REDES GENERATIVAS ADVERSÁRIAS COM CONVOLUÇÃO PROFUNDA - CONSTRUÇÃO E ANÁLISE

Rodrigo O. Schilling[1]

[1]PONTIFICIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (PUC-MG)

[2]Departamento de Ciência da Computação
Belo Horizonte, Brasil.

Abstract. *This article aims undestand, present, structure, and analyze the creation and development of the architecture of Deep Convolutional Generative Adversarial Networks (DCGANs). Generative Adversarial Networks (GANs) were initially introduced in 2014 by Ian Goodfellow and colleagues. In 2016, Radford et al. proposed incorporation of Convolutional Neural Networks (CNNs) to improve the training stability and performance of these models.*

Resumo. *Este artigo busca compreender, apresentar, estruturar e analisar a criação e o desenvolvimento da arquitutura das Redes Generativas Adversárias com Covolução Profunda (Deep Convolutional Gernerative Adversarial Networks - DCGANs). As Redes Generativas Adversárias (GANs) foram introduzidas inicialmente em 2014 por Ian Goodfellow e colaboradores. Em 2016, Radford e demais propuseram a utilização de Redes Neurais Convulocionais (CNNs) para aprimorar a estabilidade e o treinamento das GANs.*

1. INTRODUÇÃO

O aprendizado de atributos em grandes conjuntos de dados constitui uma das principais áreas de pesquisa no contexto da pesquisa em visão computacional[Radford et al. 2016]. A capacidade de aprendizado por meio de arquiteturas bem estruturadas permite a geração de imagens em alta definição. Um dos principais métodos para a geração de imagens é a aplicação das Redes Adversárias Generativa (GANs, na sigla em inglês), proposta em 2014 por Goodfellow. Esta abordagem consiste na interação entre dois componentes: um gerador (G) responsável por capturar a distribuição dos dados e gerar novas imagens, e um discriminador (D), encarregado de estimar a probabilidade de uma imagem ser real ou falsa. [Goodfellow et al. 2014]

O processo de treinamento é formulado como um jogo de soma zero, em que o gerador busca enganar o discriminador, enquanto este tenta distinguir corretamente entre imagens reais e criadas. O discriminador acaba por ser uma função de perda adaptativa, que pode ser descartada quando o gerador está treinado [T. Karras 2018]. Com a solução se aproximando de um equilíbrio de Nash [Curtó et al. 2020]

As Redes Adversárias Generativas com Convolução Profunda (DCGANs, na sigla em inglês) são uma extensão das GANs, baseadas nas Redes Neurais Convolucionais (CNNs, na sigla em inglês), e utilizam o aprendizado não supervisionado. Esta abordagem permite à rede aprender padrões diretamente dos dados, substituindo

funções de pooling por operações convolucionais que preservam mais informações [Radford et al. 2016]. Esta metodologia foi inicialmente proposta por Radford em 2016, e seus resultados demonstram maior estabilidade durante o treinamento e melhoria na média do pooling [Radford et al. 2016].

A geração de imagens constitui um dos principais desafios no campo da visão computacional [Curtó et al. 2020]. A utilização de GANs tem se mostrado promissora nesse contexto, ao viabilizar a criação de imagens sintéticas que podem ser utilizadas para o treinamento de novos modelos. Essa abordagem tem encontrado aplicações relevantes, especialmente na área médica, abrangendo sete categorias distintas: síntese, segmentação, reconstrução, detecção, redução de ruído, registro e classificação. [Kazeminia et al. 2020]

Neste trabalho, foi utilizado o conjunto de dados CelabA-HQ, apresentado inicialmente no artigo desenvolvido pela NVIDIA, intitulado "Progressive Growing of GANs for improved quality, stability, and Variation". [T. Karras 2018] Esse dataset é uma derivação de alta qualidade do CELEBA, introduzido por [Liu et al. 2015] com o objetivo de realizar reconhecimento facial e previsão de atributos. [Liu et al. 2015]

Para o presente artigo, utilizou-se uma versão redimensionado do CelebA-HQ (256x256 pixels)[1], disponibilizada na plataforma Kaggle, a qual viabilizou um treinamento mais eficiente considerando os recursos computacionais disponíveis: processador i5-11400H, placa gráfica NVIDIA GeForce GTX com 4GB de VRAM e 8GB de memória RAM. Sendo empregada a arquitetura das Redes Adversárias Generativas com Convolução Profunda.

2. SOBRE A REDE TREINADA

Conforme descrito anteriormente, as imagens utilizadas para treinamento fazem parte do conjunto de dados CelebA-HQ (256x256 pixels) o tamanho das imagens foi redimensionado para 64x64, as imagens podem ser avalinadas na Figura 1, para diminuir a GPU necessária para treinamento e tornar os cálculos mais rápidos, visando compensar as características da máquina utilizada no treinamento.

A estrutura do gerador (G) foi desenvolvida com cinco camadas de convolução transposta, responsáveis pela realização do upsampling progressivo do vetor latente até a geração da imagem falsa. A entrada do gerador consiste em vetores amostrados de uma distribuição normal, os quais são transformados em uma imagem final de 64x64 pixels[Radford et al. 2016]. A camada inicial utiliza normalização por lote e função de ativação ReLU para estabilizar o treinamento. Os três blocos seguinte expandem gradualmente a imagem até os 32x32 pixels, mantendo a estrutura de normalização e ativação. Por fim, é gerada a imagem com 64x64 pixel e aplicada a função de ativação tangente hiperbólica, que normaliza os valores em um intervalo [-1,1], compatível com a entrada do discriminador [Radford et al. 2016].

[1]Link para o Dataset: `https://www.kaggle.com/datasets/badasstechie/celebahq-resized-256x256`

O discriminador foi estruturado para processar imagens com dimensões de 64x64 pixels, em formato RGB. A rede é composta por cinco camadas convolucionais responsáveis por reduzir às dimensões da imagem, aumentando a profundidade. O objetivo do discriminador é atribuir a probabilidade de a imagem ser real ou não, assumindo valor próximo de 1 para imagens reais e próximo de 0 para imagens falsas. A camada inicial de convolucional reduz a imagem de 64x64 pixel para 32x32 pixel e aplica a função de ativação LeakyRelu com coeficiente de inclinação de 0,2. Em seguida, os três blocos adicionais realizam novas convoluções, reduzindo o tamanho da imagem até 4x4 pixels e a mesma função de ativação em cada etapa [Liu et al. 2022]. Por fim, a camada final gera um único valor por imagem, ao qual é aplicado a função de ativação Sigmoid, produzindo uma saída no intervalo [0,1], correspondente a probabilidade de a imagem ser considerada real ou falsa [Radford et al. 2016].

Foram construídos otmizadores específicos para cada uma das redes, utilizando o mesmo valor de $\beta = 0.5$ para o termo de momento, mas com taxas de aprendizados diferentes para o gerador e o discriminador. Os valores completos dos hiperparâmentros utilizados, serão apresentados no tópico seguinte, juntamente com a análise dos resultados obtidos dos três teste realizados.

Figura 1. IMAGENS UTILIZADAS NO TREINAMENTO

3. RESULTADOS OBTIDOS

No primeiro treinamento da DCGANs foi adotada uma taxa de aprendizado de 0,0001 para o discriminador e de 0,0003 para o gerador. A rede foi treinada por um total de 150 épocas. Observou-se nesse treinamento inicial, uma convergência consistente das funções de perda do discriminador e do gerador, conforme ilustrado na Figura 2.

Ao final do primeiro treinamento, a função de perda do discriminador apresentou valor de 0,8389 indicando que o discriminador ainda conseguia diferenciar razoavelmente bem às imagens reais daquelas geradas artificialmente. A função de perda do gerador foi

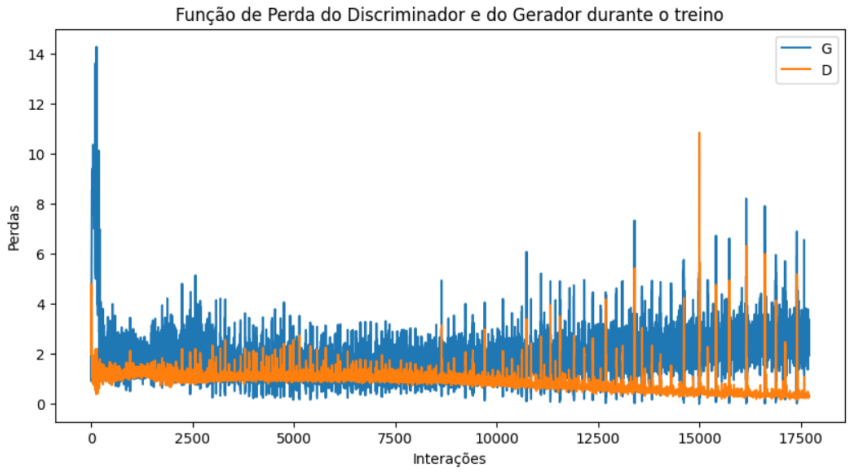

Figura 2. FUNÇÃO DE PERDA - 1º TREINAMENTO

de 0,68 sugerindo uma melhoria da qualidade das imagens geradas. A probabilidade média atribuída a imagens reais na última época foi de 0.4744, indicando uma incerteza por parte do discriminador em diferenciar as imagens reais e falsas. No entanto, é importante destacar que na época imediatamente anterior a última, o discriminador apresentava uma função de perda significativamente menor, 0,3527, e o Gerador uma perda consideravelmente maior, 2,3744, podendo indicar flutuações no treinamento e sugere que o resultado observado na última época não é suficiente para concluir que houve convergência estável da rede.

Com o objetivo de avaliar às imagens geradas do modelo foram construídas cerca de trezentas imagens para se ter uma noção das imagens construídas, estas imagens podem ser observadas na Figura 3.

Figura 3. IMAGENS GERADAS NO 1º TREINAMENTO

Observa-se que as imagens geradas pelo modelo apresentam semelhanças com o rosto humano, embora ainda se situem dentro do chamado de vale da estranheza, caracterizado pela aparência artificial das imagens, distante do aspecto natural esperado. A baixa qualidade das imagens geradas, somadas aos resultados inconclusivos observados no primeiro treinamento da DCGANs, evidenciou a necessidade de conduzir o segundo treinamento.

Nesta etapa buscou-se melhorar a capacidade do gerador de criar imagens realistas, para isso o modelo foi treinado por mais 50 épocas. Ao finalizar o treinamento destas épocas notou-se que o Discriminador teve facilidade de diferenciar as imagens geradas pelo gerador, apesar das oscilações do treinamento, como é possível observar na Figura 4.

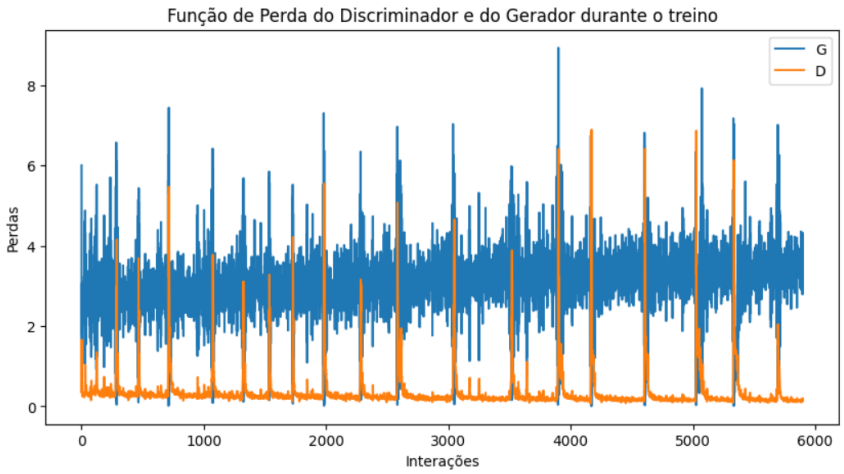

Figura 4. FUNÇÃO DE PERDA - 2º TREINAMENTO

Em termos de qualidade visual, conforme observado na Figura 5, nota-se um pequeno ganho marginal em relação ao treinamento anterior, com algumas poucas imagens apresentando semelhança com uma imagem real. Entretanto, a maioria da amostra geradas apresentam características inconsistentes, como a fusão de rostos femininos e masculinos, além de áreas de desfoque ao fundo.

Figura 5. IMAGENS GERADAS NO 2º TREINAMENTO

Os aspectos das imagens do segundo treinamento, apresentadas na Figura 5, sugerem que o gerador ainda não é capaz de produzir imagens com realismo e consistência esperados, indicando uma necessidade de mais épocas de treinamento e possíveis ajustes nos hiperparâmentros da rede. Diante do exposto, foi conduzido um terceiro treinamento, com 200 épocas, mantendo a estrutura da rede utilizada desde o primeiro treinamento. O diferencial deste novo experimento, foi a redução das taxas de aprendizado do discriminador e do gerador, para 0,00005 e 0,002, respectivamente.

A alteração proposta teve como objetivo promover um aprendizado mais estável e refinado de ambas às redes. No entanto, conforme ilustrado no Figura 6, as funções de

perda do discriminado e do gerado se mantiveram estáveis ao longo das interações, sem evidenciar uma tendência de convergência. Este fato sugere que o gerador não apresentou melhor significativa na geração de imagens, e de maneira similar o discriminador não apresentou dificuldade em distinguir as imagens geradas das reais.

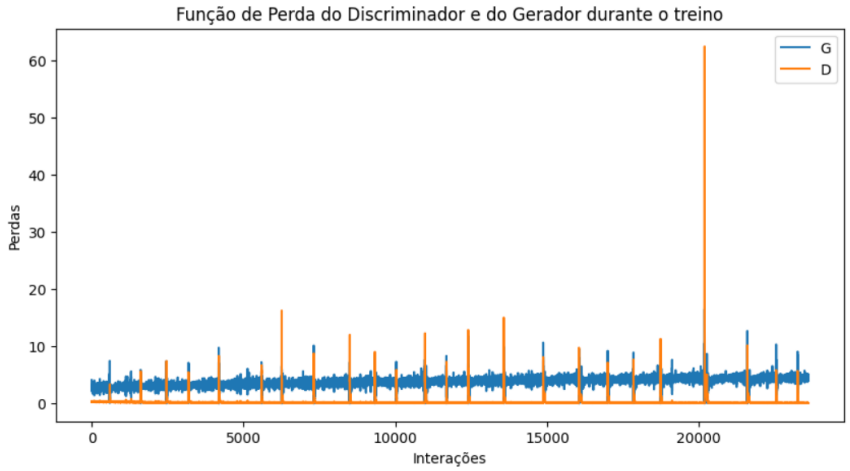

Figura 6. FUNÇÃO DE PERDA - 3º TREINAMENTO

As imagens geradas a partir do modelo resultante do terceiro treinamento também não apresentaram melhorias significativas em comparação as geradas a partir do primeiro e do segundo treinamentos, conforme demonstrado na Figura 7. Embora seja possível distinguir se o rosto gerado possui características femininas ou masculinas, a qualidade visual ainda é baixa e sem consistência estrutural. Esses resultados indicam que o modelo ainda apresenta limitações para a geração de rostos que se assemelhem aos utilizados para o treinamento.

Figura 7. IMAGENS GERADAS NO 3º TREINAMENTO

4. CONCLUSÃO

As Rede Generativa Adversária com Convolução Profunda (Deep Convolutional GANs), tem se apresentado como Uma solução robusta para a geração de imagens e textos [Heusel et al. 2018]. A DCGANs apresentada neste estudo, demonstrou certo grau de confiabilidade, mesmo com treinamento relativamente curto, 400 épocas no total, no aprendizado e na geração de rostos. No presente trabalho, os resultados obtidos nas primeiras 150 épocas de treinamento já evidenciaram um desempenho satisfatório, com o gerador se adaptando de forma eficiente às interações iniciais.

O segundo treinamento, que adicionou mais 50 épocas ao treinamento inicial, apresentou ganhos marginais na qualidade das imagens geradas. Ainda assim os resultados

das funções de perda se mantiveram estáveis durante todo o processo, indicando um comportamento controlado durante o treinamento.

No terceiro treinamento, foram adicionadas 200 épocas e alteradas as taxas de aprendizado do gerador e do discriminador. Contudo, os avanços obtidos na qualidade visual das imagens geradas foram limitados, sugerindo que aumentar o número de épocas, possa ser mais importante que alterações pontuais nos hiperparâmentros.

Este estudo buscou apresentar as etapas do treinamento de uma DCGANs e os ganhos obtidos a partir da geração e análise qualitativa das imagens e das funções de perda. No entanto, não foram realizados experimentos adicionais para avaliar a ocorrêcia de colapso do modelo [Salimans et al. 2016]. É importante ressaltar que métricas quantitativas avançadas, como o Fréchet Inception Distance (FID), que busca capturar a similaridade entre a imagem gerada e a real [Heusel et al. 2018], e o Inception Score (IS),[Salimans et al. 2016] foram consideredas para uma avaliação mais aprofundada do modelo. No entanto devido a restrições de tempo e recursos computacionais, a aplicação e análise dessas métricas ficaram de fora deste trabalho.

DISPONIBILIDADE DOS DADOS

O código utilizado para o treinamento e analise da rede estudada neste trabalho encontra-se disponiveis no repositório: `https://github.com/RoSchilling/DCGANs-Construcao-e-Analise`

References

Curtó, J. D., Zarza, I. C., Torre, F., King, I., and Lyu, M. R. (2020). *HIGH-RESOLUTION DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS*. arXiv.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). *Generative Adversarial Nets*. Departamento de pesquisa operacional.

Heusel, M., Ramsauer, J., Interthiner, T., and Nessler, B. (2018). *GANs Treined by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. arXiv.

Kazeminia, S., Baur, C., A. Kuijper, and, B. v. G., Navab, N., Albarqouni, S., and Mukhopadhyay, A. (2020). Gans for medical image analysis. In *Artificial Intelligence in Medicine*. Elsevier.

Liu, B., Lv, J., Fan, X., Luo, J., and Zou, T. (2022). Applicantion of an improved dcgan for image generation. In *Mobile Information Systems*. Hindawi.

Liu, Z., Lou, P., Wang, X., and Tang, X. (2015). *DEEP LEANING FACE ATTRIBUTES IN WILD*. arXiv.

Radford, A., Metz, L., and Chintala, S. (2016). *Unsupervised Representation LEARNING WITH DEEP CONVOLUTIONAL GENERATIVA ADVERSARIAL NETWORKS*. ICLR.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). *Improved Techniques for Training GANs*. arXiv.

T. Karras, T. Aila, S. L. J. L. (2018). *Progressive Growing of Gans for Improved Quality, Stability and Variation*. ICLR.