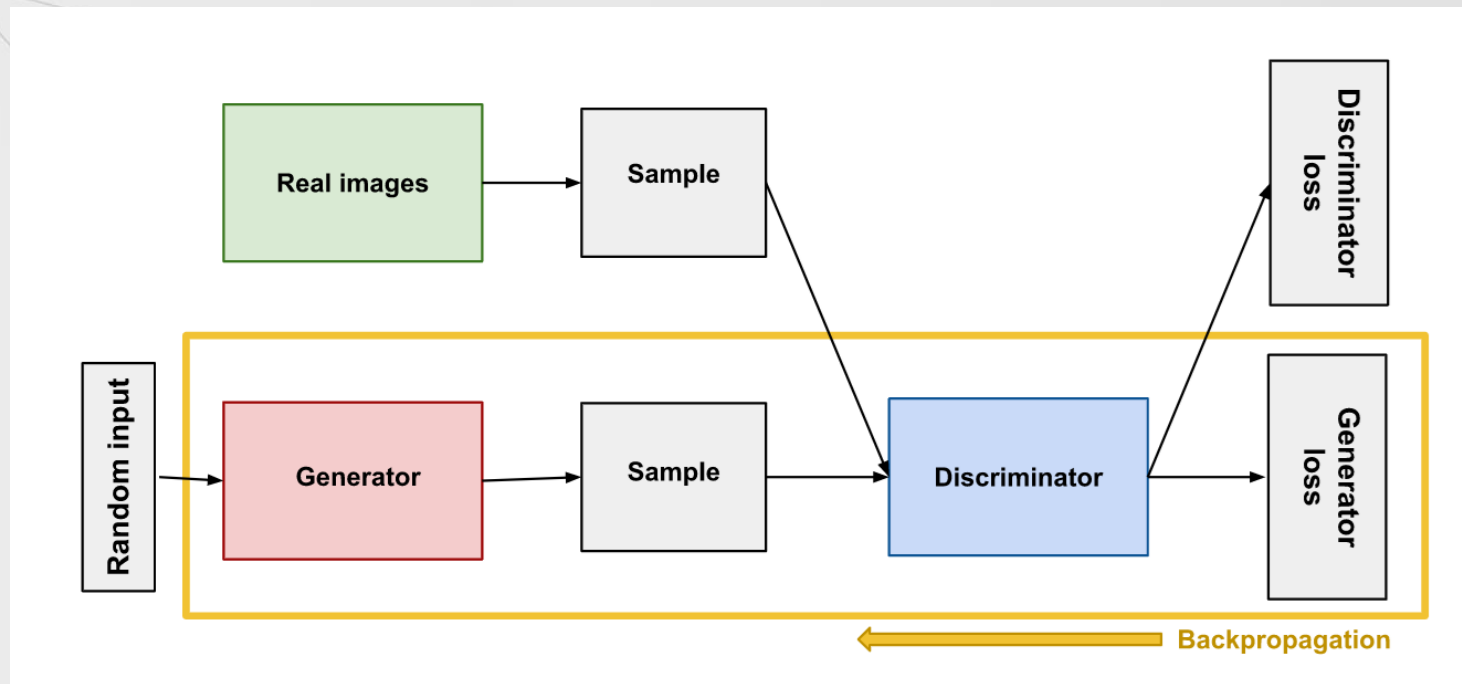
Extraído de https://developers.google.com/machine-learning/gan/gan_structure em 28 de novembro de 2022.

ESTRUTURA DE UMA GAN - DISCRIMINADOR



Extraído de https://developers.google.com/machine-learning/gan/gan_structure em 28 de Junho de 2022.

ESTRUTURA DE UMA GAN - GERADOR



PONTOS DE FALHA DAS GANs

Vanishing Gradients (Desaparecimento de gradientes): Se o discriminador for muito bom, o treinamento do gerador pode falhar por conta do desaparecimento de gradientes.

Collapse mode: Situação onde o gerador possui imagens muito similares pois está “viciado”.

POSSÍVEIS CONTORNOS

- Outras funções de perda;
- Data augmentation;
- GANs without competition.

Abstract

Training generative adversarial networks (GAN) using too little data typically leads to discriminator overfitting, causing training to diverge. We propose an adaptive discriminator augmentation mechanism that significantly stabilizes training in limited data regimes. The approach does not require changes to loss functions or network architectures, and is applicable both when training from scratch and when fine-tuning an existing GAN on another dataset. We demonstrate, on several datasets, that good results are now possible using only a few thousand training images, often matching StyleGAN2 results with an order of magnitude fewer images. We expect this to open up new application domains for GANs. We also find that the widely used CIFAR-10 is, in fact, a limited data benchmark, and improve the record FID from 5.59 to 2.42.



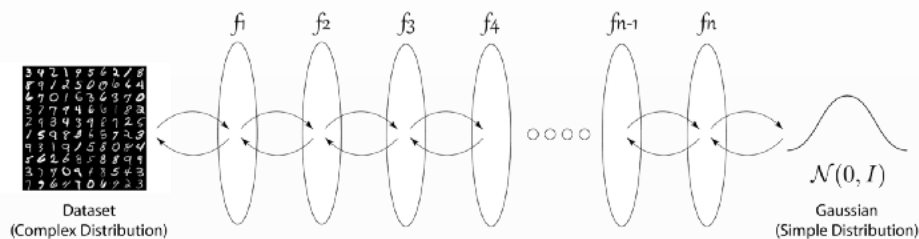
07

NORMALIZING FLOWS

O QUE É E COMO FUNCIONA

A normalização dos fluxos é um método de construção de distribuições complexas que transforma uma densidade de probabilidade através de uma série de mapeamentos invertíveis.

Essa técnica aplica repetidamente a regra de mudança de variáveis, a densidade inicial flui através da sequência de mapeamentos inversíveis. Ao final desta sequência, obtemos uma distribuição de probabilidade válida e, portanto, este tipo de fluxo é referido como um fluxo normalizante.



Nesta imagem, podemos ver que os fluxos normalizadores transformam um ponto de dados complexo - como uma imagem - para uma simples Distribuição Gaussiana ou vice-versa. Amostras aleatórias são retiradas dessa distribuição gaussiana para obter resultados do modelo durante os testes.



08

CONCLUSÕES

CONCLUSÕES

Apesar das GANs liderarem em termos de qualidade durante muito tempo, outras abordagens estão trazendo resultados satisfatórios com o passar do tempo.

Trabalhos recentes mostram que modelos híbridos oferecem um bom equilíbrio entre as vantagens e desvantagens de cada modelo isolado.

Outro ponto, é que qualquer melhoria em uma técnica específica pode significar melhorias em todos os modelos. Por exemplo, a aplicação de estratégias inovadoras de aumento de dados tem sido considerada muito benéfica em vários modelos de classes.

Concluindo, a chave para o futuro da generalização é usar modelos unificados, capazes de modelar de forma contínua dados irregulares, de comprimento arbitrários e de diferentes escalas.

09

REFERÊNCIAS



REFERÊNCIAS

- [1] LECUN, Yann; CHOPRA, Sumit; HADSELL, Raia; RANZATO, Marc'Aurelio; HUANG, Fu Jie. A tutorial on energy-based learning. MIT Press, 2006. Disponível em: <http://yann.lecun.com/>. Acesso em: 29 jun. 2022.
- [2] XIE, Jianwen; LU, Yang; ZHU, Song-Chun; WU, Ying Nian. A Theory of Generative ConvNet. (ICML 2016).
- [3] XIE, Jianwen; LU, Yang; ZHU, Song-Chun; WU, Ying Nian. Synthesizing Dynamic Pattern by Spatial-Temporal Generative ConvNet. (CVPR 2017).



The background of the slide features a complex, abstract geometric pattern. It consists of numerous thin, light gray lines that connect various points, creating a network-like structure. Some of these points are highlighted as larger, solid dark gray circles, while others are smaller dots. The overall effect is a modern, minimalist design that suggests connectivity and structure. The word "OBRIGADO" is centered in the middle of the slide, rendered in a large, bold, dark gray sans-serif font.

OBRIGADO

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.