

Redes Adversárias Generativas (GANs) para preencher regiões em imagens (inpainting)

Lucas Valdivieso Junkes¹, Gilberto Fernandes Junior¹

¹Departamento de Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

lucas.valdivieso@uel.br, gilfernandes@uel.br

Abstract. *Filling in image gaps, known as image inpainting, is a computer vision technique that involves completing empty spaces in images in a visually plausible manner. This technique is useful in various practical applications, such as image restoration and editing. The importance of this topic lies in the growing need for advanced techniques that can produce visually convincing and high-quality results. Recent advancements in the field, through Generative Adversarial Networks (GANs), have significantly improved the quality and precision of inpainting techniques. All things considered, this work aims to study GANs and implement a model capable of filling in missing regions in images, with the goal of contributing to the advancement of inpainting techniques using GANs.*

Resumo. *O preenchimento de espaços em imagens, conhecido como image inpainting, é uma técnica de visão computacional que consiste em completar espaços vazios em imagens de maneira visualmente plausível. Essa técnica é útil em diversas aplicações práticas, como restauração e edição de imagens. A importância desse tema reside na necessidade crescente de técnicas avançadas que possam produzir resultados visualmente convincentes e de alta qualidade. Avanços recentes na área, por meio de Redes Generativas Adversárias (GANs), têm aprimorado significativamente a qualidade e a precisão das técnicas de inpainting. Considerando isso, este trabalho busca estudar as GANs e implementar um modelo capaz de preencher regiões ausentes em imagens, visando contribuir para o avanço da técnica de inpainting com GANs.*

1. Introdução

A recuperação de informações perdidas ou a remoção de objetos indesejados em imagens, é um desafio significativo no campo do processamento de imagem e da visão computacional. Uma técnica central para abordar esses problemas é o *image inpainting* [1], que se refere ao processo de preencher áreas ausentes em uma imagem de forma visualmente coerente. Essa técnica possui uma ampla gama de aplicações práticas, e sua relevância se torna ainda mais evidente à medida que a quantidade de dados visuais aumenta exponencialmente [11], demandando métodos automáticos e eficientes para melhorar e preservar a integridade das imagens.

Existem diversos métodos e algoritmos para se lidar com preenchimento de imagens [19], porém, com o rápido crescimento e desenvolvimento do *deep learning* na visão computacional, vários modelos baseados em *deep learning* começaram a ganhar destaque na área [24, 17], demonstrando uma alta eficiência em tarefas como *inpainting* [14].

Nos últimos anos, as Redes Generativas Adversárias (GANs) emergiram como uma das abordagens mais promissoras para diversas tarefas de geração de imagens, incluindo o *inpainting*. Introduzidas por Goodfellow et al. em 2014, [6] as GANs consistem em dois componentes principais: um gerador, que cria novas amostras de dados, e um discriminador, que avalia a autenticidade dessas amostras. O treinamento conjunto desses dois componentes resulta em modelos capazes de gerar imagens de alta qualidade. Como consequência, rapidamente se tornaram uma das técnicas mais usadas na pesquisa de preenchimento de imagem [23].

Neste contexto, este trabalho busca explorar os fundamentos das GANs, aplicando-as no campo de preenchimento de espaços em imagens. Por meio da implementação de um modelo capaz de realizar o *inpainting*, o objetivo é contribuir para o desenvolvimento dessa área.

A estrutura desse documento está organizada da seguinte forma: A Seção 2 detalha os conceitos, métodos e aplicações abordados durante o processo de revisão bibliográfica, essenciais para o entendimento e desenvolvimento posterior do projeto. A Seção 3 especifica os objetivos gerais e específicos do projeto. A Seção 4 descreve como os objetivos da sessão anterior serão atingidos e de que modo serão utilizados os conceitos, métodos e técnicas estudadas. A Seção 5 define o cronograma previsto para a conclusão das atividades propostas. A Seção 6 identifica de que maneira, a pesquisa pode contribuir para o avanço ou a consolidação do conhecimento do leitor.

2. Fundamentação Teórico-Methodológica e Estado da Arte

2.1. Processamento de Imagem

Processamento de imagem é um campo da visão computacional que envolve a análise e manipulação de imagens digitais, com o objetivo de extrair informações úteis, melhorar a qualidade da imagem ou transformar a imagem para diversas finalidades. Este campo abrange uma ampla gama de técnicas e algoritmos, que vão desde operações básicas, como ajuste de contraste e filtragem, até métodos avançados, como reconhecimento de padrões e segmentação de imagens.

As aplicações desta área são amplas, incluindo medicina, onde é utilizado para análise de imagens médicas, como raios-X e ressonâncias magnéticas, ajudando no diagnóstico e monitoramento de doenças [5]. Na segurança, é essencial para sistemas de vigilância e reconhecimento facial [16]. Também é fundamental em indústrias, auxiliando na inspeção de qualidade de produtos e automação de processos [3]. Além disso, é utilizado em entretenimento para efeitos visuais em filmes e jogos, e na realidade aumentada e virtual para criar experiências imersivas [22].

O estado da arte em processamento de imagem tem avançado significativamente com o desenvolvimento de novas técnicas e tecnologias, como redes neurais e *deep learning*, facilitadas em grande parte pelos avanços recentes de hardware, principalmente no uso das GPUs (unidades de processamento gráfico) e processamento paralelo [10].

2.2. Deep Learning

O aprendizado profundo (*deep learning*) tem chamado muita atenção nos últimos anos, devido à sua grande capacidade de permitir que modelos computacionais com múltiplas

camadas de processamento aprendam e representem dados com vários níveis de abstração, imitando a forma como o cérebro percebe e entende informações. Essa capacidade tem possibilitado superar diversos modelos considerados estado da arte em várias áreas, sendo o campo da visão computacional um dos casos mais proeminentes [21].

O *deep learning* é uma família diversificada de métodos, incluindo redes neurais, modelos probabilísticos hierárquicos e uma variedade de algoritmos de aprendizado de características supervisionados e não supervisionados. Entre os principais, podemos destacar modelos como:

- Redes Neurais Convolucionais (CNNs): Excelentes para processamento de dados com estrutura de grade, como imagens, as CNNs utilizam camadas convolucionais para capturar características locais e reduzir a dimensionalidade por meio de operações de *pooling* [13].
- Redes Neurais Recorrentes (RNNs): Adequadas para dados sequenciais, como texto e séries temporais, as RNNs possuem conexões recorrentes que permitem a propagação de informações ao longo do tempo. LSTM (*Long Short-Term Memory*) e GRU (*Gated Recurrent Unit*) são variantes populares que mitigam problemas de *vanishing gradient* [7].
- Redes Neurais de Transformadores: Introduzidas por Vaswani em 2017, as redes de transformadores utilizam mecanismos de atenção para processar dados sequenciais de forma paralela, superando limitações das RNNs em tarefas como tradução automática e processamento de linguagem natural [20].

2.2.1. Redes Generativas Adversárias (GANs)

As Redes Adversárias Generativas (GANs) foram introduzidas em 2014 por Ian Goodfellow et al. e rapidamente se tornaram uma das abordagens mais inovadoras e influentes no campo do *deep learning*. As GANs consistem em duas redes neurais que competem em um jogo de soma zero: a rede geradora (G) e a rede discriminadora (D). Nesse jogo, a rede geradora tenta criar dados falsos que imitem os dados reais, enquanto a rede discriminadora tenta distinguir entre dados reais e falsos. O objetivo do treinamento é alcançar um equilíbrio em que a rede geradora produz dados que o discriminador não consegue distinguir dos dados reais [6].

Nesse processo, ambos os competidores são sincronizados, isto é, enquanto a G tenta gerar resultados mais realistas, a D melhora sua precisão na detecção de quais amostras são reais e quais não são. Esse processo é um jogo de minimax, pois, se a G cria uma saída melhor, será mais difícil para a D diferenciá-las. Por outro lado, se a D é mais precisa, será mais difícil para a G enganar a D. Portanto, nesse processo, a D tenta maximizar a precisão e a G tenta minimizá-la [8]. A formulação da função de perda do jogo minimax pode ser denotada como:

$$\min_G \max_D L(D, G) = \mathbb{E}_{x \sim p_r} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))] \quad (1)$$

onde $x \sim p_r$ é a distribuição dos dados reais e $z \sim p_z$ denota a distribuição de probabilidade do espaço latente de G . $z \sim p_z$ é comumente um ruído Gaussiano

ou uniforme que G usa para modelar novas amostras de dados, denotadas como $G(z)$. A função de D é diferenciar entre a distribuição real $D(x)$ e a distribuição sintetizada $D(G(x))$ [8].

GANs são amplamente utilizadas em diversos campos de pesquisa, como por exemplo, *data augmentation* [15], criação de jogos eletrônicos [12], preenchimento de espaço em imagens [2], dentre outras, nas quais são criados diferentes modelos GAN, com alterações no treinamento ou mudanças à estrutura dos modelos G ou D, para lidar com as peculiaridades de cada aplicação. Exemplos de alguns modelos são:

- Co-Mod GAN: Usado para lidar de forma mais eficiente com o preenchimento de grandes regiões ausentes na imagem [26]
- Pix2Pix: Onde o principal objetivo é fazer uma tradução de imagem para imagem, isto é, dado uma imagem de um domínio A, transformar essa imagem para um domínio B. Por exemplo, transformar uma foto de uma rua, em uma foto aérea daquela rua. [9]

2.3. Inpainting

Uma das práticas que a GAN se destaca é no preenchimento de imagens (*inpainting*) [1], que é uma técnica de processamento de imagem capaz de completar regiões em imagens de maneira imperceptível ao observador. Essa é uma técnica antiga, e suas aplicações são inúmeras, desde restaurações de fotografias até remoções ou substituições de objetos indesejáveis.

Na Europa desde o Renascimento (final do século XIII – meados do século XVII), a prática de restauração de pinturas já era amplamente utilizada [4], com o intuito de reconstruir ou retocar porções das obras de arte. Na era moderna novos modos de se realizar essas modificações surgiram, desde ferramentas de software que se utilizam do processo manual até formas totalmente automáticas de se fazer o *inpainting* digital [25].

Existem diversas técnicas para a realização do *inpainting*, que podem ser categorizadas de duas formas, técnicas tradicionais e técnicas que utilizam *deep learning*, cada uma com abordagens e aplicações específicas. Dentre as técnicas tradicionais, pode-se destacar o *inpainting* por difusão, que preenche áreas ausentes difundindo informações das bordas para dentro da área danificada. Métodos baseados em textura, que usam amostras de textura da própria imagem para preencher áreas faltantes, garantindo que o preenchimento mantenha a aparência original. Técnicas baseadas em exemplos (*example-based*) procuram regiões similares em outras partes da imagem ou em um banco de dados de imagens, usando essas amostras para preencher a área ausente [18].

Entretanto, a maioria das técnicas tradicionais, pode levar a imagens borradas e enfrenta dificuldades em manter a consistência e a naturalidade da imagem preenchida, especialmente em regiões maiores ou mais complexas. Em contraste, métodos que utilizam *deep learning*, especialmente aqueles baseados em GANs, conseguem superar essas limitações, pois são capazes de gerar preenchimentos mais realistas e coerentes, mantendo a continuidade e os detalhes finos da imagem original. Por conta de seus resultados superiores, muitos estudos são realizados em cima dessas técnicas, incentivando atualizações, ou até mesmo a implementações de novas abordagens.

Um dos modelos de *deep learning* mais avançados para esse problema é o *co-modulated generative adversarial networks* [26]. Esse modelo busca resolver uma das

maiores limitações dos algoritmos atuais, que tendem a falhar quando apresentados com uma grande parte ausente da imagem. Para resolver esse problema diversas melhorias foram feitas na arquitetura da GAN, introduzindo um mecanismo de co-modulação que permite que a rede geradora (G) utilize informações contextuais mais amplas da imagem. Esse mecanismo modula o gerador com informações obtidas de outras partes da imagem, ajudando a manter a coerência global, tornando-o capaz de gerar conteúdos diversos e consistentes não apenas para *inpainting* em pequena escala, mas também para o preenchimento de imagens em grande escala.

3. Objetivos

O objetivo deste projeto é explorar as Redes Adversárias Generativas (GANs) e aplicá-las na área de preenchimento de espaços em imagens (*image inpainting*). Para que este objetivo seja alcançado, os seguintes objetivos específicos foram definidos:

1. Estudar a fundamentação teórica, compreender o funcionamento e as diversas aplicações das Redes Adversárias Generativas (GANs), analisando sua arquitetura, mecanismos de treinamento e casos de uso em diferentes domínios.
2. Desenvolver um modelo GAN especificamente para a tarefa de *inpainting* em imagens, buscando uma abordagem eficiente e eficaz para o preenchimento de regiões ausentes.
3. Avaliar a performance do modelo proposto em comparação com outros modelos de *inpainting* já existentes na literatura.
4. Identificar possíveis melhorias, seja no modelo proposto, em modelos existentes, ou nas métricas de *benchmark* utilizadas para avaliação, visando contribuir para o avanço da técnica de *inpainting* com GANs.

4. Procedimentos metodológicos/Métodos e técnicas

Inicialmente, será realizado um estudo por meio de uma revisão bibliográfica sobre Redes Adversárias Generativas (GANs), suas arquiteturas, mecanismos de treinamento e aplicações específicas para *inpainting* em imagens. Este estudo incluirá a análise de modelos de *inpainting* existentes e as métricas de *benchmark* utilizadas para avaliar a eficácia desses modelos.

Após essa etapa, será selecionado um conjunto de dados adequado para o treinamento e teste do modelo que será desenvolvido. A escolha do conjunto de dados será baseada na sua relevância e qualidade, assegurando que ele contenha uma diversidade suficiente de imagens para permitir um treinamento robusto e uma avaliação abrangente do modelo.

Em seguida, o trabalho apresentará um modelo que utilize GAN para realizar a tarefa de preenchimento de espaços em imagens. Esse modelo será treinado, testado e avaliado para verificar como se compara com outros métodos que são referência nessa área.

Com base nas limitações e pontos de melhoria identificados durante o desenvolvimento e a comparação do modelo, serão sugeridas e implementadas melhorias no modelo desenvolvido ou nos modelos existentes, buscando aprimorar a qualidade ou a eficiência do processo de treinamento. Também serão sugeridas melhorias nas métricas de *benchmark*, se necessário, para uma avaliação mais precisa e completa dos modelos de *inpainting*.

5. Cronograma de Execução

Atividades:

1. Revisão bibliográfica;
2. Desenvolvimento do modelo GAN para *inpainting*;
3. Treinamento, teste e avaliação;
4. Comparação com outros modelos;
5. Propostas de melhorias;
6. Escrita da versão preliminar do TCC;
7. Escrita da versão final para a banca examinadora do TCC;

Tabela 1. Cronograma de Execução

	jun	jul	ago	set	out	nov	dez	jan	fev
Atividade 1	•	•							
Atividade 2			•	•					
Atividade 3				•	•				
Atividade 4					•	•			
Atividade 5						•	•		
Atividade 6			•	•	•				
Atividade 7							•	•	•

6. Contribuições e/ou Resultados esperados

Ao final do trabalho, espera-se que a pesquisa contribua para um entendimento mais profundo da área de processamento de imagens e visão computacional, além de fornecer uma visão sobre como os algoritmos de aprendizado profundo, mais especificamente os modelos GANs, são estruturados e suas aplicações práticas.

Além disso, espera-se que o estudo e a aplicação das Redes Adversárias Generativas (GANs), para a tarefa de *inpainting*, contribuam para o avanço do estado da arte e revelem limitações dessas redes no contexto do preenchimento de regiões ausentes em imagens.

7. Espaço para assinaturas

Londrina, 29 de julho de 2024.

Aluno

Orientador

Referências

- [1] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00*, page 417–424, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

- [2] Xun Dong and Ruijia Hua. Gan based image inpainting methods: A taxonomy. In *2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI)*, pages 145–150, 2022.
- [3] Charbel El Hachem, Raoul Santiago, Loïc Painvin, Gilles Perrot, and Raphaël Couturier. Brick orientation adjustment in the automotive industry using image processing techniques. In *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, volume 1, pages 729–733, 2022.
- [4] G. Émile-Mâle. *The Restorer's Handbook of Easel Painting*. Van Nostrand Reinhold, 1979.
- [5] Navnish Goel, Akhilendra Yadav, and Brij Mohan Singh. Medical image processing: A review. In *2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH)*, pages 57–62, 2016.
- [6] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- [7] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427, 2021.
- [8] Guillermo Iglesias, Edgar Talavera, and Alberto Díaz-Álvarez. A survey on gans for computer vision: Recent research, analysis and taxonomy. *Computer Science Review*, 48:100553, May 2023.
- [9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018.
- [10] Won Jeon, Gun Ko, Jiwon Lee, Hyunwuk Lee, Dongho Ha, and Won Woo Ro. Chapter six - deep learning with gpus. In Shiho Kim and Ganesh Chandra Deka, editors, *Hardware Accelerator Systems for Artificial Intelligence and Machine Learning*, volume 122 of *Advances in Computers*, pages 167–215. Elsevier, 2021.
- [11] D.A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. Challenges in visual data analysis. In *Tenth International Conference on Information Visualisation (IV'06)*, pages 9–16, 2006.
- [12] Seung Wook Kim, Yuhao Zhou, Jonah Philion, Antonio Torralba, and Sanja Fidler. Learning to simulate dynamic environments with gamegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1231–1240, 2020.
- [13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [14] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European conference on computer vision (ECCV)*, pages 85–100, 2018.

- [15] Saman Motamed, Patrik Rogalla, and Farzad Khalvati. Data augmentation using generative adversarial networks (gans) for gan-based detection of pneumonia and covid-19 in chest x-ray images. *Informatics in Medicine Unlocked*, 27:100779, 2021.
- [16] Hongxu Peng. Computer information technology and network security analysis of intelligent image recognition. In *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, pages 1103–1106, 2022.
- [17] Zhen Qin, Qingliang Zeng, Yixin Zong, and Fan Xu. Image inpainting based on deep learning: A review. *Displays*, 69:102028, 2021.
- [18] Nermin Salem. A survey on various image inpainting techniques. *Future Engineering Journal*, 2, 09 2021.
- [19] Tammineni Shanmukhaprasanthi, Swarajya Madhuri Rayavarapu, Yenneti Laxmi Lavanya, and Gottapu Sasibhusana Rao. A comprehensive study of image inpainting techniques with algorithmic approach. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, pages 1–5, 2023.
- [20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [21] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018(1):7068349, 2018.
- [22] Chunmei Wang. Research on panoramic image processing technology based on virtual reality technology. In *2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, pages 55–58, 2019.
- [23] Nanxiang Wang. A survey on improved gan based image inpainting for different aims. *Highlights in Science, Engineering and Technology*, 39:347–355, Apr. 2023.
- [24] Hanyu Xiang, Qin Zou, Muhammad Ali Nawaz, Xianfeng Huang, Fan Zhang, and Hongkai Yu. Deep learning for image inpainting: A survey. *Pattern Recognition*, 134:109046, 2023.
- [25] Yongsheng Yu, Libo Zhang, Heng Fan, and Tiejian Luo. High-fidelity image inpainting with gan inversion. In *European Conference on Computer Vision*, pages 242–258. Springer, 2022.
- [26] Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I Chang, and Yan Xu. Large scale image completion via co-modulated generative adversarial networks, 2021.