

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Redes Adversárias Generativas na Moda: Um Comparativo de Modelos na Geração de Estampas

Deyvid Fernandes da Silva

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Deyvid Fernandes da Silva

Redes Adversárias Generativas na Moda: Um Comparativo de Modelos na Geração de Estampas

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientadora: Profa. Dra. Roseli Aparecida Francelin Romero

Versão original

São Carlos

2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

S586r Silva, Deyvid Fernandes da
Redes Adversárias Generativas na Moda: Um
Comparativo de Modelos na Geração de Estampas /
Deyvid Fernandes da Silva; orientador Profa. Dra.
Roseli Aparecida Francelin Romero. -- São Carlos,
2024.
50 p.

Trabalho de conclusão de curso (MBA em
Inteligência Artificial e Big Data) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2024.

1. Generative Adversarial Networks (GAN). 2.
Redes Adversárias Generativas. 3. GANs na moda. 4.
CycleGAN. 5. Pix2Pix. I. Romero, Profa. Dra. Roseli
Aparecida Francelin , orient. II. Título.

Deyvid Fernandes da Silva

Generative Adversarial Networks in Fashion: A Comparative Study of Models in Pattern Generation

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Profa. Dra. Roseli Aparecida
Francelin Romero

Original version

São Carlos

2024

Dedico este trabalho à minha esposa e ao meu filho recém-nascido, Noah, que me deram todo o suporte e amor necessários para que eu pudesse me adaptar à nova rotina e, ainda assim, seguir com os estudos e o desenvolvimento desta obra.

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos à Profa. Dra. Roseli Aparecida Francelin Romero pela orientação e apoio durante toda a pesquisa. Agradeço também aos professores, mentores e colegas de curso pelo companheirismo e pela troca valiosa de conhecimentos.

RESUMO

SILVA, D.F. **Redes Adversárias Generativas na Moda: Um Comparativo de Modelos na Geração de Estampas**. 2024. 50 p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Este trabalho investiga a aplicação de Redes Adversárias Generativas (GANs) na criação de novas estampas para a indústria da moda. O objetivo é explorar o potencial dessas redes para gerar estampas inovadoras e contribuir para o design de roupas. O estudo compara três modelos de GANs: Pix2Pix, CycleGAN e BicycleGAN, avaliando-os quanto à qualidade e realismo das imagens geradas. Os resultados mostram que o modelo CycleGAN se destaca na geração de estampas mais realistas e consistentes, superando os modelos Pix2Pix e BicycleGAN. Além disso, o estudo avalia a influência do número de iterações no treinamento dos modelos, concluindo que o desempenho não está diretamente relacionado ao número de iterações, mas sim a outros fatores, como a arquitetura da rede e o conjunto de dados utilizado. O trabalho enfatiza a importância de continuar explorando a aplicação de GANs na indústria da moda, focando em aprimorar a qualidade dos modelos e expandir o conjunto de dados utilizado para treinamento.

Palavras-chave: Redes Adversárias Generativas (GANs); Pix2Pix; CycleGAN; BicycleGAN; Estampas; Moda; Roupas; GANs na moda;

ABSTRACT

SILVA, D.F. **Generative Adversarial Networks in Fashion: A Comparative Study of Models in Pattern Generation.** 2024. 50 p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

This work investigates the application of Generative Adversarial Networks (GANs) in creating new patterns for the fashion industry. The goal is to explore the potential of these networks to generate innovative patterns and contribute to clothing design. The study compares three GAN models: Pix2Pix, CycleGAN, and BicycleGAN, evaluating them based on the quality and realism of the generated images. The results show that the CycleGAN model excels in generating more realistic and consistent patterns, outperforming the Pix2Pix and BicycleGAN models. Additionally, the study assesses the influence of the number of iterations in model training, concluding that performance is not directly related to the number of iterations but rather to other factors, such as network architecture and the dataset used. The work emphasizes the importance of continuing to explore the application of GANs in the fashion industry, focusing on improving model quality and expanding the dataset used for training.

Keywords: Generative Adversarial Networks (GANs); Pix2Pix; CycleGAN; BicycleGAN; Patterns; Fashion; Clothing; GANs in fashion;

LISTA DE FIGURAS

Figura 1 – Imagem que ilustrativa da arquitetura do primeiro modelo GAN. Fonte: Adaptado de (Briot, 2021)	26
Figura 2 – Evolução dos modelos GANs, desde a sua introdução em 2014. Fonte: (Goodfellow <i>et al.</i> , 2020)	27
Figura 3 – Imagem que demonstra a arquitetura de uma cGAN. Fonte: Adaptado de (Mino; Spanakis, 2018) e (Briot, 2021)	28
Figura 4 – A imagem Photograph é fornecida como entrada e as demais imagens são a exemplificação do funcionamento e resultado de saída do modelo CycleGAN. A segunda figura, nomeada Monet, é a tradução da imagem de entrada nas características das pinturas de Monet. O mesmo comportamento ocorre com as figuras nomeadas Van Gogh, Cezanne e Ukiyo-e. Fonte: (Zhu <i>et al.</i> , 2017b)	29
Figura 5 – (a) Imagem noturna utilizada como entrada e (b) várias imagens de saída geradas pelo modelo BicycleGAN. Fonte: (Zhu <i>et al.</i> , 2017c) . . .	30
Figura 6 – A estrutura da rede neural utilizado para gerar sugestão de várias imagens de logotipos. Fonte: (Tian <i>et al.</i> , 2021)	31
Figura 7 – A imagem representa um pedaço da cidade de Munique, na Alemanha. E a área em vermelho é a de pegada predial e foi gerado por um modelo GAN. Fonte: (Shi; Li; Zhu, 2019)	31
Figura 8 – Alguns exemplos de imagens de roupas criadas a partir de elementos de design de roupas, como esboços de moda e padrões de tecido, geradas pelo FashionGAN. Fonte: (Cui <i>et al.</i> , 2018)	33
Figura 9 – A imagem apresentada ilustra a solução de uma forma macro.	36
Figura 10 – Exemplo de imagem de roupa superior, extraído do dataset Fashion Product Images. Fonte: Aggarwal (2019)	40
Figura 11 – Exemplo de imagem estampada. Fonte: istampa (2024)	40
Figura 12 – As imagens da esquerda e do centro fazem parte do dataset Pix2Pix, e a imagem à direita foi gerada pelo modelo treinado.	43
Figura 13 – O modelo treinado utilizou as imagens da esquerda e do centro, extraídas do dataset BicycleGAN, para gerar a imagem à direita.	43
Figura 14 – A imagem à direita foi gerada pelo modelo treinado com o dataset CycleGAN.	44

LISTA DE TABELAS

Tabela 1 – Comparação dos recursos dos ambientes de execução por tipo de máquina, fonte: GCP (2024)	35
Tabela 2 – Imagens geradas pelos modelos treinados.	45
Tabela 3 – Comparação das métricas geradas das imagens de cada modelo.	46

LISTA DE ABREVIATURAS E SIGLAS

MVP	Minimum Viable Product
GAN	Generative Adversarial Nets
cGAN	Conditional Generative Adversarial Nets
CNN	Convolutional Neural Network
ML	Machine Learning
USP	Universidade de São Paulo
USPSC	Campus USP de São Carlos
GCP	Google Cloud Platform

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Motivação	23
1.2	Objetivos	24
1.3	Objetivos específicos	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Modelos generativos	25
2.2	Tipos de GANs	25
2.2.1	GAN	26
2.2.2	cGAN	27
2.2.3	Pix2Pix	28
2.2.4	CycleGAN	29
2.2.5	BicycleGAN	29
2.3	Aplicações das GANs não relacionadas a moda	30
2.4	Aplicações de redes GANs no setor de MODA	32
3	METODOLOGIA	35
3.1	Infraestrutura	35
3.2	Fluxograma	35
3.3	Métricas de avaliação	36
3.3.1	Avaliação humana	36
3.3.2	Inception score	37
4	DESENVOLVIMENTO	39
4.1	Dataset	39
4.1.1	Imagens de roupas superiores	39
4.1.2	Imagens de estampas	40
4.1.3	Tratamento das imagens	41
4.2	Código-fonte	41
5	ANÁLISE DE RESULTADOS	43
6	CONCLUSÕES	47
	REFERÊNCIAS	49

1 INTRODUÇÃO

Ao contrário da indústria do desenvolvimento de software - onde engenheiros e gerentes de tecnologia podem entrevistar os usuários, definir requisitos, implementar e validar um *Minimum Viable Product*, também conhecido como MVP ou Produto Mínimo Viável em português, antes de investir tempo e dinheiro na produção. A indústria da moda e o processo de criação de uma nova peça de roupa dependem da criatividade do seu estilista. O estilista por sua vez precisa investir um bom tempo de estudo e aprimoramento de suas técnicas de construção, estar sintonizado com as tendências, antecipar os desejos dos consumidores e ser o mais certo possível, porque se algo der errado, dificilmente será possível realizar pequenos ajustes e relançar o mesmo produto. Além disso, com o avanço tecnológico, é necessário dominar os softwares de última geração para criação do protótipo. Em Cui *et al.* (2018), é mencionada que a exibição virtual de roupas desempenha um papel fundamental no design de moda, criando uma amostra digital da roupa, sem a necessidade de confecção, como na indústria de roupas tradicional, com o claro objetivo de mitigar os riscos de uma falha no lançamento.

No final de 2023, uma grande rede de lojas de departamento brasileira teve que retirar roupas do catálogo após muitas críticas nas redes sociais (BOLZANI, 2023), já que algumas peças lembravam o uniforme utilizado durante o período do Holocausto. O modelo produzido provavelmente foi resultado de um descuido ou infelicidade durante o processo de criação (BOLZANI, 2023). Fatos como esse trazem consequências para a empresa, que vão desde uma imagem prejudicada até a perda de engajamento de parte dos consumidores, sem falar no prejuízo financeiro ao ter que descartar todo o estoque de uma coleção recém-lançada. E a inteligência artificial generativa é uma ferramenta que pode ser utilizada para reduzir os riscos de repetir cenários como o mencionado.

De acordo com (Liu *et al.*, 2020), os seres humanos podem utilizar as roupas para expressar sua personalidade ou sentimento, além da necessidade de vestir roupas adequadas no seu dia a dia ou em ocasiões especiais. Dessa forma, alguns estilistas buscam refletir o momento atual em suas peças e estampas, usando referências do cotidiano para criar uma conexão emocional com o consumidor. A inspiração pode vir de um livro, uma obra de arte ou até mesmo de eventos, como o isolamento social. Enfim, há um vasto oceano de possibilidades.

1.1 Motivação

Como podemos ver, os desafios enfrentados por um estilista durante o planejamento e design de uma nova peça de roupa vão além de simplesmente pegar um lápis e desenhar em um papel ou usar softwares modernos para a criação. O processo exige pesquisa, estudos

aprofundados e uma monitoração constante de revistas especializadas, desejos e feedback dos consumidores.

Dado que qualquer pessoa qualificada e que domine essas técnicas pode criar uma estampa ou uma peça de roupa, a IA generativa pode ser uma aliada na produção de novos modelos, agilizando e acelerando o processo, para estilistas ou indivíduos que desejam criar um novo modelo de roupa.

Com o intuito de gerar novas representações de roupas superiores e estampadas, este estudo se baseia na comparação de várias arquiteturas de redes adversárias generativas, as conhecidas GANs, visando aprimorar o treinamento de um modelo capaz de desenhar novas peças de maneira criativa e inovadora. Os modelos GANs são constituídos por dois módulos: o módulo “gerador” e o módulo “discriminador” (Goodfellow *et al.*, 2014). Os modelos constituídos de GANs a serem investigados contemplam diferentes arquiteturas de redes CNNs, combinação de CNN com outro tipo de rede de aprendizado profundo e também variação de filtros para o módulo “gerador”. Em outras palavras, o objetivo principal desse trabalho é treinar um modelo generativo baseado em aprendizagem profunda, com um módulo gerador capaz de produzir novas peças de roupas superiores reais e perfeitas ao ponto do módulo discriminador não ser capaz de identificar que é uma imagem que não existe. E então, a imagem gerada poderá utilizada de inspiração ou como insumo para qualquer estilista ou pessoa que queira produzir uma nova peça de roupa de uma forma ágil e democrática.

1.2 Objetivos

Neste trabalho, será investigado o uso de modelos GANs no setor da Moda. Para isso, serão utilizados alguns modelos de redes GANs para geração de estampas de tecidos. Esses modelos serão comparados visando vantagens e desvantagens de cada modelo.

1.3 Objetivos específicos

Durante o desenvolvimento desta pesquisa, as seguintes hipóteses são levantadas:

- Os modelos de redes generativas adversárias produzem imagens mais interessantes, para o setor da moda, no início do treinamento, se comparados com modelos treinados com um número muito maior de épocas.
- A performance do modelo depende da arquitetura proposta e também do tipo de filtro usado no módulo gerador.

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo, são descritos os conceitos envolvidos nas redes GANs e alguns dos principais tipos de modelos propostos. Em seguida, são apresentados algumas aplicações de redes GANs, dentro e fora do setor da Moda.

2.1 Modelos generativos

Segundo (Creswell *et al.*, 2018), os modelos generativos são treinados e aprendem a capturar a distribuição estatística de dados de treinamento, permitindo criar novas amostras com base nas características e padrões da distribuição aprendida. Essas novas amostras podem ser utilizadas em tarefas como edição semântica de imagens, aumento de dados, transferência de estilos, recuperação de imagens e para muitas outras tarefas. Em outras palavras, os modelos generativos são uma vertente da área de inteligência artificial que aprendem a criar novos dados similares aos dados que foi submetido durante seu treinamento. Fazendo uma analogia, são como artistas que após observar vários exemplos de pinturas, conseguem reproduzir novas pinturas com os mesmos estilos.

Alguns exemplos de IAs generativas são o Gemini do Google, Chat GPT e DALL-E da Open AI, que geram textos e imagens com base nos dados do treinamento. O DALL-E utiliza uma síntese chamada text-to-image que tem sido uma área ativa de pesquisa desde a publicação do trabalho de (Reed *et al.*, 2016), na qual a abordagem é utilizar uma GAN condicional para geração de imagens com base em um texto de entrada.

Em uma breve analogia feita por (Goodfellow *et al.*, 2020) para explicar o conceito de uma GAN, podemos pensar como os falsificadores e a polícia: os falsificadores ganham dinheiro falso enquanto a polícia tenta prender os falsificadores e continua a permitir o gasto de dinheiro legítimo. A competição entre falsificadores e a polícia leva a dinheiro falsificado cada vez mais realista, até que eventualmente os falsificadores produzem falsificações perfeitas e a polícia não consegue distinguir entre dinheiro verdadeiro e falso. Na próxima sessão, iremos explorar um pouco mais os tipos de GANs.

2.2 Tipos de GANs

Atualmente existem muitos modelos GANs, e a seguir apresentaremos apenas alguns usados nos principais artigos referenciados neste trabalho. Ou seja, a lista é muito mais abrangente e não detalharemos todos os tipos aqui.

2.2.1 GAN

Em 2014 (Goodfellow *et al.*, 2014), propôs pela primeira vez as redes adversárias generativas, um tipo de algoritmo de inteligência artificial, que consiste em duas redes neurais: um modelo gerador G que toma variáveis de ruído como entrada para gerar novas instâncias de dados, enquanto um modelo discriminador D , interpreta e decide se cada instância de dados pertence ou não ao conjunto de dados de treinamento real. E esse é um algoritmo de machine learning que nos últimos anos, obteve grande sucesso na tradução de imagem para imagem, de acordo com (Tian *et al.*, 2021).

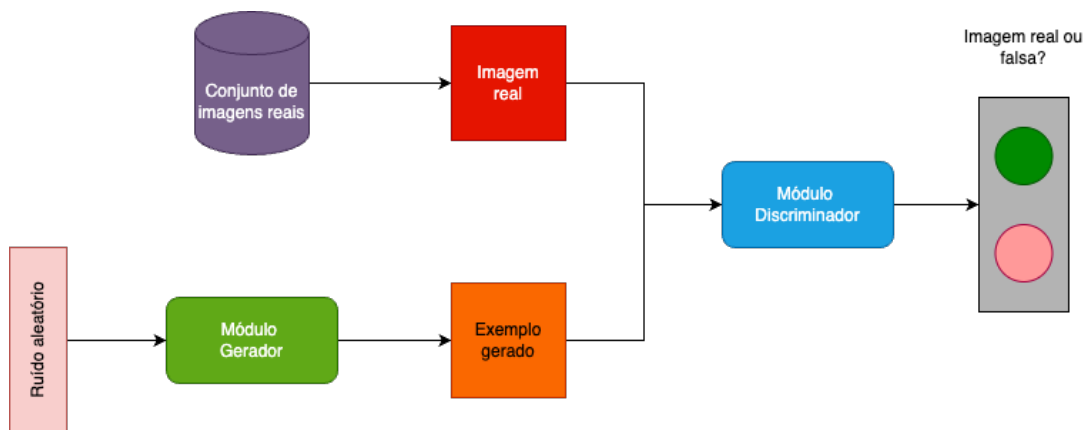


Figura 1 – Imagem que ilustrativa da arquitetura do primeiro modelo GAN. Fonte: Adaptado de (Briot, 2021)

As GANs foram aplicadas com sucesso a uma ampla variedade de tarefas (principalmente em ambientes de pesquisa), mas continuam a apresentar desafios e oportunidades de pesquisa únicas porque são baseadas na teoria dos jogos, enquanto a maioria das outras abordagens à modelagem generativa são baseadas na otimização (Goodfellow *et al.*, 2020). As redes adversárias generativas vem se tornando um dos algoritmos mais populares em Deep Learning, segundo (Tian *et al.*, 2021). A Figura 1 ilustra a arquitetura de um modelo GAN.

É difícil fornecer orientações específicas adicionais sobre os detalhes das GANs porque as GANs são uma área de pesquisa muito ativa e a maioria dos conselhos específicos ficam rapidamente desatualizados (Goodfellow *et al.*, 2020). A Figura 2 mostra a rapidez com que as capacidades das GANs progrediram nos anos desde a sua introdução.



Figura 2 – Evolução dos modelos GANs, desde a sua introdução em 2014. Fonte: (Goodfellow *et al.*, 2020)

2.2.2 cGAN

Como mencionado por (Tian *et al.*, 2021), a principal desvantagem do GAN clássico, proposto por I Goodfellow, é que a saída é incontrolável. Ou seja, após o treinamento do modelo, não conseguimos manipular ou direcionar o resultado de saída. Dado essa limitação, (Mirza; Osindero, 2014). propuseram as Redes Adversariais Gerativas Condicionais (CGAN) para otimizar o GAN clássico e mitigar essa restrição.

De acordo com (Mirza; Osindero, 2014), as redes adversárias gerativas podem ser estendidas para um modelo condicional se tanto o gerador quanto o discriminador estiverem condicionados a alguma informação extra. Que poderia ser qualquer tipo de informação auxiliar, como rótulos de classes ou dados de outras modalidades. Ainda segundo (Mirza; Osindero, 2014), em um modelo generativo não condicionado, não há controle sobre os modos de geração dos dados. Porém, ao condicionar o modelo a informações adicionais é possível direcionar o processo de geração de dados.

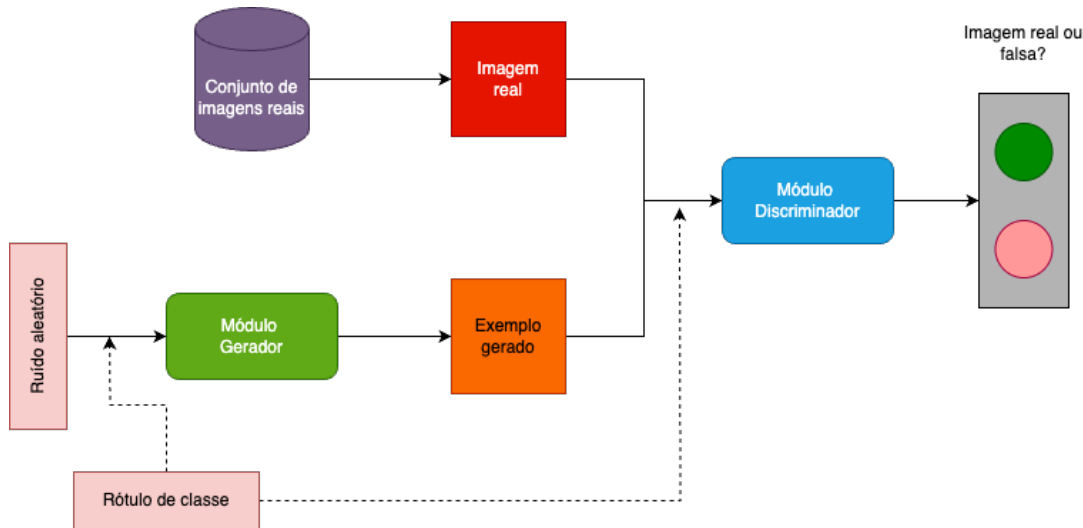


Figura 3 – Imagem que demonstra a arquitetura de uma cGAN. Fonte: Adaptado de (Mino; Spanakis, 2018) e (Briot, 2021)

Podemos ver na Figura 3 que a arquitetura é muito parecida com da GAN tradicional, ilustrada na Figura 1. Basicamente a diferença é a informação extra condicionada na entrada do módulo gerador e do módulo discriminador, no caso da Figura 3, chamado de rótulo de classe.

2.2.3 Pix2Pix

(Isola *et al.*, 2017) investigaram o uso das redes adversariais gerativas condicionais com proposito geral da criação de um framework capaz de resolver problemas de tradução de imagem para imagem. Pix2Pix GAN é uma implementação da cGAN que recebe de entrada um imagem como rótulo de classe e a utiliza para gerar o resultado de saída.

De acordo com (Isola *et al.*, 2017)

Assim como os GANs aprendem um modelo generativo de dados, os GANs condicionais (cGANs) aprendem um modelo generativo condicional. Isso torna os cGANs adequados para tarefas de tradução imagem a imagem, onde condicionamos uma imagem de entrada e geramos uma imagem de saída correspondente.

Ainda segundo (Isola *et al.*, 2017), assim como podemos expressar uma determinada frase em dois idiomas distintos, uma cena também pode ser renderizada de diferentes formas, como uma imagem RGB, ou com bordas, e até mesmo em preto e branco, etc. Em uma clara analogia à tradução de idiomas, a tradução de imagem para imagem tem como objetivo traduzir uma determinada cena em outra, com base nos dados de treinamento.

2.2.4 CycleGAN

O CycleGAN também é um modelo de tradução de imagem para imagem, assim como Pix2Pix, porém os pesquisadores propuseram uma abordagem um pouco diferente. Como mencionado por (Zhu *et al.*, 2017b), o maior desafio para treinar os modelos está no fato de que é necessário no mínimo pares de imagens correspondentes em cada domínio, ou seja, se temos uma imagem de um cavalo, deveríamos ter uma imagem correspondente para uma zebra, para fazermos a tradução de cavalo para zebra, por exemplo. Com isso, montar um conjunto de dados, com pares de imagens e que cubra todos os cenários, é uma tarefa muito difícil, o que tornando um limitante para o treinamento dos modelos.

Dado esse desafio, o CycleGAN busca aprender a tradução sem dados pareados. Em outras palavras, o modelo pode ser treinado com conjuntos de dados independentes, o que o torna mais versátil. Como mencionado por (Zhu *et al.*, 2017b), O modelo apresentado segue uma abordagem de aprender a traduzir uma imagem de um domínio de origem X para um domínio de destino Y, mesmo com a ausência de exemplos emparelhados no conjunto de dados. A Figura 4 ilustra o resultado de saída de um modelo CycleGAN.



Figura 4 – A imagem Photograph é fornecida como entrada e as demais imagens são a exemplificação do funcionamento e resultado de saída do modelo CycleGAN. A segunda figura, nomeada Monet, é a tradução da imagem de entrada nas características das pinturas de Monet. O mesmo comportamento ocorre com as figuras nomeadas Van Gogh, Cezanne e Ukiyo-e. Fonte: (Zhu *et al.*, 2017b)

2.2.5 BicycleGAN

Como mencionado por (Zhu *et al.*, 2017c), as técnicas de aprendizagem profunda evoluíram rapidamente na geração de imagens condicionais, porém a maioria dos modelos concentra-se na geração de apenas um resultado de saída e muitos acabam colapsando, que é o efeito de gerar a mesma imagem de saída para diferentes imagens de entrada.

Com base nesses desafios, (Zhu *et al.*, 2017c) propôs um modelo híbrido que combina duas abordagens cVAE-GAN e cLR-GAN. Esse modelo produz resultados visualmente atraentes e diversos. Como ilustrado na Figura 5.

De acordo com (Zhu *et al.*, 2017c)

cVAE-GAN (Conditional Variational Autoencoder GAN): Uma abordagem é primeiro codificar a imagem da verdade no espaço latente, dando ao gerador uma "espiada" ruidosa na saída desejada.

Usando isso, junto com a imagem de entrada, o gerador deve ser capaz de reconstruir a imagem de saída específica. Para garantir que a amostragem aleatória possa ser usada durante o tempo de inferência, a distribuição latente é regularizada usando a divergência KL para estar próxima de uma distribuição normal padrão.

Ainda de acordo com (Zhu *et al.*, 2017c)

cLR-GAN (Conditional Latent Regressor GAN): Outra abordagem é primeiro fornecer um vetor latente desenhado aleatoriamente ao gerador. Nesse caso, a saída produzida pode não necessariamente se parecer com o gerador (imagem verdadeira), mas deve parecer realista. Um codificador então tenta recuperar o vetor latente na imagem de saída.



Figura 5 – (a) Imagem noturna utilizada como entrada e (b) várias imagens de saída geradas pelo modelo BicycleGAN. Fonte: (Zhu *et al.*, 2017c)

2.3 Aplicações das GANs não relacionadas a moda

Uma aplicação notável de uma GAN e que não está relacionada a moda é no design de logotipos, como citado por (Tian *et al.*, 2021). O processo de design de logotipos é complexo até mesmo para designers, e a cor desempenha um papel muito importante. A colorização automática do esboço do logotipo é de grande valor e cheia de desafios. Ainda segundo (Tian *et al.*, 2021), as cores trazem diferentes sentimentos visuais e experiências psicológicas para as pessoas, o que realça a importância de ferramentas para auxiliar designers ou entusiastas da área na criação e colorização de logotipos.

Para a geração das imagens de logotipos, Tian *et al.* (Tian *et al.*, 2021) utilizaram uma imagem de esboço do logotipo como entrada alimentada diretamente no gerador. Ainda segundo (Tian *et al.*, 2021), no gerador foi adotado uma estrutura empilhada contendo vários blocos attention-based U-Net (Att-UNet), cada um dos quais possui a mesma estrutura de rede com diferentes parâmetros, sendo que cada rede gera uma imagem

de logotipo. Dessa forma, o gerador pode gerar várias imagens. A estrutura da rede neural está ilustrada na Figura 6.

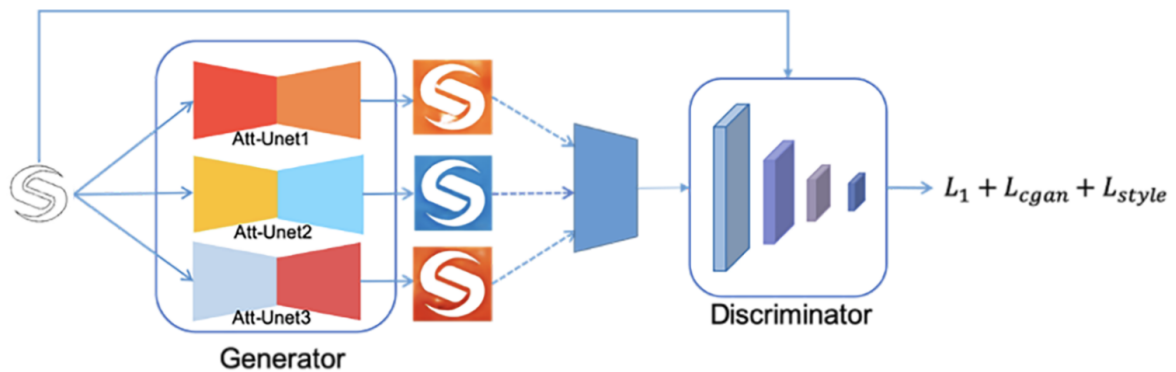


Figura 6 – A estrutura da rede neural utilizado para gerar sugestão de várias imagens de logotipos. Fonte: (Tian *et al.*, 2021)

Além disso, pesquisadores também utilizaram um modelo GAN para geração da pegada predial com base em imagens de satélite. Segundo (Shi; Li; Zhu, 2019). A geração da pegada predial é de grande importância para o planejamento e monitoramento urbano, análise do uso do solo e gestão de desastres. Na Figura 7, é mostrado o resultado gerado pela rede generativa adversária treinada para realizar o mapeamento.



Figura 7 – A imagem representa um pedaço da cidade de Munique, na Alemanha. E a área em vermelho é a de pegada predial e foi gerado por um modelo GAN. Fonte: (Shi; Li; Zhu, 2019)

No entanto, apesar do potencial transformador e inovador das GANs na área da moda, é importante reconhecer os desafios associados, como a garantia da originalidade dos designs gerados e a ética no uso de dados, evitando plágio e sem infringir os direitos

autorais. À medida que essas tecnologias seguem avançando e evoluindo, é esperado que medidas sejam tomadas para garantir esses requisitos, assim trilharemos um caminho para um futuro aonde a IA desempenha um papel central no design de moda.

2.4 Aplicações de redes GANs no setor de MODA

Nos últimos anos, a evolução das tecnologias de inteligência artificial tem revolucionado diversas indústrias e, os setores da moda não é exceção. Especificamente, Redes Geradoras Adversárias, ou em inglês *Generative Adversarial Networks* (GANs) e suas variantes condicionais, em inglês *Conditional Generative Adversarial Networks* (cGANs) emergiram como ferramentas poderosas para geração de novas imagens, com base em um conjunto de dados, oferecendo novas possibilidades para a criação de roupas ou combinação de roupas, a partir de um modelo treinado.

Um exemplo de aplicação das GANs no setor de moda é a geração automatizada de designs de roupas a partir de um rascunho. Em 2018, pesquisadores desenvolveram um sistema conhecido com FashionGAN, que utiliza as redes generativas adversárias para criar designs de roupas a partir de um rascunho e uma imagem de um tecido. Esse avanço permite que estilistas experimentem rapidamente com diferentes estilos, cores e padrões, acelerando o processo de criação e consequentemente economizando dinheiro.

Segundo (Cui *et al.*, 2018), a exibição virtual de roupas desempenha um papel importante no design de moda, pois pode mostrar diretamente o efeito do design da roupa sem a necessidade de fazer uma amostra de roupa, como na indústria de roupas tradicional. Os estudos demonstraram a capacidade e criatividade de um modelo de inteligência artificial treinado para gerar novas peças de roupas.

O modelo, conhecido como FashionGAN, foi treinado em um conjunto de dados com uma base de pouco mais de 22 mil imagens de roupas e por 160 épocas, alcançando resultados satisfatórios. Ele foi inspirado no BicycleGAN, sendo realizado algumas modificações. Além do modelo mencionado anteriormente, os pesquisadores utilizaram a mesma base e número de épocas, ou seja, as mesmas condições para treinar e comparar os resultados com outras arquiteturas como o próprio BicycleGAN.



Figura 8 – Alguns exemplos de imagens de roupas criadas a partir de elementos de design de roupas, como esboços de moda e padrões de tecido, geradas pelo FashionGAN. Fonte: (Cui *et al.*, 2018)

Apesar do resultado promissor, ilustrado na Figura 8, (Cui *et al.*, 2018) revelaram que o modelo não foi tão bem com tecidos estampados, tendo um aproveitamento melhor com tecidos lisos. Segundo os autores (Cui *et al.*, 2018), as prováveis razões são primeiramente que o codificador da rede aprendeu o mapeamento entre uma única cor ou combinação simples de cores, como, por exemplo, listras. Outro fator é que existem vários tipos de texturas irregulares, com combinações de cores complexas e dos mais variados formatos, o que dificultou o treinamento.

Além da geração de novas peças de roupas, outros pesquisadores desenvolveram um trabalho que combina peças superiores e inferiores, também a partir de um modelo de rede GAN treinado e um conjunto de imagens. Como mencionado por (Liu *et al.*, 2020), nós seres humanos temos a necessidade de vestir roupas adequadas no nosso dia a dia, seja para uma ocasião especial, encontro com amigos ou simplesmente para trabalhar, porém, nem todo mundo é especialista em moda ou possui um senso sofisticado de combinação de roupas.

A partir dessa necessidade, eles decidiram investir nesse estudo para ajudar a comunidade. Como citado por (Liu *et al.*, 2020), até o momento da publicação do artigo, e a partir dos resultados, o modelo ainda não conseguiu sintetizar bem as imagens da parte superior do vestuário e eles acreditam que esse fenômeno esteja atribuído aos padrões complicados e à rica diversidade de roupas superiores. Ainda segundo (Liu *et al.*, 2020), no futuro eles pretendem melhorar a qualidade da geração de agasalhos ampliando o conjunto de dados de treinamento e também pretendem expandir os estudos aumentando o domínio, também combinando sapatos e outros acessórios, além da parte superior e inferior já utilizada durante esse estudo. O que ressalta ainda mais o papel de destaque que as GANs vem ocupando na área da inteligência artificial com sua capacidade de gerar imagens.

O modelo proposto por (Liu *et al.*, 2020) possui uma arquitetura um pouco diferente das GANs tradicionais, pois é constituído por dois módulos geradores. Foi treinado por

200 épocas e reuniu um conjunto de dados de aproximadamente 19 mil imagens, que foram utilizadas para o treinamento e teste do modelo.

Diante do exposto, pretende-se implementar os seguintes modelos: Pix2Pix, CycleGAN e BicycleGAN. Uma análise comparativa da geração de novas estampas será realizada para determinar quais modelos são mais adequados para a tarefa proposta.

3 METODOLOGIA

Neste Capítulo, são descritos os detalhes da metodologia utilizada na concepção deste trabalho. Em seguida, são apresentadas as métricas de avaliação.

3.1 Infraestrutura

Para validarmos as hipóteses deste trabalho, implementaremos os modelos utilizando como infraestrutura o Google Colab Pro, disponível no Google Cloud Platform. O ambiente de execução escolhido foi o L4 GPU, por ter uma configuração superior ao T4 GPU e consumir menos unidades de processamento que o A100 GPU. Confira na Tabela 1 o comparativo dos recursos possíveis disponíveis para cada tipo de ambiente.

Tabela 1 – Comparação dos recursos dos ambientes de execução por tipo de máquina, fonte: GCP (2024)

Tipo de máquina	GPUs	Memória da GPU	vGPUs	Memória disponível
T4	4 GPUs	64 GB GDDR6	1 a 96 vCPUs	1 a 624 GB
L4	8 GPUs	192 GB GDDR6	96 vCPUs	384 GB
A100	16 GPUs	640 GB HBM2	96 vCPUs	1360 GB

3.2 Fluxograma

Iremos treinar cada modelo por três números de interações diferentes, com o objetivo de compararmos os resultados entre os modelos e também os resultados obtidos para o mesmo modelo com números de interações diferentes.

Após o treinamento, iremos avaliar os resultados por meio de métricas geradas por voluntários e uma métrica gerada automaticamente. Mais detalhes das métricas de avaliação são apresentados no próximo tópico. A Figura 9 é um diagrama de como funcionará o experimento.

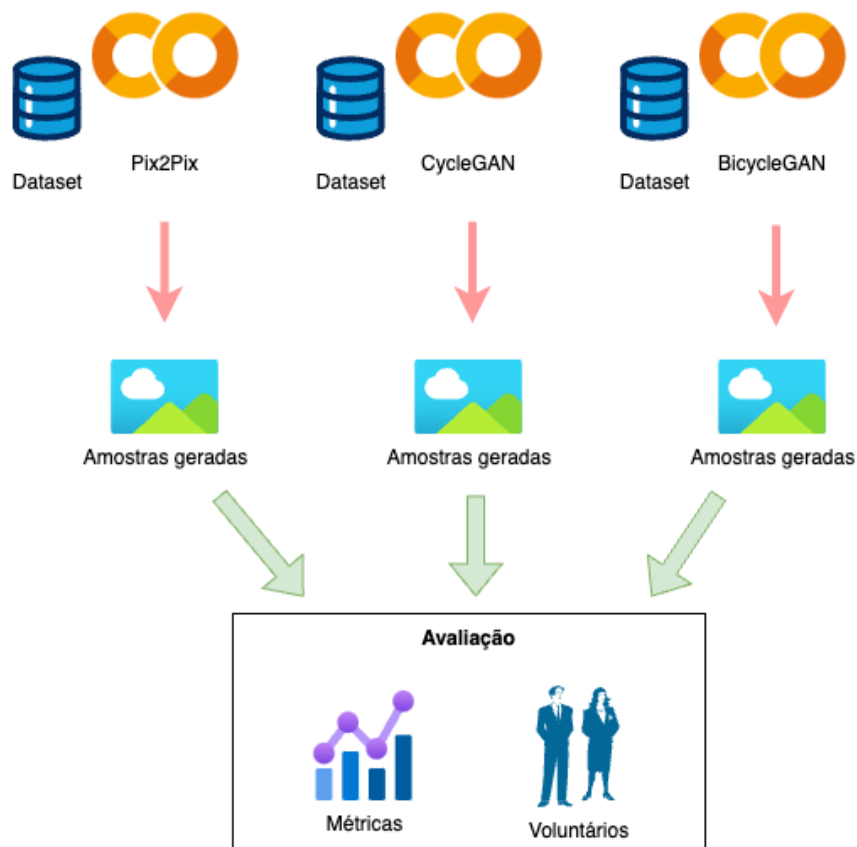


Figura 9 – A imagem apresentada ilustra a solução de uma forma macro.

3.3 Métricas de avaliação

A avaliação será baseada em dois modelos, os detalhes são apresentados na sequência.

3.3.1 Avaliação humana

Para avaliar os resultados mapeados da geração das imagens dos diferentes métodos, alguns voluntários irão receber 9 pares de imagens geradas por cada modelo, sendo 3 imagens para cada número de interação. Na sequência, irão avaliar as imagens com valores de 0 a 5, sendo 0 a nota mais baixa e 5 a mais alta. Os critérios de avaliação serão: Realismo e harmonia.

Com o objetivo de mantermos a parcialidade, o número mínimo de voluntários será 10.

No final, iremos calcular a média dessas pontuações.

3.3.2 Inception score

(Salimans *et al.*, 2016) propôs uma alternativa para a avaliação humana, conhecida como inception score. Um método automático para avaliar amostras geradas, que comparado com a avaliação humana obteve resultados mais assertivos.

A formula da métrica proposta é representada pela expressão:

$$\exp(\text{ExKL}(p(y|x) || p(y))) \quad (3.1)$$

onde x é uma amostra e $p(y|x)$ é a saída da camada softmax do classificador treinado. $p(y)$ é a distribuição geral dos rótulos das amostras geradas. O resultado é um valor entre 0 e ∞ , quanto maior a pontuação, melhor a qualidade e diversidade da imagem gerada, como mencionado por (Cui *et al.*, 2018).

Também utilizaremos essa métrica como forma de avaliar a qualidade das imagens geradas pelo modelo treinado.

4 DESENVOLVIMENTO

Neste Capítulo, são apresentados os métodos utilizados no desenvolvimento desta pesquisa, tais como: a criação do dataset apropriado para a aplicação almejada e aplicação de filtros nas imagens.

4.1 Dataset

Em relação à base de imagens para o treinamento, utilizaremos um dataset público e imagens obtidas da internet, sendo o dataset composto por imagens de roupas superiores e estampas.

No caso das imagens estampadas, aplicaremos um filtro para que a imagem tenha o formato de uma roupa superior. Pois os datasets dos modelos Pix2Pix e BicycleGan precisam de pares de imagem.

4.1.1 Imagens de roupas superiores

Utilizaremos o dataset Fashion Product Images (Aggarwal, 2019) como base de imagens de roupas superiores, disponível no site Kaggle. A base em questão também contém imagens de roupas inferiores e acessórios. Para o estudo proposto neste trabalho, iremos utilizar apenas as imagens de roupas superiores, como ilustrado na Figura 10. Ou seja, filtraremos por *Shirts* e *Tshirts* ou camisas e camisetas em português.

Separaremos o conjunto de imagens em 80% das imagens para treino e 20% para avaliação.



Figura 10 – Exemplo de imagem de roupa superior, extraído do dataset Fashion Product Images. Fonte:Aggarwal (2019)

4.1.2 Imagens de estampas

Para formar o dataset de imagens de estampas, utilizamos uma técnica conhecida como *web scraping*, que consiste em navegar e extrair imagens de uma página web. O site istampa (2024) foi utilizado como fonte. Assim como nas imagens de roupas superiores, iremos separar 80% das imagens para treino e 20% para validação. A Figura 11 ilustra uma imagem extraída da página web.



Figura 11 – Exemplo de imagem estampada. Fonte:istampa (2024)

4.1.3 Tratamento das imagens

Tanto as imagens estampadas quanto as de roupas superiores, deverão ser submetidas a um redimensionamento para que todas as imagens do dataset tenham 256x256 pixels.

Além do redimensionamento, aplicaremos um filtro nas imagens estampadas, para ter o formato de uma roupa.

Após os tratamentos mencionados anteriormente, as imagens do dataset CycleGan serão separadas em duas pastas, uma para as imagens superiores e outras para as estampadas. Já as imagens dos dataset Pix2Pix e BicycleGan serão agrupadas, tendo o tamanho 512x256 pixels.

4.2 Código-fonte

O código-fonte dos modelos Pix2Pix (Isola *et al.*, 2016), CycleGAN (Zhu *et al.*, 2017a) e BicycleGAN (Zhu *et al.*, 2018), disponíveis no GitHub, será a base para o desenvolvimento e treinamento dos modelos propostos. Esses modelos serão adaptados para utilizar o dataset criado e detalhado no tópico anterior, assim como as métricas de avaliação dos resultados. A principal biblioteca utilizada para o treinamento e teste dos modelos é a TensorFlow.

Todo o código-fonte, incluindo as adaptações e modificações realizadas, estará disponível no meu repositório do GitHub: <https://github.com/deyviddfs/mba-usp-ia-big-data>.

5 ANÁLISE DE RESULTADOS

Nesta sessão, discutiremos os resultados obtidos através dos experimentos. Para cada experimento, geramos as métricas Inception Score e também enviamos as imagens geradas pelos modelos para um grupo de voluntários avaliar. Ao final, calculamos a média das avaliações.

Treinamos cada modelo com três números de iterações diferentes, conforme apresentado na Tabela 2 e de modo progressivo para validar a hipótese de que um número menor de iterações poderia gerar imagens de melhor qualidade. Com exceção do CycleGAN, os modelos Pix2Pix e BicycleGAN não conseguiram produzir imagens nítidas de roupas estampadas. A Figura 12 e Figura 13 exibem as imagens geradas com base nos correspondentes datasets.



Figura 12 – As imagens da esquerda e do centro fazem parte do dataset Pix2Pix, e a imagem à direita foi gerada pelo modelo treinado.



Figura 13 – O modelo treinado utilizou as imagens da esquerda e do centro, extraídas do dataset BicycleGAN, para gerar a imagem à direita.

Apesar de as imagens do modelo CycleGAN serem um pouco mais nítidas que as dos outros modelos, a qualidade e o realismo das imagens geradas ainda ficaram abaixo do esperado. A Figura 14, ilustra um exemplo de imagem gerada pelo modelo treinado.



Figura 14 – A imagem à direita foi gerada pelo modelo treinado com o dataset CycleGAN.

Nos modelos CycleGAN, Pix2Pix e BicycleGAN, os principais hiperparâmetros ajustados incluem o tamanho do batch que foi definido como 1 para os modelos CycleGAN e Pix2Pix, e 8 para o BicycleGAN. O tamanho das imagens foi de 256x256 pixels para o CycleGAN e Pix2Pix, enquanto no BicycleGAN foi utilizado 128x128 pixels, visando um equilíbrio entre qualidade visual e eficiência computacional. Em todos os modelos, adotou-se o otimizador Adam, com uma taxa de aprendizado de $2e-4$ e beta igual a 0.5. Outro hiperparâmetro relevante, como o número de épocas estão detalhados na Tabela 2.

O resultado completo deste experimento pode ser visualizado na Tabela 2, que exibe o número de iterações utilizadas para treinamento e três imagens geradas após o treinamento, com base no dataset de testes.

Tabela 2 – Imagens geradas pelos modelos treinados.

Modelos	N. Iterações	Imagens geradas
Pix2Pix	4.000	
Pix2Pix	8.000	
Pix2Pix	16.000	
BicycleGan	2.000	
BicycleGan	4.000	
BicycleGan	8.000	
CycleGan	500	
CycleGan	1.000	
CycleGan	1.500	

Com base nas imagens apresentadas anteriormente, submetemos os resultados à avaliação de um grupo de voluntários e geramos métricas automáticas. Os resultados podem ser visualizados na Tabela 3. Para facilitar a análise, destacamos em negrito a maior pontuação de cada modelo em cada métrica, seja o Inception Score ou a média da avaliação dos voluntários.

Tabela 3 – Comparação das métricas geradas das imagens de cada modelo.

Modelos	N. Iterações	Inception Score	Avaliação dos voluntários
Pix2Pix	4.000	1,0451233386993408	1,153846154
Pix2Pix	8.000	1,076288104057312	1,846153846
Pix2Pix	16.000	1,0744099617004395	1,846153846
BicycleGan	2.000	1,004276156425476	1,307692308
BicycleGan	4.000	1,0063849687576294	1,615384615
BicycleGan	8.000	1,0016096830368042	1,076923077
CycleGan	500	1,0927993059158325	2
CycleGan	1.000	1,19273042678833	2,538461538
CycleGan	1.500	1,0091439485549927	1,923076923

Como podemos observar, o modelo CycleGAN obteve as melhores pontuações tanto na avaliação dos voluntários quanto nas métricas automatizadas, indicando um melhor desempenho na geração de imagens de roupas superiores estampadas, utilizando o mesmo conjunto de dados que os modelos Pix2Pix e BicycleGAN. Vale destacar que, embora os resultados tenham sido obtidos com datasets adaptados para cada modelo, todos eles contavam com as mesmas imagens de base.

Uma das hipóteses deste trabalho era que um número menor de iterações poderia gerar imagens mais realistas. No entanto, durante os experimentos, essa hipótese não foi confirmada. Observou-se que, nas primeiras iterações, as imagens apresentavam baixa qualidade, e, após um determinado número de épocas, a qualidade das imagens também começava a deteriorar-se. Isso indica que a quantidade de iterações não é o principal fator determinante para a qualidade das imagens geradas, sendo influenciada por outros elementos, como a arquitetura da rede e o conjunto de dados.

Outro ponto relevante é que os modelos, em sua maioria, não apresentaram melhor desempenho com um maior número de iterações, exceto o modelo Pix2Pix, que apresentou um empate técnico na avaliação dos voluntários entre o segundo e o último número de iterações. As demais métricas também não indicaram um melhor aproveitamento com um maior número de repetições de treinamento.

Isso sugere que o número ideal de iterações depende mais do modelo utilizado e do conjunto de dados do que do número máximo de repetições.

6 CONCLUSÕES

Neste trabalho, investigamos a aplicação de três diferentes modelos GANs no contexto de produção de imagens, com foco em sua capacidade de gerar imagens de alta qualidade a partir de imagens estampadas e roupas superiores. Os resultados obtidos demonstram que, apesar de o modelo CycleGAN não ter alcançado a produção de imagens completamente realistas, seu desempenho superou significativamente os modelos BicycleGAN e Pix2Pix, tanto em termos de qualidade visual quanto em consistência de resultados.

Além disso, uma observação relevante feita durante os experimentos foi a independência entre o número de interações e a obtenção dos melhores resultados para cada modelo. Esse fator sugere que o desempenho dos modelos pode ser mais influenciado por outros parâmetros, como a arquitetura da rede ou o conjunto de imagens, do que pela simples repetição dos processos de treinamento. Outro fator importante está relacionado ao número de interações, sendo observado que cada modelo alcançou o melhor resultado com um número diferente de interações.

Para futuras investigações, sugere-se a exploração e comparação do CycleGAN com o modelo FashionGAN, mencionado no referencial teórico deste trabalho. Além disso, recomenda-se melhorar a qualidade do dataset, tanto em termos de quantidade de imagens quanto de padronização, a fim de potencializar os resultados obtidos. Esses passos poderão contribuir significativamente para o avanço dos estudos relacionados à geração de imagens de roupas estampadas.

REFERÊNCIAS

AGGARWAL, P. **Fashion Product Images Dataset**. 2019. Accessed: 2024-06-02. Disponível em: <https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-dataset>.

BOLZANI, I. **Comparada a uniforme de campo de concentração, roupa gera críticas nas redes; Riachuelo vai retirar peças**.

2023. Disponível em: <https://g1.globo.com/economia/noticia/2023/09/11/comparada-a-uniforme-de-campo-de-concentracao-roupa-gera-criticas-nas-redes-riachuelo-vai-retirar-pecas.html>. Acesso em: 8 jan. 2024.

BRIOT, J.-P. From artificial neural networks to deep learning for music generation: history, concepts and trends. **Neural Computing and Applications**, v. 33, p. 1–27, 01 2021.

CRESWELL, A. *et al.* Generative adversarial networks <i>an overview</i>. IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, 445 HOES LANE, PISCATAWAY, NJ 08855-4141 USA, v. 35, n. 1, p. 53–65, JAN 2018. Disponível em: <https://www-webofscience.ez67.periodicos.capes.gov.br/wos/woscc/full-record/WOS:000422751500008>. Acesso em: 13 abr. 2024.

CUI, Y. R. *et al.* Fashiongan: Display your fashion design using conditional generative adversarial nets. OCT 2018. 26th International Pacific Conference on Computer Graphics and Applications (Pacific Graphics), City Univ Hong Kong, Kowloon Tong, HONG KONG, OCT 08-2018. Disponível em: <https://www-webofscience.ez67.periodicos.capes.gov.br/wos/woscc/full-record/WOS:000448166700011>. Acesso em: 20 mar. 2024.

GCP. **Comparativo GPU**s. 2024. <https://cloud.google.com/compute/docs/gpus?hl=pt-br>. Accessed: 2024-06-02.

GOODFELLOW, I. *et al.* Generative adversarial nets. 2014. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf. Acesso em: 20 mar. 2024.

GOODFELLOW, I. *et al.* Generative adversarial networks. ASSOC COMPUTING MACHINERY, 1601 Broadway, 10th Floor, NEW YORK, NY USA, NOV 2020. Disponível em: <https://www-webofscience.ez67.periodicos.capes.gov.br/wos/woscc/full-record/WOS:000585011100041>. Acesso em: 20 mar. 2024.

ISOLA, P. *et al.* **pix2pix**. 2016. <https://github.com/phillipi/pix2pix>. Accessed: 2024-05-02.

ISOLA, P. *et al.* Image-to-image translation with conditional adversarial networks. *In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. [S.l.: s.n.], 2017.

ISTAMPA, B. **Estampas Exclusivas**. 2024. <https://bellaistampa.com.br/estampas-exclusivas?page=1>. Accessed: 2024-06-02.

LIU, L. *et al.* Collocating clothes with generative adversarial networks cosupervised by categories and attributes: A multidiscriminator framework. **IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS**, SEPT 2020.

Disponível em: <https://www-webofscience.ez67.periodicos.capes.gov.br/wos/woscc/full-record/WOS:001080441800026>. Acesso em: 20 mar. 2024.

MINO, A.; SPANAKIS, G. Logan: Generating logos with a generative adversarial neural network conditioned on color. 10 2018.

MIRZA, M.; OSINDERO, S. **Conditional Generative Adversarial Nets**. 2014.

REED, S. E. *et al.* Generative adversarial text to image synthesis. **CoRR**, abs/1605.05396, 2016. Disponível em: <http://arxiv.org/abs/1605.05396>.

SALIMANS, T. *et al.* **Improved Techniques for Training GANs**. 2016.

SHI, Y.; LI, Q.; ZHU, X. X. Building footprint generation using improved generative adversarial networks. **IEEE GEOSCIENCE AND REMOTE SENSING LETTERS**, IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, 445 HOES LANE, PISCATAWAY, NJ 08855-4141 USA, v. 16, n. 4, p. 603–607, APR 2019. Disponível em: <https://www-webofscience.ez67.periodicos.capes.gov.br/wos/woscc/full-record/WOS:000462443300022>. Acesso em: 20 mar. 2024.

TIAN, N. *et al.* Colorization of logo sketch based on conditional generative adversarial networks. **ELECTRONICS**, MDPI, v. 10, n. 4, FEB 2021. Disponível em: <https://www-webofscience.ez67.periodicos.capes.gov.br/wos/woscc/full-record/WOS:000623373800001>. Acesso em: 20 mar. 2024.

ZHU, J.-Y. *et al.* **CycleGAN**. 2017. <https://github.com/junyanz/CycleGAN>. Accessed: 2024-05-02.

ZHU, J.-Y. *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. *In: Computer Vision (ICCV), 2017 IEEE International Conference on*. [S.l.: s.n.], 2017.

ZHU, J.-Y. *et al.* Toward multimodal image-to-image translation. *In: Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017.

ZHU, J.-Y. *et al.* **BicycleGAN**. 2018. <https://github.com/junyanz/BicycleGAN>. Accessed: 2024-05-02.