

REVISTA
DESAFIOS

ISSN: 2359-3652

V.13,n.3, maio/2026–DOI:10.20873/vol13n3pibic202517

**INTELIGÊNCIA ARTIFICIAL NA GERAÇÃO DE IMAGENS REALISTAS:
TÉCNICAS E APLICAÇÕES**

*ARTIFICIAL INTELLIGENCE IN GENERATING REALISTIC IMAGES:
TECHNIQUES AND APPLICATIONS*

*INTELIGENCIA ARTIFICIAL EN LA GENERACIÓN DE IMÁGENES REALISTAS:
TÉCNICAS Y APLICACIONES*

Kalil Garcia Canuto

E-mail: kalil.garcia@mail.uft.edu.br

Warley Gramacho

E-mail: wgramacho@mail.uft.edu.br

ABSTRACT:

This project investigates the application of Deep Convolutional Generative Adversarial Networks (DCGANs) for the generation of realistic synthetic images. Utilizing the 102 Flower Category Dataset, a model was implemented and trained with the objective of generating 128x128 pixel color images of flowers. The model architecture was based on the principles proposed by Radford, Metz, and Chintala (2016), featuring a Generator that uses transposed convolutions and a Discriminator that employs strided convolutions. The training was stabilized through Batch Normalization and ReLU/LeakyReLU activation functions. The quality of the generated samples was evaluated both visually and quantitatively using the Fréchet Inception Distance (FID) metric. The results demonstrate the architecture's ability to learn the distribution of a complex dataset, achieving an optimal FID score of 808.74 at epoch 600. This result serves as a quantitative baseline and shows the difficulties of using this architecture to generate highly realistic images with this dataset.

KEYWORDS: DCGAN; deep learning; artificial intelligence

RESUMO:

Este projeto investiga a aplicação de Redes Gerativas Adversariais Convolucionais Profundas (DCGANs) para a geração de imagens sintéticas realistas. Utilizando o 102 Flower Category Dataset, foi implementado e treinado um modelo com o objetivo de gerar imagens coloridas de flores com resolução de 128x128 pixels. A arquitetura do modelo foi baseada nos princípios propostos por Radford, Metz e Chintala (2016), com um Gerador que utiliza convoluções transpostas e um Discriminador que emprega convoluções com passo. O treinamento foi estabilizado por meio de Normalização em Lote e funções de ativação ReLU/LeakyReLU. A qualidade das amostras geradas foi avaliada visualmente e, quantitativamente, pela métrica Fréchet Inception Distance (FID). Os resultados demonstram a capacidade da arquitetura de aprender a distribuição de um dataset complexo, atingindo um score FID de 808.74 na época 600. Este resultado serve como um ponto de partida quantitativo e mostra as dificuldades de usar esta arquitetura para gerar imagens muito realistas com este dataset.

PALAVRAS-CHAVE: DCGAN; aprendizado profundo; inteligência artificial.

RESUMEN:

Este proyecto investiga la aplicación de Redes Generativas Adversarias Convolucionales Profundas (DCGANs) para la generación de imágenes sintéticas realistas. Utilizando el 102 Flower Category Dataset, se implementó y entrenó un modelo con el objetivo de generar imágenes en color de flores con una resolución de 128x128 píxeles.

La arquitectura del modelo se basó en los principios propuestos por Radford, Metz y Chintala (2016), con un Generador que utiliza convoluciones transpuestas y un Discriminador que emplea convoluciones con paso (strided convolutions). El entrenamiento se estabilizó mediante Normalización por Lotes (Batch Normalization) y funciones de activación ReLU/LeakyReLU.

La calidad de las muestras generadas se evaluó visualmente y, cuantitativamente, mediante la métrica Fréchet Inception Distance (FID). Los resultados demuestran la capacidad de la arquitectura para aprender la distribución de un conjunto de datos complejo, alcanzando una

puntuación FID de 808.74 en la época 600. Este resultado sirve como un punto de partida cuantitativo y evidencia las dificultades de utilizar esta arquitectura para generar imágenes altamente realistas con este dataset en particular.

PALABRAS CLAVE: DCGAN; aprendizaje profundo; inteligencia artificial.

INTRODUÇÃO

O campo da inteligência artificial tem apresentado uma evolução notável, transitando de modelos com capacidade de análise para sistemas capazes de gerar conteúdo novo e sintético. Um marco nesse avanço são as Redes Gerativas Adversariais (GANs), introduzidas por Goodfellow et al. (2014), que se destacam pela sua arquitetura baseada em uma competição adversarial entre dois modelos neurais: um Gerador, responsável por criar amostras a partir de um vetor de ruído, e um Discriminador, treinado para distinguir as amostras sintéticas das reais.

Contudo, o treinamento de GANs convencionais é frequentemente marcado por instabilidades, como o colapso de modo (mode collapse) e a dificuldade em convergir, resultando em imagens de baixa fidelidade (RADFORD; METZ; CHINTALA, 2016). Para superar tais limitações, a arquitetura Deep Convolutional Generative Adversarial Network (DCGAN) foi proposta, incorporando camadas convolucionais que preservam as relações espaciais dos dados. A DCGAN introduz um conjunto de diretrizes arquitetônicas, como a substituição de camadas de pooling por convoluções com passo, o uso de Normalização em Lote (Batch Normalization) para estabilizar o aprendizado (IOFFE; SZEGEDY, 2015), e a aplicação de funções de ativação específicas.

Uma das aplicações mais relevantes da geração de imagens realistas é o aumento de dados (data augmentation), uma técnica utilizada para ampliar artificialmente a variedade de um conjunto de treinamento. Em domínios onde a obtenção de dados é custosa ou limitada, a geração de amostras sintéticas de alta qualidade pode ser fundamental para melhorar o desempenho e a capacidade de generalização de modelos de aprendizado de máquina. A obtenção de bases de dados extensas e diversificadas, no entanto, ainda representa um obstáculo comum para pesquisadores e desenvolvedores (WAN; WANG; CHUNG; WANG, 2023).

Neste contexto, o presente trabalho se propõe a implementar e treinar um modelo de Deep Convolutional Generative Adversarial Network (DCGAN) na tarefa de geração de imagens realistas de flores, utilizando o 102 Flower Category Dataset. O objetivo é validar a capacidade da arquitetura em aprender as características de um dataset complexo e diverso, e analisar a qualidade e a diversidade das imagens geradas. Como objetivos específicos temos:

- Implementar as arquiteturas do Gerador e do Discriminador seguindo as diretrizes propostas no artigo de referência da DCGAN.
- Realizar o pré-processamento do 102 Flower Category Dataset para o treinamento do modelo.
- Conduzir o treinamento do modelo, monitorando as funções de perda para avaliar a estabilidade e a convergência do processo adversarial.
- Avaliar a qualidade das imagens geradas visualmente e quantitativamente, utilizando a métrica Fréchet Inception Distance (FID).

METODOLOGIA

COLETA DE DADOS:

O estudo utilizou o 102 Flower Category Dataset, composto por 8.189 imagens coloridas. Esse conjunto contém imagens de 102 espécies de flores, com cada classe apresentando entre 48 e 258 imagens. O dataset foi dividido em dois subconjuntos:

Conjunto de treinamento: 80% das imagens (6.551 imagens)

Conjunto de validação: 20% das imagens (1.638 imagens)

Essa divisão é uma prática comum que permite ajustar os hiperparâmetros e monitorar a generalização sem que o conjunto de validação interfira no aprendizado do modelo (GOODFELLOW; BENGIO; COURVILLE, 2016).

PRÉ-PROCESSAMENTO DAS IMAGENS:

Seguindo o artigo original, nenhum pré-processamento foi aplicado às imagens originais exceto pela normalização dos valores dos pixels de cada imagem para o intervalo $[-1, 1]$, para adequar os dados à função de ativação tanh da camada final do Gerador.

ARQUITETURA DO MODELO:

O modelo foi implementado utilizando a biblioteca TensorFlow. As arquiteturas do Gerador e do Discriminador foram projetadas para serem simétricas e totalmente convolucionais, conforme a filosofia da DCGAN descrita por Radford, Metz e Chintala (2016).

- **Gerador:** A arquitetura recebe como entrada um vetor de ruído de 100 dimensões, que é projetado por uma camada densa para um tensor e remodelado para uma base espacial de

4x4x1024. A partir desta base, cinco camadas de convolução transposta (Conv2DTranspose) com passo (stride) de 2 são utilizadas para progressivamente aumentar a resolução espacial até a saída final de 128x128x3. A Normalização em Lote e a função de ativação ReLU são aplicadas após cada camada intermediária, enquanto a camada de saída utiliza a função tanh.

- **Discriminador:** A arquitetura recebe uma imagem de 128x128x3 e a processa através de cinco camadas convolucionais (Conv2D) com passo de 2, que reduzem a resolução espacial e extraem características hierárquicas. A função de ativação LeakyReLU é utilizada em todas as camadas convolucionais. A Normalização em Lote é aplicada a partir da segunda camada para estabilizar o treinamento. Por fim, o mapa de características resultante é achatado (Flatten) e processado por uma camada densa com uma única saída para classificar a imagem como real ou falsa.

TREINAMENTO:

O treinamento foi conduzido por um total de 1000 épocas, com um tamanho de lote (batch size) de 128. Para a otimização dos modelos, foi utilizado o otimizador Adam (KINGMA; BA, 2017). Os hiperparâmetros foram ajustados conforme as recomendações do artigo de referência da DCGAN para promover a estabilidade, definindo-se a taxa de aprendizado em 0.0002 e o parâmetro de momento, β_1 , em 0.5 (RADFORD; METZ; CHINTALA, 2016). As funções de perda utilizadas foram baseadas na Entropia Cruzada Binária (BinaryCrossentropy), configuradas para calcular o erro a partir dos logits de saída do Discriminador. O progresso do treinamento foi salvo periodicamente através de checkpoints.

AVALIAÇÃO:

A avaliação dos resultados foi realizada por meio de duas abordagens:

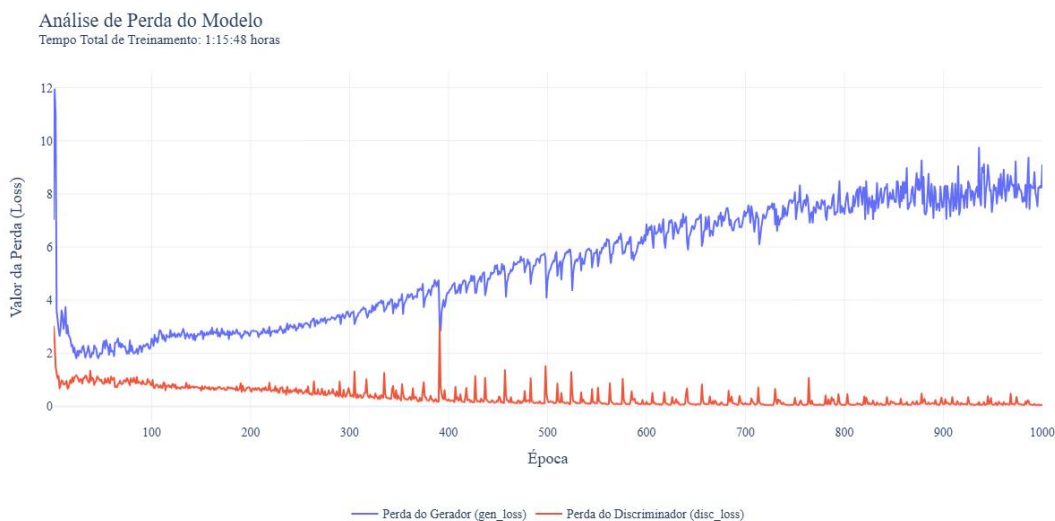
1. **Análise Qualitativa:** Inspeção visual das imagens geradas em diferentes épocas para avaliar a coerência, a qualidade estética e a diversidade das amostras. Os checkpoints das épocas com as melhores imagens foram então usados para gerar um lote de 1600 imagens para que pudessem passar pela próxima etapa de análise.
2. **Análise Quantitativa:** A qualidade e o realismo das imagens geradas foram avaliados quantitativamente utilizando a métrica Fréchet Inception Distance (FID). Proposta por Heusel et al. (2018), esta métrica compara as características das imagens sintéticas (geradas pelo modelo treinado até a época do checkpoint selecionado na etapa da análise visual) com as características de imagens reais do conjunto de validação, indicando o

quão realistas e diversas as imagens geradas são em relação às reais. Um valor de FID menor demonstra maior qualidade e realismo.

RESULTADOS E DISCUSSÃO

Neste primeiro experimento, o modelo DCGAN foi treinado utilizando os hiperparâmetros de otimização propostos por Radford, Metz e Chintala (2016), com o objetivo de estabelecer uma linha de base (baseline) para o desempenho da arquitetura no 102 Flower Category Dataset. A evolução das funções de perda do Gerador e do Discriminador ao longo do treinamento é apresentada na Figura 1.

Gráfico 1 - Gráfico da evolução das funções de perda do Gerador e do Discriminador ao longo das 1000 épocas de treinamento. As perdas são calculadas utilizando a Entropia Cruzada Binária.



Fonte: Autoria própria

A análise do gráfico revela três fases distintas. Inicialmente, entre as épocas 1 e 200, observa-se uma dinâmica adversarial estável, na qual as perdas de ambos os modelos possuem uma diferença de valores menores entre si. Durante esta fase, o Gerador começou a aprender características visuais relevantes, como texturas, cores e formas primárias análogas a pétalas, embora as imagens resultantes ainda fossem abstratas e pouco definidas.

Figura 1 - Grade de 16 imagens de flores sintéticas geradas pelo Gerador DCGAN após 100 épocas de treinamento. As imagens demonstram que o modelo começa a gerar formas, cores e texturas características do 102 Flower Category Dataset, embora ainda sejam imagens irreconhecíveis.



Fonte: Autoria própria

A partir da época 300, inicia-se uma segunda fase caracterizada pela convergência acentuada do Discriminador. Sua função de perda (*disc_loss*) diminui progressivamente, indicando um aumento em sua capacidade de distinguir amostras reais das sintéticas. Consequentemente, a perda do Gerador (*gen_loss*) inicia uma tendência de crescimento contínuo, demonstrando a crescente dificuldade do Gerador em enganar o Discriminador. É neste intervalo, por volta da época 440, que se observou a maior qualidade visual das amostras geradas. As imagens (Figura 2) apresentavam flores com estruturas mais coerentes e distinção entre pétalas e folhagem, embora ainda contivessem artefatos e distorções.

Figura 2 - Grade de 16 imagens de flores sintéticas geradas pelo Gerador DCGAN após 440 épocas de treinamento. As imagens demonstram a capacidade do modelo em capturar formas, cores e texturas características do 102 Flower Category Dataset.



Fonte: Autoria própria

A terceira e última fase, aproximadamente após o pico de melhora da época 600 (Figura 3), é marcada pela divergência do Gerador. O contínuo aumento de sua função de perda corresponde a uma degradação na qualidade visual das imagens. Nesta etapa, observou-se a geração de amostras distorcidas, irreconhecíveis e com baixa diversidade (Figura 4), um sintoma consistente com o fenômeno de instabilidade conhecido como colapso de modo (*mode collapse*). Isso sugere que treinar o modelo por 1000 épocas é excessivo para o tamanho do dataset.

Figura 3 - Grade de 16 imagens de flores sintéticas geradas pelo Gerador DCGAN após 600 épocas de treinamento (ponto de melhor FID). As imagens demonstram a capacidade do modelo em capturar formas, cores e texturas características do 102 Flower Category Dataset.



Fonte: Autoria própria

Figura 4 - Grade de 16 imagens de flores sintéticas geradas pelo Gerador DCGAN após 800 épocas de treinamento. A imagem ilustra a perda de coerência do modelo em comparação com os resultados das épocas 440 e 600, confirmando a instabilidade do treinamento tardio.



Fonte: Autoria própria

Os resultados deste experimento indicam que, embora a arquitetura DCGAN seja capaz de aprender a distribuição de um dataset complexo, os hiperparâmetros de treinamento propostos no artigo original não foram suficientes para garantir a estabilidade a longo prazo neste cenário específico. O fortalecimento excessivo do Discriminador levou a um problema de gradientes desvanecentes (*vanishing gradients*), no qual o Gerador deixa de receber um sinal de erro informativo para aprimorar seus pesos, resultando na estagnação e subsequente degradação da qualidade das amostras.

A comparação visual direta (Figura 5) revela aspectos fundamentais sobre a capacidade generativa do modelo. Embora as imagens sintéticas (linha superior), revelem que o modelo conseguiu capturar adequadamente as formas e características de flores reais, como estruturas básicas de pétalas, notam-se inconsistências significativas. Algumas das imagens geradas mostram-se distorcidas e incompreensíveis, enquanto em outras, partes da flor são representadas corretamente, mas com pétalas flutuantes ou desproporcionais. Outro desafio para o modelo reside na representação de estruturas complexas como as estruturas reprodutivas centrais da flor, como o pistilo e os estames, que tendem a ser simplificados ou distorcidos.

Figura 5 - Grade de 8 imagens, 3 de flores sintéticas geradas pelo Gerador DCGAN após 600 épocas de treinamento (Linha de cima da grade), e 4 de imagens do 102 Flowers Category Dataset (Linha de baixo).



Fonte: Autoria própria

Tabela 1 - Pontuação da Métrica Fréchet Inception Distance (FID) em diferentes épocas do treinamento. O menor valor indica o melhor desempenho do modelo.

Resultados Quantitativos:

RESULTADOS FINAIS DO FID	
Época do Checkpoint	Pontuação FID (Menor é Melhor)
100	1328.57
200	1106.23
300	859.17
400	908.08
440	812.08
500	1007.29
600 (Melhor Modelo)	808.74
700	1090.53
800	1319.33
900	1135.34
1000	1301.98

Fonte: Autoria própria

A avaliação quantitativa final comparou as amostras geradas em diferentes épocas com um conjunto de validação de 6.551 imagens. Os resultados, exibidos na Tabela 1, indicam uma curva de aprendizado evidente, com a performance aprimorando até a época 440, decaindo na época 500 e atingindo seu melhor desempenho na época 600. Após este ponto, a pontuação volta a aumentar, indicando que o equilíbrio do treinamento foi perdido e a qualidade das amostras entrou em um processo de degradação.

CONSIDERAÇÕES FINAIS

Este trabalho conseguiu implementar, treinar e avaliar um modelo de rede neural (DCGAN) para gerar imagens de flores. O objetivo de testar essa arquitetura em um dataset complexo foi atingido, mostrando tanto o que a técnica pode fazer quanto suas limitações.

A principal descoberta foi que o melhor desempenho do modelo aconteceu no meio do treinamento. A análise dos números, usando a medida FID (Heusel et al., 2018), confirmou o que se via nas imagens: a época 600 foi o melhor momento, com um FID de 808.74. Depois disso, a

qualidade piorou, mostrando que o treinamento desse tipo de modelo é instável quando se prolonga demais. Este comportamento é consistente com o fenômeno de overfitting do discriminador, um desafio documentado no treinamento de GANs com dados limitados (KARRAS et al., 2020). Os resultados deste trabalho indicam que, quando o discriminador atinge um nível de eficácia muito alto, o gerador deixa de receber informações úteis, o que impede o avanço do processo de aprendizagem.

É neste contexto que a pontuação FID obtida, de 808.74, embora numericamente elevada, pode ser contextualizada através da literatura existente sobre limitações da arquitetura DCGAN e características do dataset utilizado. Trabalhos como o de Heusel et al. (2018) demonstram que valores de FID variam drasticamente entre diferentes datasets e arquiteturas, com modelos mais antigos como DCGAN tipicamente apresentando pontuações na faixa de centenas a milhares em datasets complexos. Jayasumana et al. (2024) também destacam que a métrica FID deve ser interpretada relativamente ao contexto específico do experimento, considerando a diversidade e complexidade do dataset.

O resultado alcançado estabelece uma linha de base quantitativa importante para o 102 Flower Category Dataset, utilizando a arquitetura DCGAN. Isso serve como um ponto de referência para futuras comparações e melhorias arquitetônicas neste domínio específico. Contudo, mesmo no seu melhor desempenho, o modelo desenvolvido não pode ser empregado para expandir o conjunto de dados em outros estudos (data augmentation), devido à significativa presença de imagens ininteligíveis ou distorcidas que poderiam introduzir ruído nos dados de treinamento e comprometer o desempenho de classificadores.

As dificuldades encontradas apontam para o que pode ser feito no futuro. Usar arquiteturas generativas mais avançadas, como StyleGAN ou Modelos de Difusão, seria um passo lógico para melhorar a qualidade e conseguir FIDs menores. Além disso, criar uma versão condicional do modelo (cGAN) permitiria gerar tipos específicos de flores, dando mais controle e aumentando a utilidade da ferramenta.

Agradecimentos

À equipe de estudantes. Ao apoio financeiro recebido do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil.

REFERÊNCIAS

GOODFELLOW, Ian J. et al. *Generative adversarial networks*. 2014. Disponível em: <https://arxiv.org/abs/1406.2661>. Acesso em: 4 set. 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. Cambridge: MIT Press, 2016. Disponível em: <http://www.deeplearningbook.org>. Acesso em: 4 set. 2025.

HEUSEL, Martin; RAMSAUER, Hubert; UNTERTHINER, Thomas; NESSLER, Bernhard; HOCHREITER, Sepp. *GANs trained by a two time-scale update rule converge to a local Nash equilibrium*. 2018. Disponível em: <https://arxiv.org/abs/1706.08500>. Acesso em: 30 ago. 2025.

IOFFE, Sergey; SZEGEDY, Christian. *Batch normalization: accelerating deep network training by reducing internal covariate shift*. 2015. Disponível em: <https://arxiv.org/abs/1502.03167>. Acesso em: 22 ago. 2025.

JAYASUMANA, Sadeep et al. *Rethinking FID: towards a better evaluation metric for image generation*. 2024. Disponível em: <https://arxiv.org/abs/2401.09603>. Acesso em: 30 ago. 2025.

KARRAS, Tero; AITTALA, Miika; HELLSTEN, Janne; LAINE, Samuli; LEHTINEN, Jaakko; AILA, Timo. *Training generative adversarial networks with limited data*. 2020. Disponível em: <https://arxiv.org/abs/2006.06676>. Acesso em: 4 set. 2025.

KINGMA, Diederik P.; BA, Jimmy. *Adam: a method for stochastic optimization*. 2017. Disponível em: <https://arxiv.org/abs/1412.6980>. Acesso em: 30 ago. 2025.

RADFORD, Alec; METZ, Luke; CHINTALA, Soumith. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2015. Disponível em: <https://arxiv.org/abs/1511.06434>. Acesso em: 22 ago. 2025

WAN, Zhijing; WANG, Zhixiang; CHUNG, CheukTing; WANG, Zheng. *A survey of dataset refinement for problems in computer vision datasets*. 2023. Disponível em: <https://arxiv.org/abs/2210.11717>. Acesso em: 30 ago. 2025.