UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
PRÓ-REITORIA DE PESQUISA E EXTENSÃO
COORDENAÇÃO DE PESQUISA

RELATÓRIO DE ATIVIDADES DO ALUNO

Programa: PIBIC

Título do Projeto: Geração de Imagens Artificiais para Aumento de Dados Utilizando Redes Adversárias Generativas

Aluno: Alysson Machado de Oliveira Barbosa

Orientadora: Luciana Ribeiro Veloso

Alysson Machado de O. Barbosa

Sr. Alysson Machado de Oliveira Barbosa
Bolsista

Dra. Luciana Ribeiro Veloso
Orientadora

Campina Grande, 9 de setembro de 2023

Introdução

Modelos de inteligência artificial (*artificial intelligence*, IA) têm se destacado em diversas áreas devido aos avanços tecnológicos recentes, que permitiram o aumento do processamento e do armazenamento dos computadores modernos, além da exorbitante quantidade de dados produzidos pelos sistemas modernos (MAKRIDAKIS, 2017). Nesse contexto, técnicas de aprendizagem de máquina e aprendizado profundo (*deep learning*, DL) têm revolucionado o estado da arte em diversos domínios, como visão computacional, processamento de sinais e imagens, processamento de linguagem natural, entre outros.

Uma das áreas que mais se beneficia dos avanços tecnológicos do DL é a visão computacional, cujo objetivo é modelar algoritmos capazes de compreender as informações contidas em vídeos ou imagens digitais com alto nível de precisão (SONKA; HLAVAC; BOYLE, 2014). Os algoritmos canônicos em DL para a visão computacional são a detecção de objetos (e.g. OUYANG et al., 2016; DIBA et al., 2017), rastreamento de movimento (e.g. DOULAMIS; VOULODIMOS, 2016; DOULAMIS, 2018), reconhecimento de gestos (e.g. LIN et al., 2016; CAO; NEVATIA, 2016), estimativa de pose humana (e.g. TOSHEV; SZEGEDY, 2014; CHEN; YUILLE, 2014) e segmentação semântica e de instâncias (e.g. GEUS; MELETIS; DUBBELMAN, 2018; RUIZ-SANTAQUITERIA et al., 2020).

As técnicas utilizadas na área da visão computacional incluem processar, analisar e compreender as características contidas nas imagens digitais para produzir informações simbólicas ou numéricas que fazem sentido para os processos de pensamento lógico (KLETTE, 2014). Essas informações correspondem a resultados preditivos que agregam valor aos algoritmos supracitados de DL. O estado da arte do DL tem evoluído e apresentado avanços significativos ao longo dos anos, tornando-se uma área promissora e sofisticada para auxiliar na resolução de problemas da engenharia.

Desde que Yann LeCun propôs um algoritmo baseado em redes neurais para reconhecimento de dígitos escritos à mão, em 1989 (LECUN et al., 1989), até os algoritmos utilizados em carros autônomos na atualidade (FAN et al., 2020), os algoritmos de DL têm influenciado e moldado profundamente a indústria e a interação do ser humano com as novas tecnologias (MAKRIDAKIS, 2017). Entretanto, à medida que a complexidade e a capacidade de resolução de problemas dos algoritmos de DL cresceram, a demanda por bases de dados mais ricas em qualidade e quantidade também aumentou (NANDY; DUAN; KULIK, 2022), indicando que técnicas focadas na produção de mais dados representativos são relevantes para fomentar o crescimento dos algoritmos de DL.

Considerando a problemática em questão para a realização do aumento de dados, os métodos baseados em aprendizagem generativa têm se mostrado relevantes para melhorar as bases de dados quantitativamente e qualitativamente. Especialmente, ao considerar que as pesquisas sobre essa temática evoluíram notavelmente nos últimos nove anos (SALEHI; CHALECHALE; TAGHIZADEH, 2020). Dentre os métodos contemplados na categoria de aprendizagem, as redes neurais adversárias generativas (*generative adversarial networks*, GANs) têm se destacado, pois através da modelagem implícita da distribuição dos dados de alta dimensão (por exemplo, imagens digitais), torna-se pos-

sível gerar novas amostras de dados de forma sintética e realística (GOODFELLOW et al., 2014). Entretanto, mesmo considerando os avanços notáveis nesta área de pesquisa, muitos fatores relacionados às perspectivas futuras do uso dessa tecnologia devem ser analisados e ponderados para que seu uso seja mais eficaz e consonante com o progresso dos algoritmos de DL (TURHAN; BILGE, 2018).

A utilização de grandes quantidades de dados para treinar modelos de IA é fundamental para garantir um bom desempenho de modelos aplicáveis de DL, sendo tão importante quanto o desenvolvimento arquitetural dos algoritmos (ZHANG et al., 2018). Além da quantidade, a qualidade representativa dos dados também é relevante. A partir das características presentes nos dados, o algoritmo aprende a tomar uma decisão lógica para a realização de uma tarefa específica, tais como classificação, localização e segmentação (POUYANFAR et al., 2018). Entretanto, o processo de obtenção dos dados pode apresentar obstáculos que dificultam o desenvolvimento de um modelo de alto desempenho. Em relação à utilização de imagens em aplicações de DL, é possível relatar que:

1. A disponibilidade de conjuntos de imagens públicas na internet é escassa, especialmente no desenvolvimento de projetos de DL que envolvam aplicações mais específicas ou de grau de complexidade mais elevado (HASAN et al., 2022);

2. A criação de conjuntos de imagens é um processo que abrange várias etapas, envolvendo a captação de dados, rotulagem, organização estrutural e revisão dos dados coletados e tratados. Dependendo do tipo de aplicação, como em aplicações do domínio da radiologia e do sensoriamento remoto, a consulta de especialistas é necessária para lidar com essa tarefa. Portanto, a criação de conjuntos de imagens é um processo complexo e potencialmente oneroso (WANG et al., 2017).

Visando contornar essas dificuldades, diversas técnicas foram propostas para viabilizar o uso de pequenos conjuntos de imagens em aplicações de DL, tais como: técnicas de aumento de dados e técnicas de transferência de aprendizagem.

As técnicas de aumento de dados, que utilizam algoritmos de processamento de imagens digitais, são úteis para ampliar a base de dados e melhorar o treinamento dos algoritmos de DL. Nesse processo, transformações geométricas (e.g. rotação, translação, cisalhamento, etc.) e não geométricas (e.g. aplicações de filtros espaciais de suavização e aguçamento) são utilizadas para aumentar a quantidade de imagens em um conjunto de dados original (SHORTEN; KHOSHGOFTAAR, 2019).

As técnicas de transferência de aprendizagem consistem na reutilização dos parâmetros (pesos sinápticos e bias) de um modelo pré-treinado em um novo modelo que desempenha uma tarefa distinta do modelo pré-treinado. Em linhas gerais, trata-se de treinar um modelo previamente utilizando um conjunto de imagens robustas X_D e reaproveitá-lo em uma segunda tarefa que envolve um conjunto de imagens Y_D. Esse processo permite um progresso rápido e eficaz ao modelar a segunda tarefa (PAN; YANG, 2009).

As Figuras 1 e 2 ilustram as técnicas descritas de aumento de dados e transferência de aprendizagem, respectivamente. Os procedimentos citados são bastante comuns em pesquisas da área de DL, permitindo o desenvolvimento de aplicações complexas e de grande relevância, mesmo quando as bases de dados disponíveis para essas aplicações apresentam limitações.

Figura 1 – Um exemplo de aumento de dados utilizando transformações geométricas e não geométricas pode ser encontrado em uma aplicação de classificação automatizada de plantas com o uso de DL (PAWARA et al., 2017).

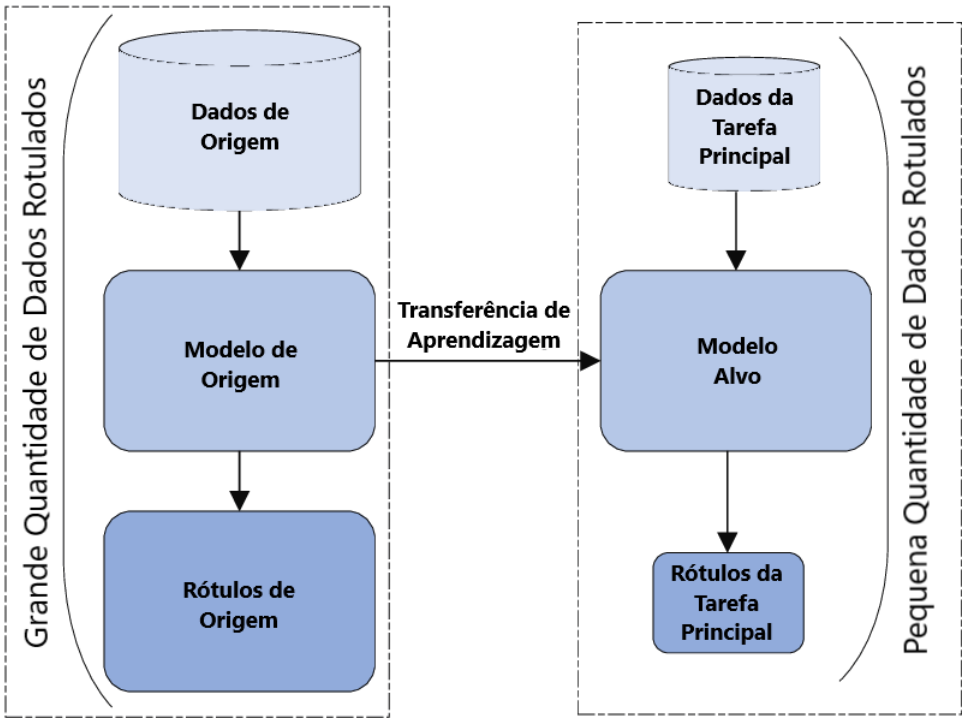


Figura 2 – Diagrama esquemático da técnica de transferência de aprendizagem para otimização de modelos (ALZUBAIDI et al., 2020).

Entretanto, as técnicas mencionadas referem-se as práticas que envolvem manipulação de conjuntos de imagens já existentes, seja de maneira direta (aumento de dados) ou indireta (transferência de aprendizagem), a fim de permitir que os algoritmos de DL extraiam as características discrimina-

tórias do problema abordado de forma mais eficiente. Diante desse cenário, modelos capazes de gerar amostras inéditas com a mesma distribuição estatística do conjunto original têm se tornado cada vez mais relevantes no estado da arte de DL (TURHAN; BILGE, 2018). Esses modelos são conhecidos como modelos generativos, e as GANs, pertencentes ao domínio desses tipos de modelos, destacam-se por sua robustez e capacidade de lidar adequadamente com a geração de imagens de alta dimensão (TURHAN; BILGE, 2018). As GANs são capazes de preservar as informações sobre os padrões e as qualidades dos dados originais da base ao gerar novas amostras.

As GANs possuem o potencial de aprender a imitar qualquer distribuição de conjunto de imagens, a fim de gerar novos dados sintéticos semelhantes aos padrões e com a qualidade realística dos dados originais. Como resultado, as GANs são capazes de realizar o procedimento de aumento de dados sem envolver transformações diretas das imagens da base de dados. Em vez disso, elas utilizam as características e detalhes presentes nos dados originais para recriar imagens inusitadas e exclusivas (sistemas generativos de imagens) que possuem coerência visual e lógica (CRESWELL et al., 2018). Assim, a estratégia de aumento de dados com GANs para geração de imagens artificiais tem ganhado relevância entre a comunidade de pesquisadores da área de DL (SALEHI; CHALECHALE; TAGHIZADEH, 2020), principalmente ao se considerar que esses procedimentos são mais eficazes em lidar com imagens mais complexas do que as técnicas tradicionais citadas anteriormente (BROCK; DONAHUE; SIMONYAN, 2018; TANAKA; ARANHA, 2019).

As GANs foram propostas por Ian Goodfellow em 2014 (GOODFELLOW et al., 2014), um dos principais incentivadores do avanço da teoria de DL no meio científico. Yann LeCun, diretor de pesquisa no *Facebook*, expandiu a teoria das GANs aplicando ideias em diversos domínios de DL (LECUN, 2016). Desde 2014, houve um rápido crescimento no número de artigos publicados sobre as GANs, permitindo que os modelos generativos ganhassem robustez e relevância no desenvolvimento de projetos de DL. As Figuras 3 e 4 ilustram o aumento no número de publicações sobre as GANs de 2014 a 2019, e a evolução desses sistemas generativos de imagens ao longo dos anos, respectivamente.

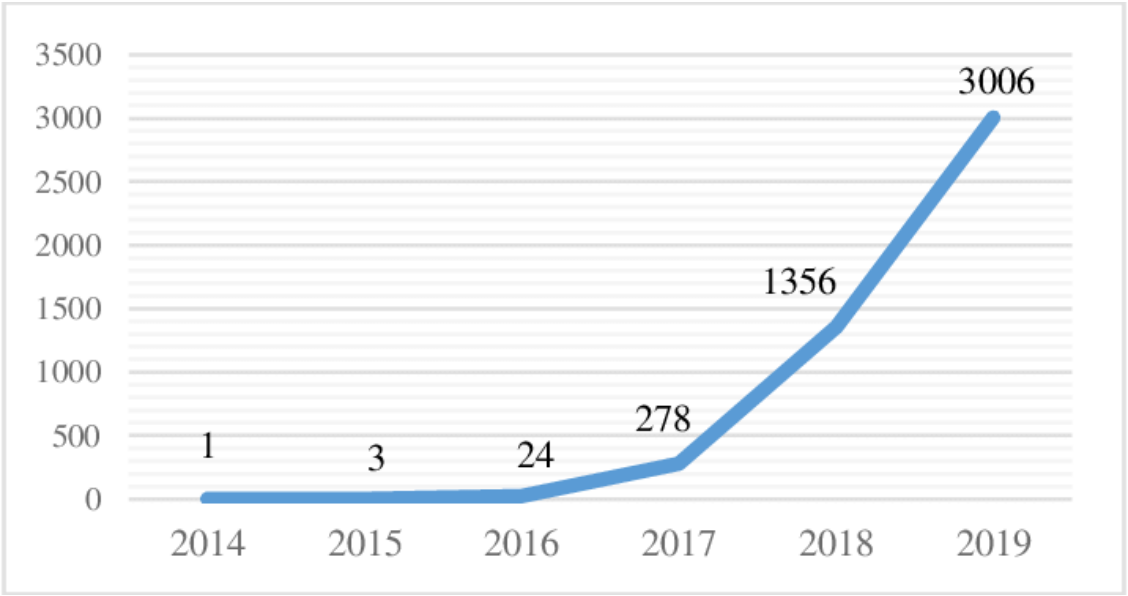


Figura 3 – Número de artigos indexados pela *Scopus* sobre GANs de 2014 a 2019 (SALEHI; CHALECHALE; TAGHIZADEH, 2020).

Portanto, o desenvolvimento de algoritmos voltados para métodos generativos de criação de imagens artificiais utilizando GANs contribui para o desenvolvimento de soluções de aumento de dados em aplicações de DL. Assim, o projeto proposto possibilita a consolidação de aspectos práticos, como a elaboração de uma técnica que permite otimizar algoritmos de DL em cenários em que os conjuntos de imagens para treinamento dos modelos são escassos.

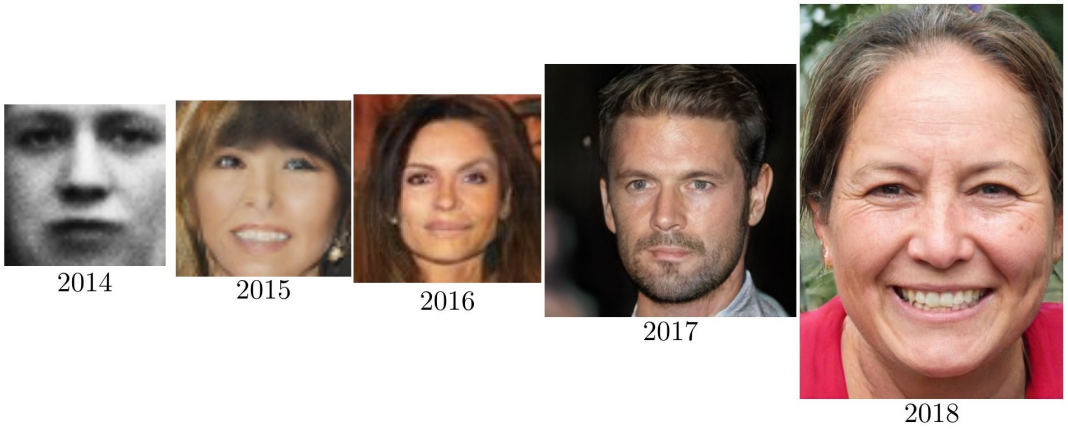


Figura 4 – Evolução da aplicação das GANs para geração de rostos artificiais ao longo dos anos. 2014 (GOODFELLOW et al., 2014), 2015 (RADFORD; METZ; CHINTALA, 2015), 2016 (LIU; TUZEL, 2016), 2017 (KARRAS et al., 2017) e 2018 (KARRAS; LAINE; AILA, 2019).

Dentro deste contexto, o presente projeto PIBIC tem como objetivo a geração de imagens artificiais para aumento de dados utilizando GANs, investigando a qualidade da representação visual, a diversidade e a fidelidade das imagens geradas através desta técnica. Para tal, serão utilizadas bases de imagens disponíveis na internet com diferentes graus de resolução, quantidade e complexidade da representação visual dos objetos presentes nas imagens. Neste trabalho, considera-se a geração de imagens artificiais como a capacidade de gerar imagens sintéticas e inéditas que imitam a distribuição de características de um conjunto de dados original.

Objetivos

O objetivo deste projeto é o desenvolvimento de um sistema capaz de aumentar conjuntos de imagens por meio da geração de amostras sintéticas utilizando GANs e outros algoritmos de sistemas generativos. Espera-se que o sistema gere imagens artificiais com boa representação visual, diversidade e fidelidade em relação ao grau de detalhamento presente nas imagens originais das bases de dados testadas. Os objetivos específicos deste projeto são:

I. Identificar e estudar as técnicas existentes que possibilitem o aumento de dados com imagens artificiais por meio das GANs e outros sistemas generativos de imagens;

II. Realizar um levantamento de bases de imagens disponíveis na internet, considerando o uso de bases de imagens com diferentes graus de detalhamento da informação (e.g. quantidade e complexidade de elementos visuais presentes nas imagens);

III. Selecionar, implementar e treinar arquiteturas das GANs para o aumento de dados nas bases de imagens pesquisadas;

IV. Avaliar o desempenho dos modelos produzidos com relação as novas imagens geradas artificialmente, considerando a qualidade da representação visual, diversidade e fidelidade com as principais características das bases de imagens;

V. Verificar o ganho de desempenho proporcionado pelos algoritmos de aumento de base dados realizando testes em tarefas canônicas do DL (e.g. classificação, localização, segmentação, etc.) com base em diversas métricas avaliativas.

Material e Métodos/Metodologia

A metodologia deste projeto baseia-se no desenvolvimento de arquiteturas de GANs para aumentar dados em conjuntos de imagens. Inicialmente, foram realizados estudos sobre as técnicas mais recentes em GANs, com especialização na plataforma *Coursera* (ZHOU; ZELIKMAN, 2019). Além disso, foi realizado um estudo para aprender a utilizar o *framework PyTorch* (PASZKE et al., 2019) para projetar as GANs. Em seguida, foi realizada uma revisão sistemática dos artigos mais relevantes para o desenvolvimento das GANs e a seleção de quatro bases de dados para testar as arquiteturas a serem construídas posteriormente. Cinco tipos de arquiteturas de GANs foram selecionadas e, por fim, as arquiteturas foram treinadas nas bases de dados selecionadas. Esses procedimentos serão descritos detalhadamente ao longo da metodologia.

A primeira etapa do projeto consistiu na especialização em GANs. A especialização em GANs do *Coursera* é composto por três cursos (ZHOU; ZELIKMAN, 2019), que fornecem uma ampla e completa introdução ao campo das GANs. O primeiro curso explora a arquitetura básica da GAN e seu treinamento, enquanto o segundo curso aprofunda a compreensão de GANs em escala e técnicas de regularização, além de apresentar diversas métricas avaliativas. O terceiro curso aborda as aplicações de GANs em problemas do mundo real, incluindo a síntese de imagens e a geração de texto. Além disso, os cursos oferecem atividades práticas aplicáveis em projetos do mundo real. Ao finalizar o curso, foi estabelecida uma compreensão sólida das técnicas de GANs, possibilitando uma boa visão para a construção das arquiteturas ao longo do projeto.

Na segunda etapa da pesquisa, foi realizada uma especialização no *framework PyTorch* utilizando a linguagem de programação *Python* (PASZKE et al., 2019). A especialização foi conduzida por meio de ampla pesquisa de tutoriais disponíveis na internet e de documentações sobre o uso desse *framework*. Durante o processo de aprendizado, foram exploradas as principais funcionalidades do *PyTorch*, incluindo a criação de modelos de redes neurais artificiais e diversas formas de otimizá-los. As habilidades adquiridas foram aplicadas no desenvolvimento de modelos de GANs e de DL em geral, contribuindo para o avanço do projeto de pesquisa.

Na terceira etapa da pesquisa, a revisão sistemática dos principais artigos sobre GANs foi realizada com o objetivo de fornecer uma compreensão teórica abrangente sobre a construção dessas

arquiteturas. Ela permitiu a identificação e análise crítica de artigos relacionados às GANs, possibilitando a obtenção de informações cruciais para a construção de modelos e solução de problemas práticos na área. Além disso, a revisão sistemática possibilitou a análise quantitativa e qualitativa do trabalho realizado durante o projeto e forneceu subsídios para a escrita de artigos acadêmicos em etapas posteriores. Essa revisão foi fundamental para a compreensão da literatura existente em GANs, e possibilitou a construção de uma base sólida para o projeto.

Na quarta etapa da pesquisa, foram selecionadas quatro bases de dados úteis para o treinamento de arquiteturas de GANs. A seleção das bases de dados foi realizada após a análise de diversos artigos que as utilizam para avaliar a capacidade dessas arquiteturas. A escolha das bases de dados foi baseada em uma revisão sistemática que levou em consideração não apenas a qualidade dos dados, mas também a limitação de hardware atual, que se restringe a GPUs não muito potentes fornecidas pelos servidores da plataforma *Google Colaboratory* (GOOGLE, 2023). As bases de dados escolhidas foram o MNIST (LECUN et al., 1998), Fashion MNIST (XIAO; RASUL; VOLLGRAF, 2017), EMNIST Letters (COHEN et al., 2017) e CelebA (LIU et al., 2015), amplamente utilizadas na literatura para avaliar o desempenho das GANs.

O MNIST é uma base de dados de dígitos manuscritos que consiste em 60.000 imagens de treinamento e 10.000 imagens de teste, cada uma delas em preto e branco e com dimensão de 28x28 pixels. Fashion MNIST é uma base de dados de roupas, também composta por 60.000 imagens de treinamento e 10.000 imagens de teste, em preto e branco e com as mesmas dimensões do MNIST. O EMNIST Letters é uma base de dados de letras manuscritas, contendo 145.600 imagens de treinamento e 24.400 imagens de teste, todas em preto e branco e com dimensões de 28x28 pixels. Por fim, o CelebA é uma base de dados de rostos de celebridades, com 202.599 imagens de rostos em cores e tamanho de 178x218 pixels. Como as arquiteturas de GANs têm finalidade distinta de outros algoritmos de DL (e.g. classificação, localização, segmentação, etc.), as imagens de teste também foram utilizadas no treinamento para tornar as arquiteturas mais robustas, uma vez que as métricas avaliativas dos modelos não necessitam dessas imagens como referência.

Na quinta etapa da pesquisa, diversas GANs foram treinadas para o projeto. As arquiteturas escolhidas para serem projetadas foram a GAN simples (GOODFELLOW et al., 2014), DCGAN (RADFORD; METZ; CHINTALA, 2015), WGAN-GP (ARJOVSKY; CHINTALA; BOTTOU, 2017; GULRAJANI et al., 2017), SNGAN (MIYATO et al., 2018) e WGAN-GP + SNGAN (ARJOVSKY; CHINTALA; BOTTOU, 2017; GULRAJANI et al., 2017; MIYATO et al., 2018), que é uma mistura das duas arquiteturas anteriores. Cada uma dessas arquiteturas apresenta distinções em sua forma de construção, implicando em resultados distintos de geração de imagens. Além disso, todas as arquiteturas selecionadas possuem controle condicional, que é uma modificação arquitetural que permite controlar o valor de cada classe. Essa abordagem permite que os geradores produzam imagens mais precisas e realistas, ao mesmo tempo em que oferece maior controle sobre o resultado final.

A arquitetura de uma GAN simples envolve a criação de duas redes neurais que competem entre si (GOODFELLOW et al., 2014). Tanto o gerador quanto o discriminador possuem arquitetura com camadas densamente conectadas, responsáveis por mapear as entradas para as saídas por meio de

cinco blocos de camadas densamente conectadas. Esses blocos são compostos por camadas lineares, normalização de lotes e função de ativação ReLU. Ao final desses blocos densamente conectado, é aplicada uma camada linear com a função de ativação sigmoide, e a função de custo utilizada é a entropia cruzada binaria. No gerador, as camadas densamente conectadas são seguidas por camadas que expandem a quantidade de parâmetros para transformar um vetor de ruído em uma imagem. No caso do discriminador, essas camadas são seguidas por uma camada de saída que produz um único valor entre 0 e 1, indicando a probabilidade de a entrada ser real, no caso, uma imagem. Durante o treinamento, a rede geradora é treinada para enganar a rede discriminadora, gerando amostras que se assemelham às imagens reais do conjunto de dados. Enquanto isso, a rede discriminadora é treinada para distinguir entre as imagens reais e as produzidas pelo gerador. Esse processo é repetido até que a rede geradora seja capaz de gerar amostras indistinguíveis das imagens reais. Devido à simplicidade dessa arquitetura, a base de dados CelebA não foi treinada devido a complexidade das características presentes nesse conjunto de imagens. A Figura 5 ilustra o diagrama arquitetural de uma GAN simples.

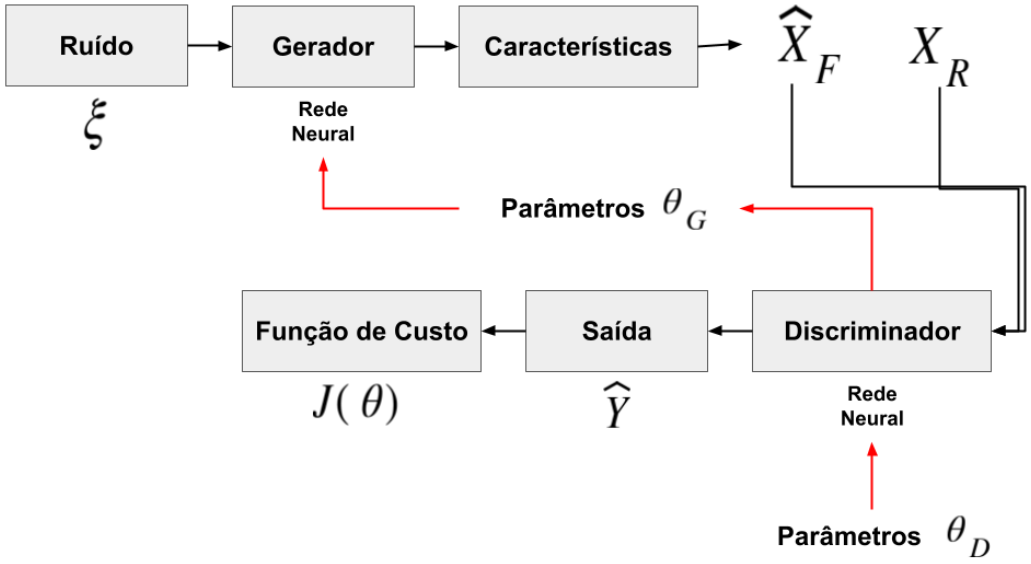


Figura 5 – Diagrama arquitetural básico de uma GAN Simples (ZHOU; ZELIKMAN, 2019). ξ representa o vetor de ruido, θ_G os parâmetros do gerador, θ_D os parâmetros do discriminador, $J(\theta)$ a função de custo, $\hat{X}_F$ as imagens criadas pelo gerador, X_R as imagens reais e $\hat{Y}$ a classificação feita pelo discriminador.

Para gerar imagens mais diversificadas e com maior fidelidade, é possível modificar o vetor de ruído utilizado como entrada do gerador. Alterações nesse vetor podem mudar as características da imagem gerada, como cores, formas e texturas. Uma modificação que foi adicionada em quase todas as arquiteturas criadas é a capacidade de controlar a classe alvo, ou seja, especificar na entrada da rede que tipo de imagens queremos com base nas possíveis classes disponíveis no conjunto de imagens treinadas. Essa técnica é chamada de controle condicional (MIRZA; OSINDERO, 2014).

Para realizar o controle condicional em uma GAN, é necessário modificar a arquitetura da rede. Uma das formas de fazer isso é incluir uma informação adicional, como um vetor codificado categoricamente com a classe alvo ativa, como entrada tanto para o gerador quanto para o discriminador. No gerador, esse processo ocorre apenas concatenando esse vetor codificado ao vetor de ruído.

No discriminador, esse processo acontece expandindo a quantidade de camadas de entrada da imagem. As camadas adicionais às já existentes na imagem consistem de um tensor composto de matrizes de apenas 1s ou 0s, em que a matriz de 1s representa a classe alvo e as matrizes de 0s representam as classes não alvo. Como essa técnica exige a utilização de camadas convolucionais no discriminador, considerando que é uma manipulação nos canais das imagens, o controle condicional não foi realizado na arquitetura de uma GAN simples, mas sim nas demais arquiteturas que foram construídas com base na teoria de redes neurais convolucionais (*convolutional neural networks*, CNNs). A Figura 6 ilustra como o controle condicional é aplicado nas arquiteturas de GANs.

A arquitetura DCGAN (*Deep Convolutional GAN*) é uma extensão da arquitetura básica do GAN Simples que busca melhorar a qualidade e a estabilidade dos resultados gerados pelo modelo (RADFORD; METZ; CHINTALA, 2015). Para alcançar esse objetivo, a DCGAN utiliza uma abordagem de CNN em vez de uma rede neural totalmente conectada. Essa abordagem permite que o modelo capture características mais complexas e abstratas das imagens, como bordas e texturas, o que resulta em imagens mais realistas e detalhadas. Além disso, a DCGAN incorpora outras modificações, como a utilização de camadas de normalização e a eliminação de camadas densamente conectadas, que ajudam a regularizar o modelo e a reduzir o tempo de treinamento. No geral, a DCGAN é uma arquitetura mais avançada e sofisticada do que a GAN simples, que se tornou uma das mais populares para a geração de imagens de alta qualidade. A Figura 6 além de ilustrar o funcionamento do controle condicional, ilustra o diagrama arquitetural de uma DCGAN.

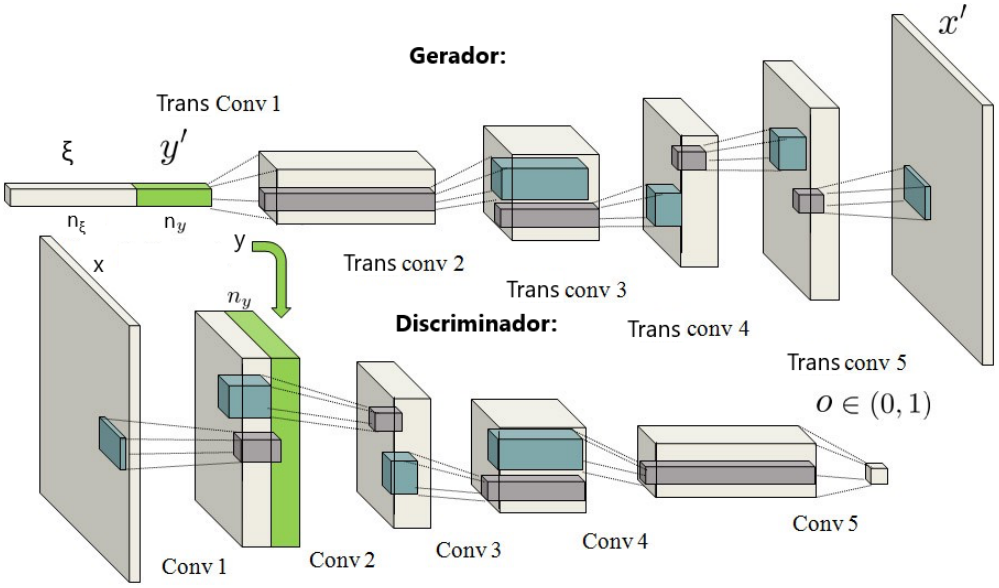


Figura 6 – Controle condicional aplicado a uma arquitetura de DCGAN. *Trans Conv i* é a convolução transposta aplicada ao i-ésimo bloco do gerador, *Conv i* é a convolução aplicada ao i-ésimo bloco do discriminador, ξ é o vetor de ruido, y representa as classes do conjunto de imagens, y' é o vetor codificado das classes do conjunto de imagens, n_ξ é a dimensão do vetor de ruído e n_y é a dimensão do vetor de classe codificado. (MIRZA; OSINDERO, 2014).

A arquitetura WGAN-GP (*Wasserstein GANs with Gradient Penalty*) é uma extensão da arquitetura DCGAN, que visa melhorar o desempenho do modelo na geração de imagens de alta qualidade (ARJOVSKY; CHINTALA; BOTTOU, 2017; GULRAJANI et al., 2017). A principal novidade dessa

arquitetura é a introdução de duas técnicas, a primeira é a utilização da função de perda *Wassertein*, que é uma métrica que mede a distância entre duas distribuições de probabilidade, a distribuição real dos dados e a distribuição produzida pela rede generativa. A segunda técnica é a aplicação do gradiente penalizado (GP), que é uma forma de suavizar a superfície de perda da função de perda *Wassertein*, permitindo um melhor treinamento da rede. A combinação dessas técnicas torna a WGAN-GP mais robusta em relação a problemas de convergência e instabilidade do treinamento, além de produzir resultados de qualidade superior aos modelos GAN convencionais. A Figura 7 ilustra o diagrama arquitetural da WGAN-GP.

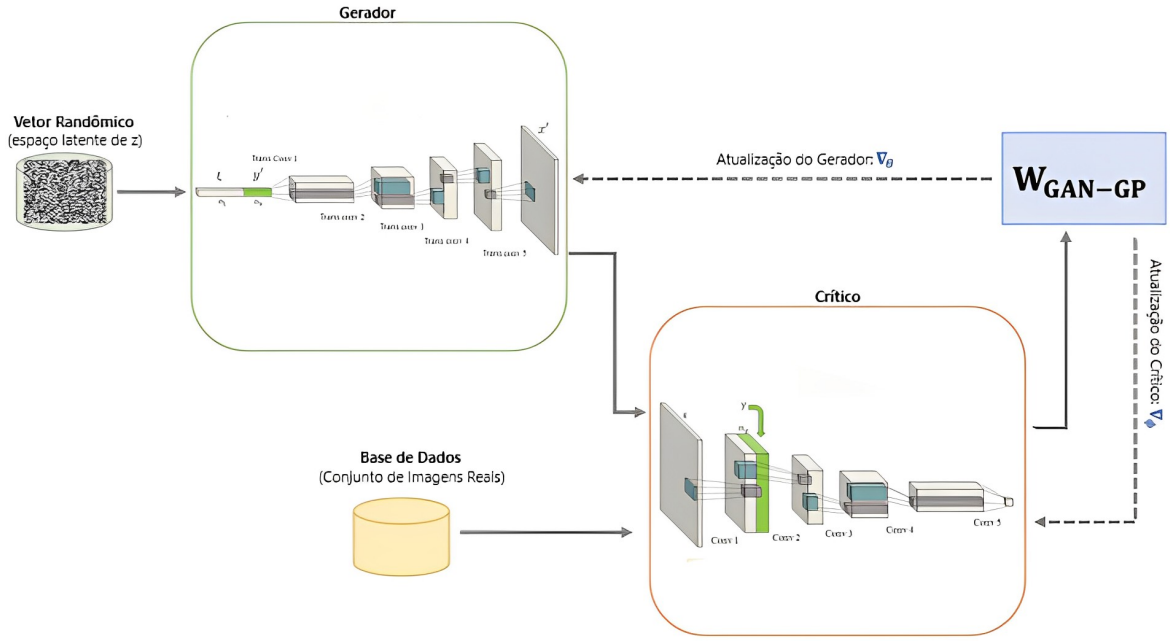


Figura 7 – Diagrama arquitetural básico da WGAN-GP (ARJOVSKY; CHINTALA; BOTTOU, 2017; GULRAJANI et al., 2017). A rede geradora tenta produzir amostras que se assemelhem às amostras reais. O crítico (ou rede discriminadora) tenta distinguir entre as amostras geradas e as amostras reais. Sua saída é linear, sem função de ativação, com o objetivo de utilizar a função de perda *Wassertein*, que é uma métrica que mede a distância entre duas distribuições de probabilidade, a distribuição real dos dados e a distribuição produzida pela rede generativa.

A arquitetura SNGAN (*Spectral Normalization GAN*) é uma extensão da arquitetura GAN que visa melhorar a estabilidade e a qualidade da geração de imagens (MIYATO et al., 2018). A principal inovação da SNGAN é a técnica de normalização espectral, que consiste em aplicar uma restrição na norma espectral das camadas da rede discriminadora. Isso ajuda a controlar o crescimento explosivo das normas das matrizes de peso, que pode levar a problemas de instabilidade durante o treinamento. Além disso, a arquitetura SNGAN também utiliza a função de perda de entropia cruzada binária. A combinação dessas técnicas permite que a rede discriminadora seja treinada de forma mais estável e, portanto, melhora a qualidade da geração de imagens. A Figura 8 ilustra a aplicação da normalização espectral nas camadas de uma arquitetura de GAN.

A arquitetura que combina as ideias do WGAN-GP e da SNGAN é conhecida como WGAN-GP + SNGAN, a qual utiliza a técnica de normalização espectral da SNGAN na rede discriminadora,

juntamente com a função de perda de *Wassertein* utilizada na arquitetura WGAN-GP (ARJOVSKY; CHINTALA; BOTTOU, 2017; GULRAJANI et al., 2017; MIYATO et al., 2018). Isso ajuda a controlar o crescimento explosivo das normas das matrizes de peso e, ao mesmo tempo, evita o problema de desaparecimento do gradiente e instabilidade do treinamento associados à função de perda baseada em entropia cruzada binária. A combinação dessas técnicas tem mostrado melhorias significativas na qualidade da geração de imagens em comparação com as arquiteturas WGAN-GP e SNGAN individuais.

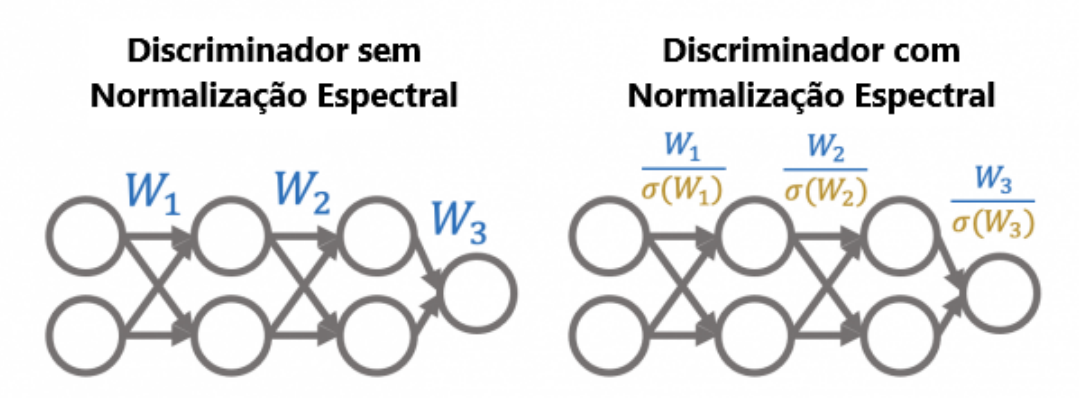


Figura 8 – Aplicação da normalização espectral na rede discriminadora de uma GAN (LIN; SEKAR; FANTI, 2021). A normalização espectral divide os pesos W_i por suas normas espectrais $\sigma(W_i)$ (ou seja, o maior valor singular de W_i).

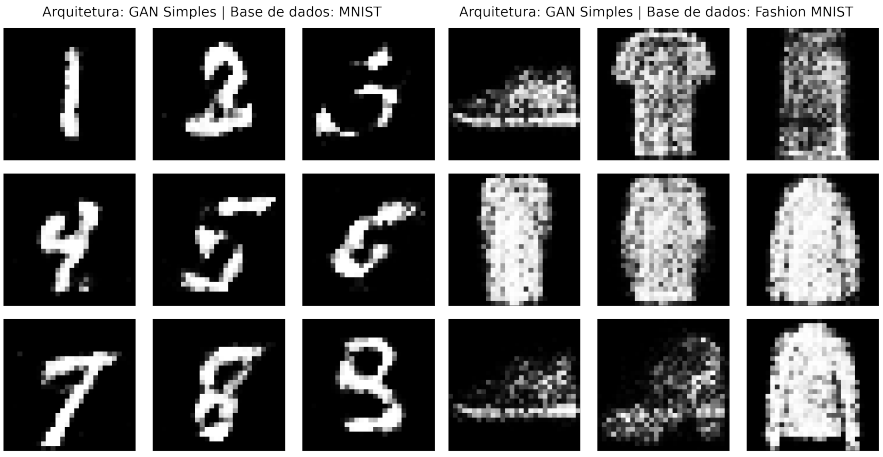
Todas as arquiteturas projetadas, incluindo a GAN simples, DCGAN, WGAN-GP, SNGAN e WGAN-GP + SNGAN, foram treinadas utilizando as bases de dados previamente pesquisadas. A MNIST, Fashion MNIST, EMNIST Letters e CelebA foram usadas para avaliar o desempenho de cada uma das arquiteturas. É importante destacar que a arquitetura da GAN simples não foi treinada na base de dados CelebA, sendo a única exceção no conjunto de arquiteturas utilizadas (ao todo, 19 modelos foram treinados). Vale lembrar que cada arquitetura foi configurada de acordo com as especificidades de cada base de dados, a fim de alcançar o melhor desempenho possível em relação aos resultados obtidos.

As próximas etapas da pesquisa envolvem a adição de métricas avaliativas para avaliar quantitativamente o desempenho das GANs treinadas. Essas métricas incluem *Inception Score* (SALIMANS et al., 2016), *Fréchet Distance* (HEUSEL et al., 2017), precisão e sensibilidade (BORJI et al., 2019). Além disso, será realizado o treinamento de um sistema generativo com o algoritmo chamado de autoencoder variacional (KINGMA; WELLING, 2013). Em seguida, experimentos de aumento de dados serão realizados com os modelos de sistemas generativos treinados (GANs e autoencoder variacional), para verificar se essa estratégia de aumento de dados sintéticos traz benefícios para outros domínios do aprendizado profundo, como a classificação. Por fim, pretende-se expandir as técnicas utilizadas para modelar arquiteturas de GANs, como o método pix-2-pix (ISOLA et al., 2017) e CycleGAN (ZHU et al., 2017), buscando expandir a qualidade, tanto em diversidade quanto em fidelidade, dos sistemas generativos propostos para aumento de dados. Com essas etapas, espera-se obter resultados significativos para contribuir no avanço das pesquisas com sistemas generativos.

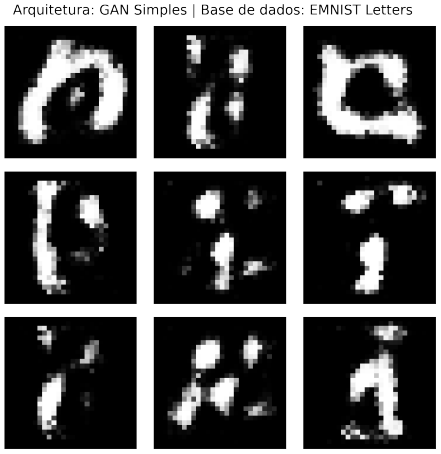
Resultados Parciais

Como resultado parcial deste projeto PIBIC, são descritos os resultados obtidos com o treinamento das arquiteturas de GANs Simples, DCGAN, WGAN-GP, SNGAN e WGAN-GP + SNGAN, utilizando as bases de dados selecionadas: MNIST, Fashion MNIST, EMNIST Letters e CelebA. No total, foram treinadas 19 redes de GANs. Os próximos parágrafos apresentam os principais resultados parciais e contribuições deste projeto, bem como as análises realizadas para avaliar a capacidade das GANs para geração de novas amostras de dados.

Os resultados obtidos no treinamento da arquitetura de GAN Simples revelaram baixa fidelidade na geração de novas imagens, o que não permite um aumento efetivo nos conjuntos de imagens estudados. Observou-se que a principal causa dessa falta de fidelidade está relacionada à não utilização de CNNs, que têm a capacidade de lidar com imagens de forma mais eficaz do que as camadas totalmente conectadas presentes na GAN Simples. Esses resultados indicam a necessidade de explorar outras arquiteturas de GANs que contemplem a utilização de CNNs, com o objetivo de melhorar a qualidade e a fidelidade das imagens geradas para a ampliação dos conjuntos de dados. A Figura 9 ilustra alguns exemplos de imagens geradas pela GAN Simples.



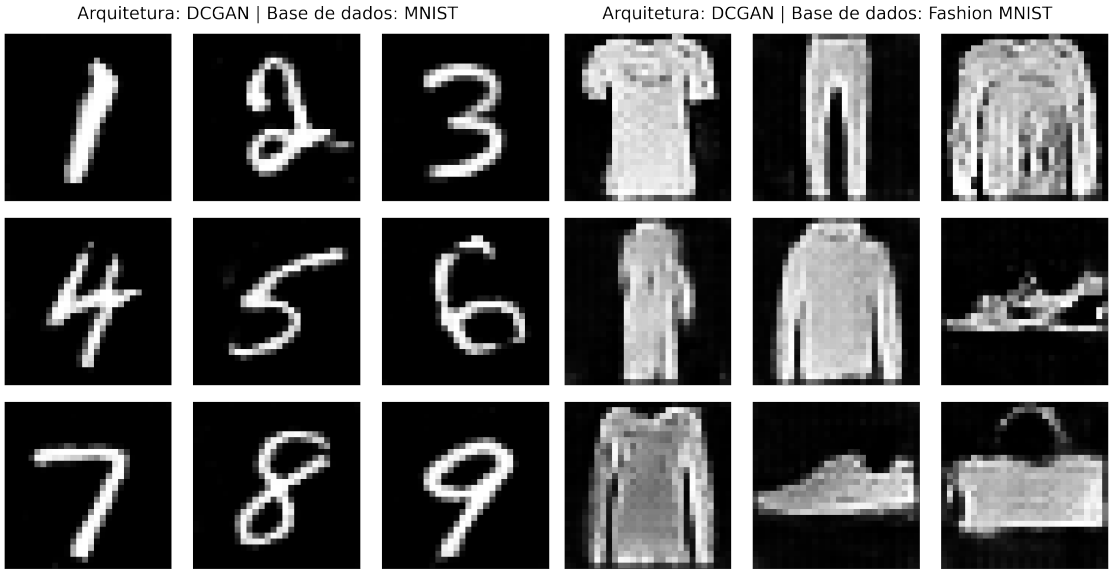
(a) Base de dados MNIST. (b) Base de dados Fashion MNIST.



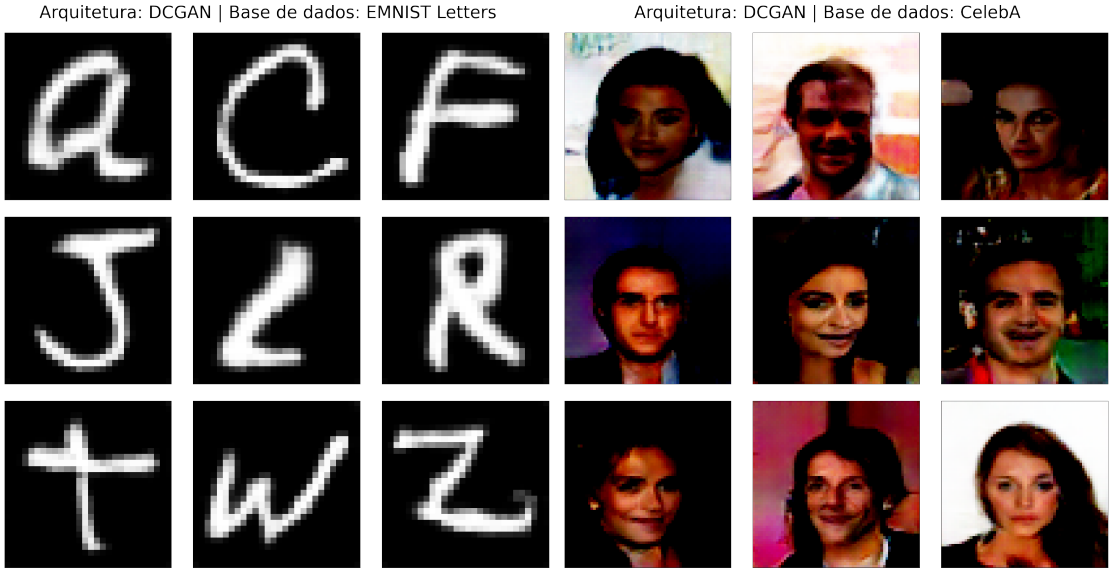
(c) Base de dados EMNIST Letters.

Figura 9 – Imagens geradas pela arquitetura de GAN Simples treinada com as bases de dados (a) MNIST, (b) Fashion MNIST e (c) EMNIST Letters.

Os resultados obtidos após o treinamento da arquitetura DCGAN nas bases de dados selecionadas foram satisfatórios. A utilização de camadas de convolução na rede geradora permitiu melhorar significativamente a qualidade e a fidelidade das imagens geradas, possibilitando sua utilização para a realização de aumento de dados nos conjuntos de imagens estudados. De maneira particular, observou-se que a DCGAN conseguiu gerar imagens de faces na base de dados CelebA com boa fidelidade à representação humana, com detalhes como a silhueta, os cabelos, o nariz e a boca bem definidos. Essas descobertas reforçam a importância da utilização de arquiteturas de GAN que levem em consideração as particularidades dos conjuntos de dados em análise, para que ocorra um bom aprendizado por parte das redes para reconstrução das características presentes na distribuição dos dados dos conjuntos de imagens. A Figura 10 ilustra alguns exemplos de imagens geradas pela DCGAN.



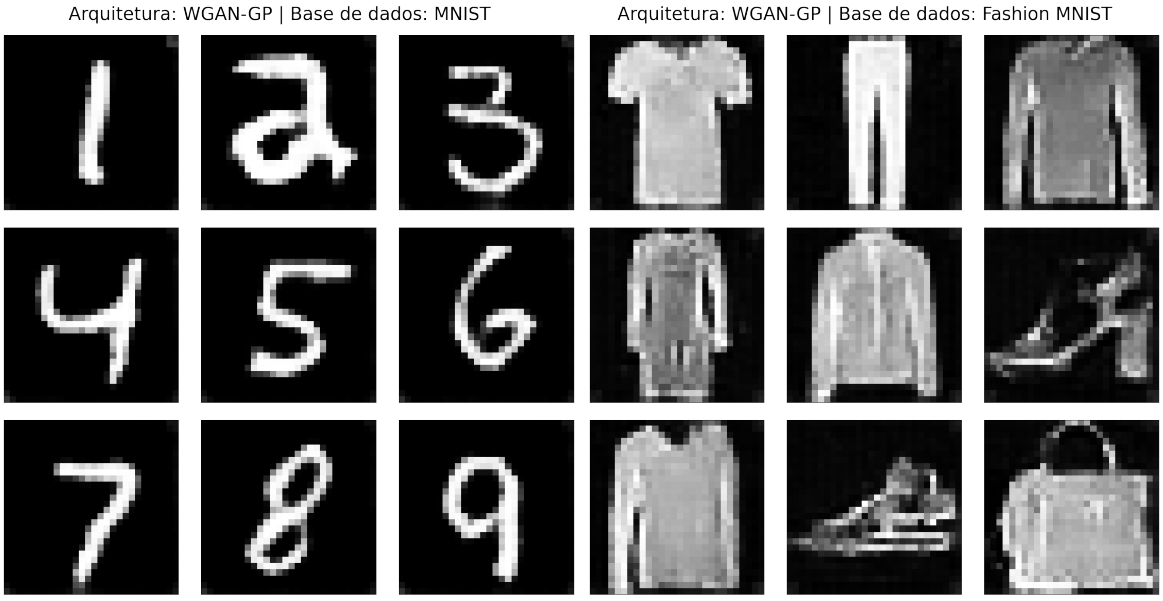
(a) Base de dados MNIST. (b) Base de dados Fashion MNIST.



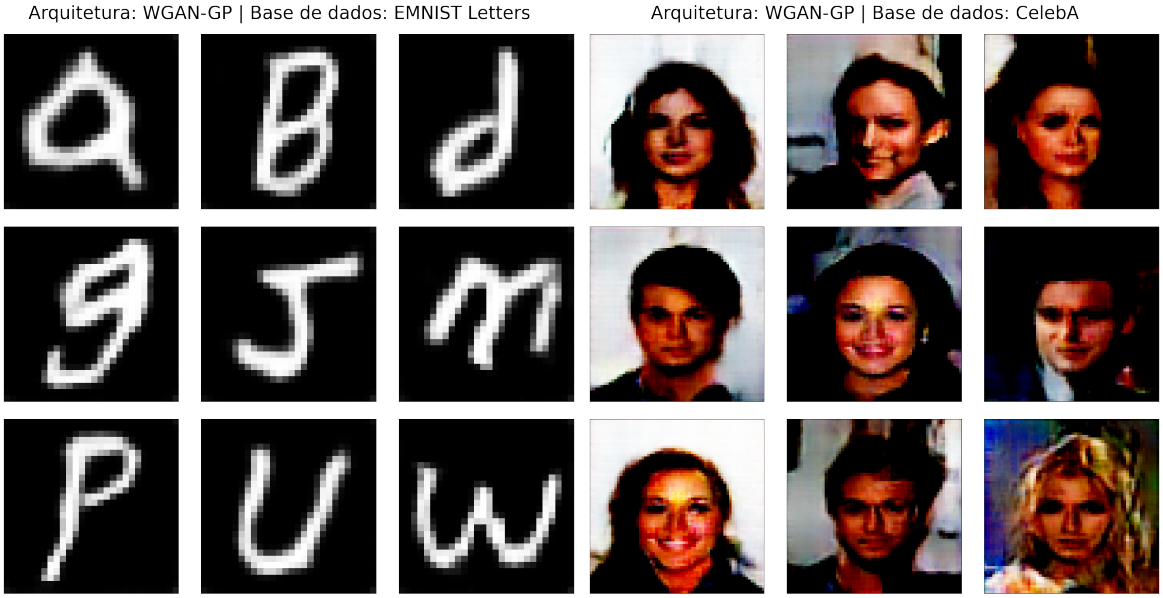
(c) Base de dados EMNIST Letters. (d) Base de dados CelebA.

Figura 10 – Imagens geradas pela arquitetura DCGAN treinada com as bases de dados (a) MNIST, (b) Fashion MNIST, (c) EMNIST Letters e (d) CelebA.

Os resultados obtidos após o treinamento da arquitetura WGAN-GP nas bases de dados selecionadas foram considerados adequados e atenderam às expectativas do estudo. A utilização da perda de Wasserstein como função de custo, juntamente com a aplicação do gradiente penalizado (GP), tornou a WGAN-GP mais robusta em relação a problemas de convergência e instabilidade do treinamento. Em termos qualitativos, os resultados obtidos pela WGAN-GP superaram os resultados obtidos com a DCGAN em todas as bases de dados. O Fashion MNIST, por exemplo, apresentou uma melhoria significativa de performance. Na base de dados CelebA, a WGAN-GP foi capaz de gerar imagens com mais diversidade em relação aos resultados obtidos pela DCGAN, o que representa uma vantagem significativa para a técnica de aumento de dados proposta. A Figura 11 ilustra alguns exemplos de imagens geradas pela WGAN-GP.



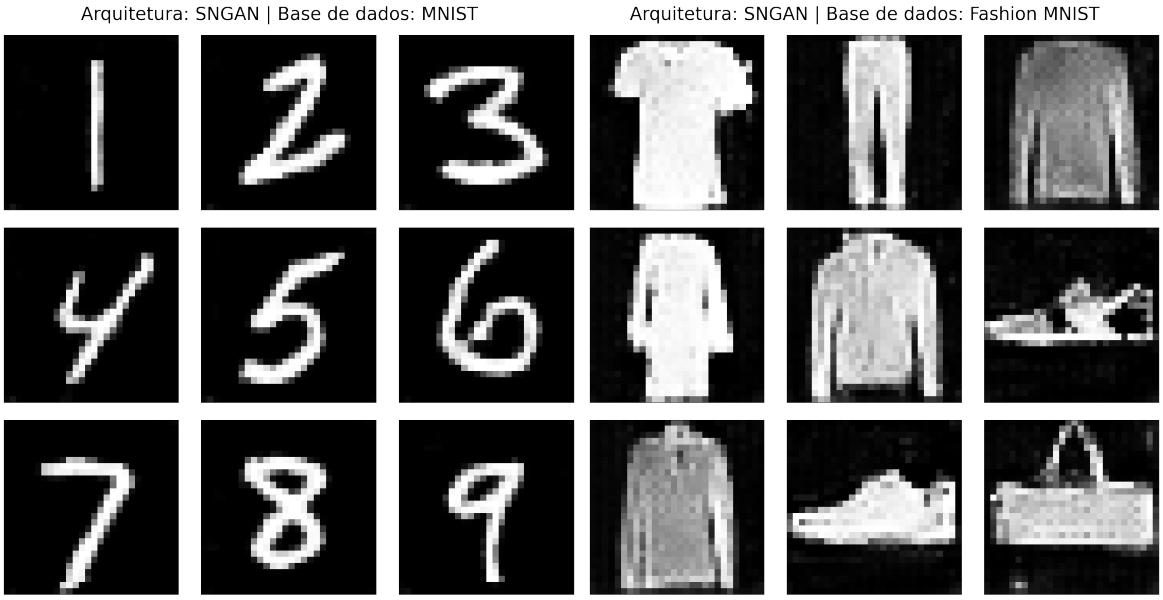
(a) Base de dados MNIST. (b) Base de dados Fashion MNIST.



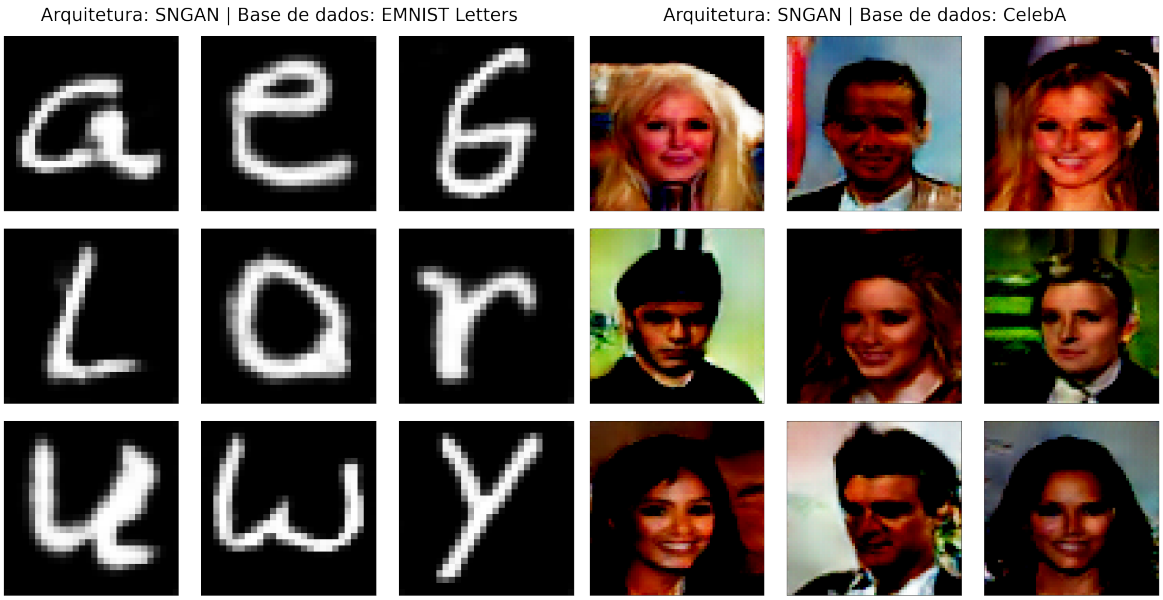
(c) Base de dados EMNIST Letters. (d) Base de dados CelebA.

Figura 11 – Imagens geradas pela arquitetura WGAN-GP treinada com as bases de dados (a) MNIST, (b) Fashion MNIST, (c) EMNIST Letters e (d) CelebA.

Os resultados obtidos com o treinamento da arquitetura SNGAN foram conforme o esperado. A aplicação da normalização espectral nas camadas de convolução contribuiu para um treinamento mais estável em todas as bases de dados, resultando em uma melhoria significativa das imagens geradas em relação às produzidas pela DCGAN. Além disso, em comparação com a arquitetura WGAN-GP, os resultados foram semelhantes em todas as bases de dados analisadas, sugerindo que ambas as arquiteturas são capazes de melhorar significativamente os resultados obtidos pela DCGAN. A Figura 12 ilustra alguns exemplos de imagens geradas pela SNGAN.



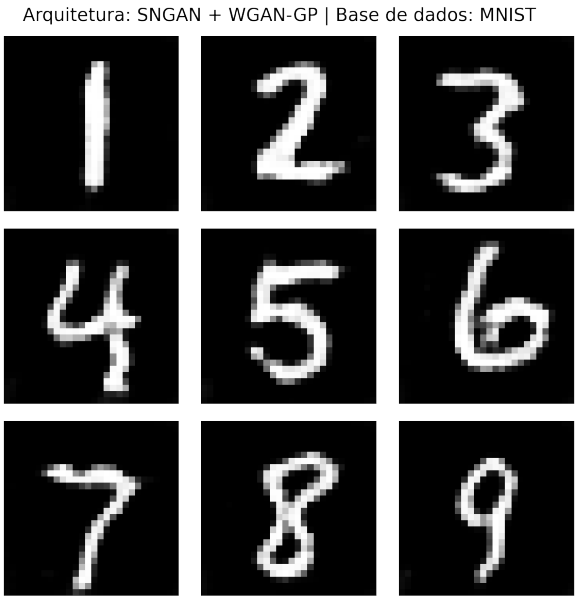
(a) Base de dados MNIST. (b) Base de dados Fashion MNIST.



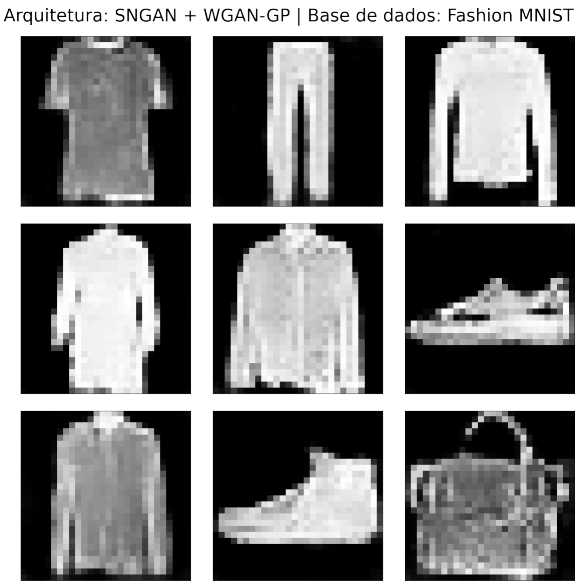
(c) Base de dados EMNIST Letters. (d) Base de dados CelebA.

Figura 12 – Imagens geradas pela arquitetura SNGAN treinada com as bases de dados (a) MNIST, (b) Fashion MNIST, (c) EMNIST Letters e (d) CelebA.

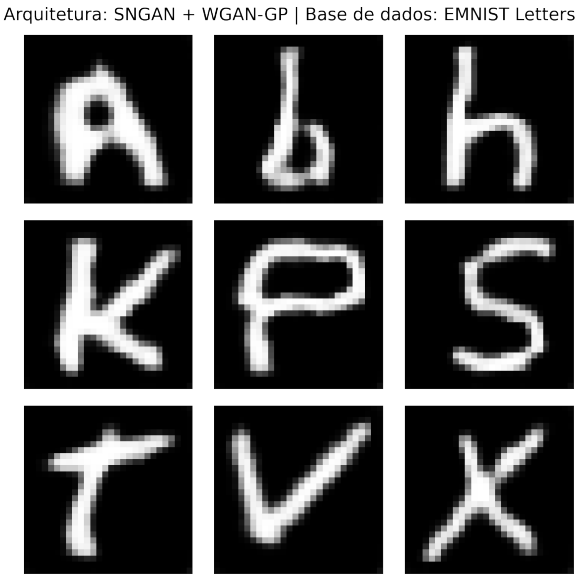
A combinação da arquitetura SNGAN com a WGAN-GP foi um experimento bem-sucedido no treinamento de bases de dados de imagens para o aumento de dados com imagens sintéticas. Essa união de técnicas permitiu obter resultados ainda mais promissores do que quando aplicadas individualmente. A utilização dessas técnicas de composição de arquitetura de GAN permitiu que o treinamento ocorresse de forma estável, proporcionando melhorias na qualidade e diversidade das imagens geradas. Como resultado, o treinamento dessa combinação de técnicas obteve os melhores resultados em termos de fidelidade e diversidade. A Figura 13 ilustra alguns exemplos de imagens geradas pela WGAN-GP + SNGAN.



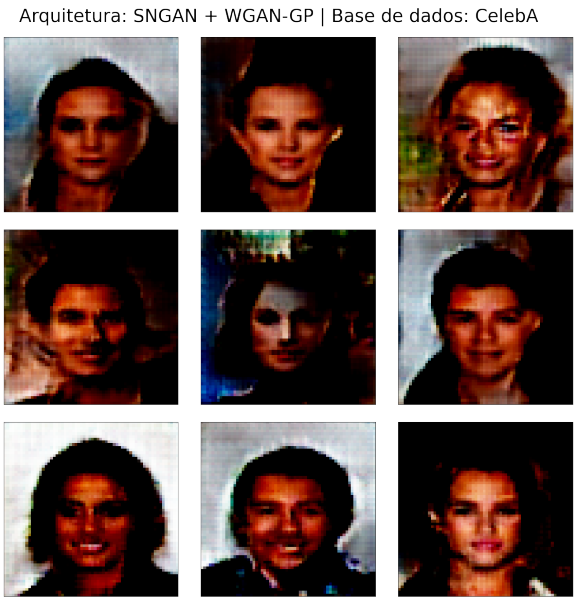
(a) Base de dados MNIST.



(b) Base de dados Fashion MNIST.



(c) Base de dados EMNIST Letters.



(d) Base de dados CelebA.

Figura 13 – Imagens geradas pela arquitetura WGAN-GP + SNGAN treinada com as bases de dados (a) MNIST, (b) Fashion MNIST, (c) EMNIST Letters e (d) CelebA.

Cronograma

Para alcançar os objetivos do projeto, o desenvolvimento do trabalho originalmente foi organizado nas seguintes etapas:

1. Estudo das técnicas avançadas de algoritmos de DL aplicado a sistemas generativos de imagens, com ênfase em GANs.
2. Estudo das técnicas existentes que possibilitem realizar o aumento de dados de forma artificial utilizando GANs.
3. Levantamento das principais bases públicas de imagens e obtenção das mais adequadas para o projeto.
4. Implementação e construção dos modelos de GANs com base nas técnicas pesquisadas na etapa 2, utilizando a linguagem de programação *Python* e *framework PyTorch*.
5. Treinamento dos modelos implementados da etapa 4 utilizando as bases de imagens pesquisadas na etapa 3.
6. Elaboração do relatório parcial e apresentação de seminário.
7. Avaliação da abordagem implementada através da realização de testes qualitativos e quantitativos com os modelos treinados na Etapa 5.
8. Redação de relatórios e de artigo para submissão a congressos e periódicos, o que será feito através do registro cotidiano de todas as atividades desenvolvidas ao longo do projeto.

Uma vez que o projeto tem vigência de setembro de 2022 a agosto de 2023, o cronograma apresentado na Tabela 1, distribui as atividades concluídas e as que estão em curso no período mencionado.

	2022				2023							
	SET	OUT	NOV	DEZ	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO
1												
2												
3												
4												
5												
6												
7												
8												

Tabela 1 – Cronograma das atividades a serem realizadas no projeto. Elementos em cinza representam tarefas em curso e elementos em verde representam tarefas concluídas.

Referências

ALZUBAIDI, L. et al. Towards a better understanding of transfer learning for medical imaging: a case study. *Applied Sciences*, Multidisciplinary Digital Publishing Institute, v. 10, n. 13, p. 4523, 2020.

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

BORJI, A. et al. Pros: A dataset for evaluating deep learning models on protein structures. *Bioinformatics*, Oxford University Press, v. 35, n. 11, p. 1982–1989, 2019.

BROCK, A.; DONAHUE, J.; SIMONYAN, K. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

CAO, S.; NEVATIA, R. Exploring deep learning based solutions in fine grained activity recognition in the wild. In: IEEE. *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.], 2016. p. 384–389.

CHEN, X.; YUILLE, A. L. Articulated pose estimation by a graphical model with image dependent pairwise relations. *Advances in neural information processing systems*, v. 27, 2014.

COHEN, G. et al. Emnist: an extension of mnist to handwritten letters. *arXiv preprint arXiv:1702.05373*, 2017.

CRESWELL, A. et al. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, IEEE, v. 35, n. 1, p. 53–65, 2018.

DIBA, A. et al. Weakly supervised cascaded convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 914–922.

DOULAMIS, N. Adaptable deep learning structures for object labeling/tracking under dynamic visual environments. *Multimedia Tools and Applications*, Springer, v. 77, n. 8, p. 9651–9689, 2018.

DOULAMIS, N.; VOULODIMOS, A. Fast-mdl: Fast adaptive supervised training of multi-layered deep learning models for consistent object tracking and classification. In: IEEE. *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*. [S.l.], 2016. p. 318–323.

FAN, R. et al. Computer stereo vision for autonomous driving. *arXiv preprint arXiv:2012.03194*, 2020.

GEUS, D. D.; MELETIS, P.; DUBBELMAN, G. Panoptic segmentation with a joint semantic and instance segmentation network. *arXiv preprint arXiv:1809.02110*, 2018.

GOODFELLOW, I. et al. Generative adversarial nets. *Advances in neural information processing systems*, v. 27, 2014.

GOOGLE. *Google Colaboratory*. 2023. <https://colab.research.google.com>. Cloud-based development environment for training neural networks and other machine learning models.

GULRAJANI, I. et al. Improved training of wasserstein gans. *Advances in Neural Information Processing Systems*, v. 30, p. 5767–5777, 2017.

HASAN, M. K. et al. Challenges of deep learning methods for covid-19 detection using public datasets. *Informatics in Medicine Unlocked*, Elsevier, v. 30, p. 100945, 2022.

HEUSEL, M. et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, v. 30, p. 6626–6637, 2017.

ISOLA, P. et al. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 1125–1134.

KARRAS, T. et al. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

KARRAS, T.; LAINE, S.; AILA, T. A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. [S.l.: s.n.], 2019. p. 4401–4410.

KINGMA, D. P.; WELLING, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

KLETTE, R. *Concise computer vision*. [S.l.]: Springer, 2014. v. 1.

LECUN, Y. *RL seminar: The next frontier in AI: Unsupervised learning*. [S.l.]: Technical Report, 2016.

LECUN, Y. et al. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, v. 2, 1989.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LIN, L. et al. A deep structured model with radius–margin bound for 3d human activity recognition. *International Journal of Computer Vision*, Springer, v. 118, n. 2, p. 256–273, 2016.

LIN, Z.; SEKAR, V.; FANTI, G. Why spectral normalization stabilizes gans: Analysis and improvements. *Advances in neural information processing systems*, v. 34, p. 9625–9638, 2021.

LIU, M.-Y.; TUZEL, O. Coupled generative adversarial networks. *Advances in neural information processing systems*, v. 29, 2016.

LIU, Z. et al. Deep learning face attributes in the wild. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2015. p. 3730–3738.

MAKRIDAKIS, S. The forthcoming artificial intelligence (ai) revolution: Its impact on society and firms. *Futures*, Elsevier, v. 90, p. 46–60, 2017.

MIRZA, M.; OSINDERO, S. Conditional generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680.

MIYATO, T. et al. Spectral normalization for generative adversarial networks. In: *International Conference on Learning Representations*. [s.n.], 2018. Disponível em: <https://openreview.net/forum?id=B1QRgziT->.

NANDY, A.; DUAN, C.; KULIK, H. J. Audacity of huge: overcoming challenges of data scarcity and data quality for machine learning in computational materials discovery. *Current Opinion in Chemical Engineering*, Elsevier, v. 36, p. 100778, 2022.

OUYANG, W. et al. Deepid-net: Object detection with deformable part based convolutional neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 39, n. 7, p. 1320–1334, 2016.

PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2009.

PASZKE, A. et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019.

PAWARA, P. et al. Data augmentation for plant classification. In: SPRINGER. *International conference on advanced concepts for intelligent vision systems*. [S.l.], 2017. p. 615–626.

POUYANFAR, S. et al. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 51, n. 5, p. 1–36, 2018.

RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

RUIZ-SANTAQUITERIA, J. et al. Semantic versus instance segmentation in microscopic algae detection. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 87, p. 103271, 2020.

SALEHI, P.; CHALECHALE, A.; TAGHIZADEH, M. Generative adversarial networks (gans): An overview of theoretical model, evaluation metrics, and recent developments. *arXiv preprint arXiv:2005.13178*, 2020.

SALIMANS, T. et al. Improved techniques for training gans. *Advances in Neural Information Processing Systems*, v. 29, p. 2234–2242, 2016.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of big data*, Springer, v. 6, n. 1, p. 1–48, 2019.

SONKA, M.; HLAVAC, V.; BOYLE, R. *Image processing, analysis, and machine vision*. [S.l.]: Cengage Learning, 2014.

TANAKA, F. H. K. d. S.; ARANHA, C. Data augmentation using gans. *arXiv preprint arXiv:1904.09135*, 2019.

TOSHEV, A.; SZEGEDY, C. Deeppose: Human pose estimation via deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 1653–1660.

TURHAN, C. G.; BILGE, H. S. Recent trends in deep generative models: a review. In: IEEE. *2018 3rd International Conference on Computer Science and Engineering (UBMK)*. [S.l.], 2018. p. 574–579.

WANG, X. et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2097–2106.

XIAO, H.; RASUL, K.; VOLLGRAF, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

ZHANG, Q. et al. A survey on deep learning for big data. *Information Fusion*, Elsevier, v. 42, p. 146–157, 2018.

ZHOU, S.; ZELIKMAN, E. *Generative Adversarial Networks (GANs) Specialization*. 2019. Acesso em: 20 de março de 2023. Disponível em: <https://www.coursera.org/specializations/generative-adversarial-networks-gans>.

ZHU, J.-Y. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2223–2232.