Refinamento de Imagens de Satélite: Uma Abordagem de Super-Resolução para Preparação de Datasets

Vitor Yuske Watanabe
Faculdade de Computação - FACOM
UFMS
Campo Grande - MS, Brasil
vitor.y@ufms.br

Pedro Arfux Pereira Cavalcante de Castro
Faculdade de Computação - FACOM
UFMS
Campo Grande - MS, Brasil
pedroarfux.castro@gmail.com

Wesley Nunes Gonçalves
Faculdade de Computação - FACOM
UFMS
Campo Grande - MS, Brasil
wesley.goncalves@ufms.br

***Abstract*—Este trabalho investiga o uso de técnicas de super-resolução em imagens capturadas por Veículos Aéreos Não Tripulados (VANTs), com o objetivo de aprimorar artificialmente a resolução espacial por meio de algoritmos de inteligência artificial. A proposta busca explorar metodologias de estado da arte para superar limitações na qualidade das imagens, facilitando sua aplicação em tarefas como monitoramento ambiental, mapeamento urbano e análise agrícola. Os resultados obtidos demonstram o potencial da super-resolução para melhorar a qualidade dos conjuntos de dados, contribuindo para a aplicação de visão computacional em cenários que exigem alta precisão na identificação de objetos e características pequenas.**

***Index Terms*—super-resolução, visão computacional, drones.**

I. Introdução

A identificação de objetos pequenos em imagens de satélite é um desafio crucial em diversas aplicações, como monitoramento ambiental e planejamento urbano [3], [13]. Imagens de satélite, embora eficientes em cobrir vastas áreas com alta frequência, frequentemente apresentam resolução limitada, dificultando a identificação de objetos menores e comprometendo tarefas críticas [14]. Este problema é agravado pelas limitações tecnológicas de sensores remotos e pelas condições adversas do ambiente de aquisição, que introduzem ruídos e distorções [4]. Assim, a necessidade de melhorar a resolução de tais imagens é uma questão relevante em aplicações de sensoriamento remoto.

Em contrapartida, imagens capturadas por drones (VANTs - Veículos Aéreos Não Tripulados) são amplamente reconhecidas por sua alta resolução e riqueza de detalhes, tornando-as valiosas para aplicações que exigem precisão, como mapeamento de pequenos objetos e monitoramento ambiental detalhado. No entanto, essas vantagens são limitadas por sua cobertura espacial restrita e baixa frequência de aquisição, causadas por fatores como autonomia de voo, capacidade de bateria e custo operacional elevado, tornando-as menos adequadas para monitoramento em larga escala [14]. Tais limitações em opções de coleta dificultam a obtenção de imagens de boa qualidade e com grande cobertura espacial, o que trouxe a necessidade de desenvolver soluções com o intuito de melhorar a qualidade das imagens de satélite.

Nesse contexto, métodos de super-resolução (SR), em especial aqueles baseados em aprendizado profundo, têm mostrado grande potencial para superar essas limitações. Técnicas modernas, como redes neurais convolucionais profundas (CNNs), permitem reconstruir imagens de alta resolução a partir de versões de baixa qualidade, preservando detalhes e melhorando a acurácia em tarefas de visão computacional [6], [8]. Modelos como SRCNN [4] e VDSR [7] foram pioneiros ao introduzir o aprendizado residual e aumentar a profundidade das redes, enquanto abordagens mais recentes, como EDSR [10], eliminaram camadas como batch normalization para preservar a integridade das cores e melhorar o desempenho em super-resolução. Recentemente, o *Local Implicit Image Function* (LIIF) permitiu a reconstrução de imagens de alta resolução de forma contínua, ou seja, o modelo treinado consegue operar em um intervalo de ampliações e pode ser utilizado para resoluções maiores do que as usadas no treino [2].

Neste trabalho, propomos o uso de técnicas de super-resolução, especificamente o LIIF, para aumentar artificialmente imagens de baixa resolução, possibilitando uma melhoria na identificação de objetos pequenos, e permitindo a viabilidade de uso de imagens de satélite em projetos de visão computacional que necessitem de imagens de alta resolução. Para validar a proposta, imagens capturadas por VANTs foram pré-processadas para reduzir intencionalmente sua qualidade, gerando uma imagem de baixa resolução. Em seguida, o LIIF foi aplicado para aumentar a resolução, sendo que esse método usa a imagem original como *ground truth* durante o treinamento. Os resultados demonstraram que o LIIF apresentou métricas de precisão superiores em comparação com o método tradicional de redimensionamento de imagens, como o implementado na biblioteca OpenCV [1]. Esses resultados evidenciam o potencial da super-resolução na preparação de conjuntos de dados para aplicações de visão computacional que requerem alta acurácia.

II. Referencial Teórico

A fundamentação teórica deste trabalho é composta por dois eixos principais: os avanços recentes em super-resolução de imagens e as métricas utilizadas para avaliar a qualidade das imagens geradas.

A. Métodos de Super-resolução

Super-resolução é o processo de aumentar a resolução espacial de imagens de forma artificial, buscando preservar ou aprimorar detalhes visuais [5]. Métodos tradicionais, como interpolação bicúbica (implementada no OpenCV), frequentemente falham em recuperar detalhes finos, levando ao desenvolvimento de abordagens baseadas em aprendizado profundo, que revolucionaram a área ao oferecer melhorias significativas na reconstrução de imagens de alta resolução [14].

O uso de redes neurais profundas, como proposto em trabalhos seminalistas como SRCNN [4] e EDSR [10], introduziu arquiteturas especializadas para a reconstrução de detalhes em imagens. A introdução de técnicas como VDSR [7] e DRCN [8] demonstrou que redes profundas com treinamento residual poderiam melhorar ainda mais os resultados, enquanto modelos como o LIIF focaram em representações contínuas para imagens, permitindo ampliação em escalas arbitrárias [2].

No contexto deste trabalho, foi adotado o modelo LIIF proposto por Chen et al. [2]. Este modelo representa imagens de forma contínua ao prever valores RGB em coordenadas específicas, utilizando coordenadas espaciais e características profundas 2D locais como entrada. Essa abordagem permite reconstruções em resoluções arbitrárias, oferecendo flexibilidade em diversas aplicações. O LIIF é treinado de maneira auto-supervisionada, utilizando pares de imagens de alta e baixa resolução para minimizar perdas durante a reconstrução.

No campo do sensoriamento remoto, essas tecnologias têm sido aplicadas para melhorar a análise de imagens de satélite, que, embora abrangentes em termos de área, frequentemente carecem de resolução suficiente para identificar pequenos objetos [3], [13]. Métodos como MetaSR também contribuíram para a ampliação de imagens em resoluções arbitrárias, mas o LIIF demonstrou uma capacidade superior de extrapolação para resoluções muito maiores do que aquelas apresentadas durante o treinamento [2], além de apresentar resultados com menos alucinações em relação aos outros modelos mencionados.

B. Métricas de Desempenho

Para avaliar a qualidade das imagens geradas, foram utilizadas métricas amplamente reconhecidas na literatura de visão computacional, garantindo uma análise quantitativa e qualitativa do desempenho dos modelos de super-resolução. Três métricas principais foram consideradas: o Peak Signal-to-Noise Ratio (PSNR), o Structural Similarity Index (SSIM) e o Mean Absolute Error (MAE).

Antes de explicar sobre o PSNR, SSIM e MAE, devemos mencionar o MSE, que é o *Mean Squared Error*, dado por:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(I_i - \hat{I}_i\right)^2,$$

onde I_i e $\hat{I}_i$ são, respectivamente, os valores dos pixels da imagem original e da imagem reconstruída, e N é o número total de pixels.

O PSNR mede a relação entre o sinal máximo possível de uma imagem e o ruído presente na imagem reconstruída. Ele é calculado em decibéis (dB) e é definido conforme a Equação 1.

$$PSNR = 10 \cdot \log_{10}\left(\frac{\text{MAX}^2}{\text{MSE}}\right), \tag{1}$$

onde MAX representa o valor máximo de intensidade dos pixels (normalmente 255. Valores mais altos de PSNR indicam menor distorção e, consequentemente, maior qualidade da reconstrução.

O SSIM, por sua vez, é uma métrica perceptual que avalia a similaridade estrutural entre duas imagens, sendo mais sensível à percepção humana (Equação 2). Ele considera três componentes: luminância (l), contraste (c) e estrutura (s).

$$SSIM(I,\hat{I}) = \left[l(I,\hat{I})\right]^{\alpha} \cdot \left[c(I,\hat{I})\right]^{\beta} \cdot \left[s(I,\hat{I})\right]^{\gamma}, \tag{2}$$

No qual os termos são calculados por:

$$l(I,\hat{I}) = \frac{2\mu_I\mu_{\hat{I}}}{\mu_I^2 + \mu_{\hat{I}}^2}, \tag{3}$$

$$c(I,\hat{I}) = \frac{2\sigma_I\sigma_{\hat{I}}}{\sigma_I^2 + \sigma_{\hat{I}}^2}, \tag{4}$$

$$s(I,\hat{I}) = \frac{\sigma_{I\hat{I}}}{\sigma_I\sigma_{\hat{I}}}. \tag{5}$$

Aqui, μ_I e $\mu_{\hat{I}}$ são as médias, σ_I e $\sigma_{\hat{I}}$ são os desvios-padrão, $\sigma_{I\hat{I}}$ é a covariância entre as imagens. O valor de SSIM varia entre -1 e 1, com valores mais próximos de 1 indicando maior similaridade.

Por fim, o MAE avalia a diferença absoluta média entre os valores de pixel da imagem original e da reconstruída. Ele é definido como:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|I_i - \hat{I}_i\right|,$$

onde valores menores indicam maior precisão na reconstrução, uma vez que refletem diferenças menores entre as duas imagens.

III. Materiais e Métodos

A. Conjunto de Dados

O conjunto de dados utilizado neste estudo, ilustrado na Figura 1, é composto por 3637 recortes de imagens capturadas por drones, todas com resolução de 960x960 pixels. Para o particionamento do conjunto, 3519 imagens (97%) foram destinadas ao treinamento, enquanto 118 imagens (3%) foram reservadas para validação. Essa divisão foi feita para garantir que regiões do treino não estejam em regiões de teste. Cada imagem apresenta uma resolução espacial correspondente a 0,5 metros por pixel, proporcionando um nível de detalhe adequado para tarefas de super-resolução.

Fig. 1: Exemplo de imagens do conjunto de dados

B. Metodologia

O diagrama apresentado na Figura 2 descreve o fluxo de trabalho utilizado para avaliar o desempenho do modelo de super-resolução. Inicialmente, a imagem de entrada é processada pelo método proposto, resultando em uma imagem predita. Paralelamente, a imagem original, utilizada como referência, é comparada tanto com a imagem predita quanto com uma versão ampliada da imagem gerada pelo algoritmo tradicional (*resized*) de redimensionamento bicúbico da biblioteca OpenCV [1]. A comparação é realizada utilizando as métricas quantitativas *Peak Signal-to-Noise Ratio* (PSNR), *Structural Similarity Index* (SSIM) e *Mean Absolute Error* (MAE), que fornecem uma avaliação objetiva da qualidade da reconstrução de cada método. Por fim, os resultados obtidos são consolidados, permitindo uma análise quantitativa e qualitativa da abordagem proposta em relação a técnicas convencionais.

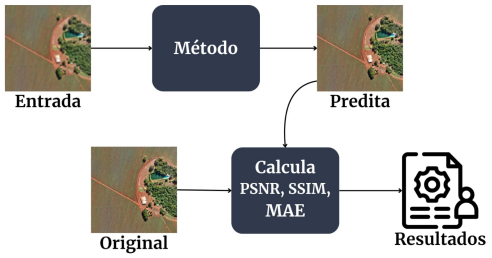


Fig. 2: Fluxo de processo para avaliação do modelo de super-resolução.

Para preparar as imagens para o processo de super-resolução, as imagens originais de alta resolução são degradadas para gerar versões de baixa resolução. Utilizamos um fator de downsampling específico (2x, 3x e 4x), aplicando interpolação bicúbica para simular imagens de resolução reduzida. Resumidamente, podemos descrever as possibilidades de imagens de alta e baixa resolução da metodologia na Tabela I, descrevendo o número do treinamento 1, 2, 3 e 4, as resluções de entrada (baixa rasolução) e saída (alta resolução).

Treinamento	Baixa $\rightarrow$ Alta
1	480x480 $\rightarrow$ 960x960
2	320x320 $\rightarrow$ 960x960
3	240x240 $\rightarrow$ 960x960
4	96x96 $\rightarrow$ 384x384

TABLE I: Possibilidades de imagens de baixa e alta resolução usadas na metodologia.

Ademais, em um downsampling de 2x, as imagens de 960x960 (chamadas de imagens de alta resolução) foram redimensionadas para 480x480 (imagens chamadas de baixa resolução). As imagens de baixa resolução são dadas como entrada para um método de super-resolução que prediz a imagem de alta resolução. Durante o treino, a imagem de alta resolução original é comparada com a imagem predita para corrigir o método. Para verificar a possibilidade de imagens ainda menores, um novo treino foi realizado com imagens de 384x384 como imagens de alta resolução e 96x96 como sua versão em baixa resolução. Optou-se por utilizar imagens de 96x96 ampliadas para 384x384 para melhorar a qualidade visual em relação aos treinamentos anteriores e avaliar o impacto de uma nova resolução alvo. Além disso, o foco em uma ampliação de 4x visou testar a robustez do modelo em reconstruções de maior escala, atendendo a cenários que demandam imagens de alta resolução.

Neste trabalho, nós empregamos o método LIIF proposto por Chen et al. [2] para realizar a super-resolução das imagens degradadas. O LIIF é uma abordagem que representa imagens de forma contínua, permitindo a geração de imagens em resoluções arbitrárias a partir de uma representação latente aprendida. O LIIF consiste em duas componentes principais:

- Codificador (Encoder): Utilizamos a arquitetura EDSR [10] como base para o codificador, dada sua eficiência

comprovada em tarefas de super-resolução. O codificador extrai características profundas das imagens de baixa resolução, produzindo um mapa de características latentes que preserva informações essenciais para a reconstrução.
- Função de Decodificação Implícita: Implementada como uma rede neural *Multi-Layer Perceptron* (MLP), esta função recebe como entrada as coordenadas espaciais contínuas dos pixels e as características latentes correspondentes. Ela prevê os valores RGB para cada ponto na imagem de alta resolução desejada, permitindo a reconstrução da imagem com detalhes refinados.

Os métodos de desdobramento de recursos (*feature unfolding*), conjunto local (local ensemble) e decodificação de células (*cell decoding*) são técnicas fundamentais no modelo LIIF para aprimorar a super-resolução de imagens. O desdobramento de recursos enriquece a informação contida nos códigos latentes ao integrar os valores de códigos latentes vizinhos, formando um contexto mais amplo para as predições. O conjunto local resolve a questão da descontinuidade nas previsões, utilizando múltiplos códigos latentes próximos e combinando suas predições por meio de uma votação ponderada, garantindo transições suaves entre as regiões da imagem. Já a decodificação de células adapta o processo de predição ao considerar o tamanho do pixel, não apenas o valor central, mas também sua área circundante, resultando em predições mais precisas e contextuais. Esses métodos combinados permitem que o LIIF gere imagens de alta resolução de forma mais eficaz, proporcionando maior fidelidade e continuidade nas transições das imagens reconstruídas. .

O treinamento foi realizado de forma auto-supervisionada, utilizando pares de imagens de alta e baixa resolução, com a minimização da perda *L1* entre os valores preditos e os valores reais. A *L1 Loss*, também conhecida como *Mean Absolute Error* (MAE), calcula a diferença absoluta média entre os valores preditos pelo modelo e os valores reais, sendo definida na Equação 5, onde y_i representa o valor real e $\hat{y}_i$ o valor predito para a i-ésima amostra. Essa função de perda foi escolhida por sua simplicidade e robustez a outliers, o que é particularmente importante em tarefas de reconstrução, como a super-resolução de imagens, onde a preservação de detalhes finos é essencial para garantir a qualidade das imagens geradas.

Para garantir a reprodutibilidade e a comparabilidade dos resultados, as configurações de treinamento seguiram os mesmos parâmetros utilizados no artigo original do LIIF [2]. Isso incluiu a adoção das mesmas taxas de aprendizado e o otimizador específico empregado no treinamento do modelo original. Essas configurações foram implementadas para assegurar condições ideais de treinamento e avaliar o desempenho do modelo de forma consistente com os resultados previamente publicados.

C. Protocolo Experimental

O framework MMagic [11] do OpenMMLab foi escolhido para definir, treinar e validar os modelos, devido à sua simplicidade e por fornecer opções no estado da arte. Com o objetivo de replicar os experimentos conduzidos pelos autores do LIIF [2], o treinamento foi realizado utilizando o otimizador Adam [9], amplamente reconhecido por sua eficiência em cenários de aprendizado profundo. A configuração do treinamento foi estabelecida para durar 10^3 épocas, totalizando aproximadamente $2,2 \cdot 10^6$ iterações, com lotes de tamanho 16. Para a preparação dos dados, a imagem de entrada foi gerada a partir da imagem de referência original por meio da função *Bicubic Resizing* do PyTorch [12], garantindo a obtenção de amostras com resoluções adequadas para o treinamento. Além disso, a taxa de aprendizado foi programada para sofrer um decaimento multiplicativo de 0,5 a cada 200 épocas, assegurando uma convergência progressiva durante o processo de otimização. Essa configuração foi adotada para manter consistência metodológica e permitir comparações diretas com os resultados apresentados no estudo original.

IV. Experimentos e Resultados

Os resultados quantitativos apresentados nas Tabelas II e III permitem uma análise detalhada do desempenho da metodologia proposta em comparação com o método de redimensionamento da biblioteca OpenCV. A avaliação abrange três métricas amplamente utilizadas na literatura: PSNR, SSIM e MAE. Cada escala de ampliação (2x, 3x e 4x) foi analisada individualmente para entender as vantagens e limitações do modelo LIIF.

- **Escala 2x (480x480 $\rightarrow$ 960x960)**: A ampliação de 2x apresentou o melhor desempenho geral para ambas as abordagens, com o modelo LIIF superando significativamente o OpenCV. O LIIF alcançou um PSNR de 26,44 dB, enquanto o OpenCV obteve 24,97 dB, demonstrando uma reconstrução mais precisa em termos de preservação de detalhes. O SSIM do LIIF (0,7921) foi substancialmente superior ao do OpenCV (0,6598), refletindo maior similaridade estrutural e percepção visual. Além disso, o MAE do LIIF foi o menor registrado entre todas as escalas, com um valor de 0,0350, contra 0,0428 do OpenCV. Esses resultados indicam que a ampliação de 2x é menos desafiadora e permite uma reconstrução mais fiel.
- **Escala 3x (320x320 $\rightarrow$ 960x960)**: Para a ampliação de 3x, o LIIF manteve sua superioridade em todas as métricas, com PSNR de 23,72 dB e SSIM de 0,5562, em comparação aos 23,36 dB e 0,4924 do OpenCV. O MAE do LIIF também foi inferior (0,0490 contra 0,0510). No entanto, a diferença percentual entre os dois métodos diminuiu em relação à ampliação de 2x, indicando que a tarefa de reconstrução em 3x é mais desafiadora, com maiores limitações na recuperação de detalhes finos.
- **Escala 4x (240x240 $\rightarrow$ 960x960 e 96x96 $\rightarrow$ 384x384)**: A ampliação de 4x foi a mais desafiadora para ambas as abordagens, como esperado devido à maior necessidade de extrapolação de detalhes. O LIIF obteve PSNR de 22,94 dB e 26,48 dB para as configurações de entrada 240x240 e 96x96, respectivamente, superando consistentemente o OpenCV (22,62 dB e 25,62 dB). Em termos de SSIM, o LIIF também se destacou, com valores de

0,4516 e 0,5732, em comparação aos 0,4017 e 0,5206 do OpenCV. No entanto, o aumento do MAE em relação às escalas menores (0,0533 contra 0,0554 para 240x240 e 0,0337 contra 0,0366 para 96x96) evidencia a maior dificuldade em reconstruir texturas finas e padrões estruturais em ampliações maiores.

Além dos indicadores quantitativos, a avaliação qualitativa reforça a superioridade do modelo treinado. As imagens reconstruídas pelo modelo apresentam maior nitidez e preservação de detalhes, enquanto as geradas pela função 'resized' do OpenCV frequentemente exibem artefatos visuais e suavização de bordas. Essa diferença pode ser observada nas Figuras 2 e 3, e é perceptível em regiões de alto contraste, onde o modelo treinado é capaz de recuperar texturas e padrões com maior precisão. Essas observações confirmam que as métricas objetivas refletem a qualidade visual observada nas imagens ampliadas, o que é crucial para aplicações práticas que demandam fidelidade visual.

Finalmente, a análise dos desempenhos entre diferentes treinamentos revela a influência de parâmetros específicos na qualidade dos resultados. Apesar das variações, o modelo treinado mostrou consistência em superar a abordagem do OpenCV em todas as configurações avaliadas. Observa-se, ainda, que os gráficos das métricas indicam uma tendência de melhoria contínua ao longo das iterações de treinamento. Isso evidencia o sucesso do ajuste do modelo e reforça a necessidade de comparações adicionais com outros algoritmos para consolidar sua eficácia.

Entrada	Saída	Escala	PSNR (dB)	SSIM	MAE
480x480	960x960	2x	26.4426	0.7921	0.0350
320x320	960x960	3x	23.7158	0.5562	0.0490
240x240	960x960	4x	22.9405	0.4516	0.0533
96x96	384x384	4x	26.4772	0.5732	0.0337

TABLE II: Resultados da metodologia

Entrada	Saída	Escala	PSNR (dB)	SSIM	MAE
480x480	960x960	2x	24.9713	0.6598	0.0428
320x320	960x960	3x	23.3649	0.4924	0.0510
240x240	960x960	4x	22.6248	0.4017	0.0554
96x96	384x384	4x	25.6229	0.5206	0.0366

TABLE III: Resultados da função *resized* da OpenCV

(a) Entrada (b) Original

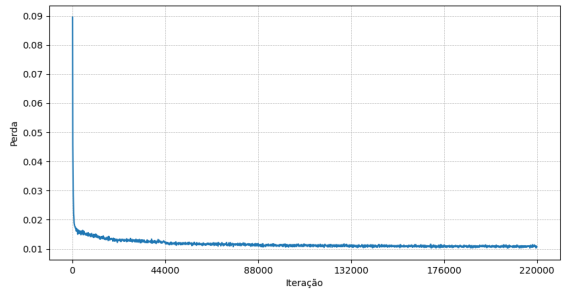
(c) Predita (d) Redimensionada

Fig. 3: Exemplo de resultado obtido com conjunto de validação

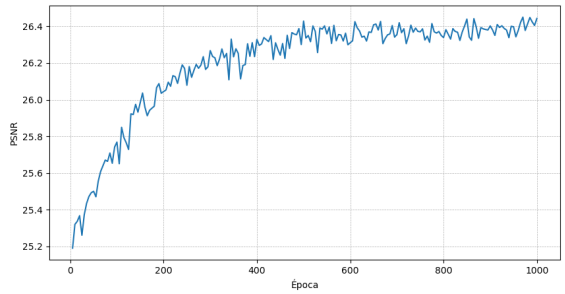
(a) Entrada (b) Original

(c) Predita (d) Redimensionada

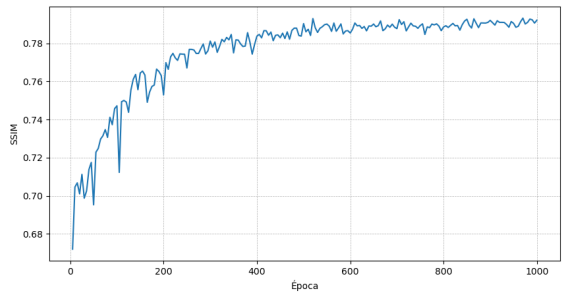
Fig. 4: Exemplo de resultado obtido com conjunto de validação



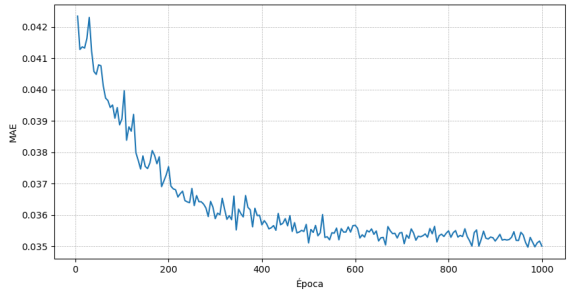
(a) Perda (loss)



(b) PSNR

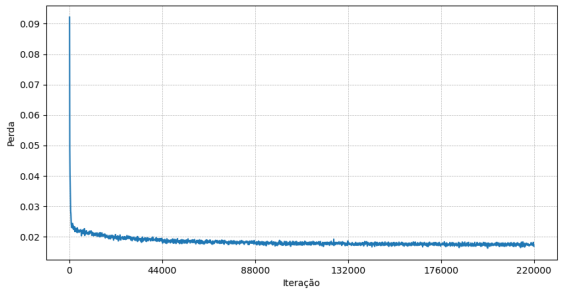


(c) SSIM

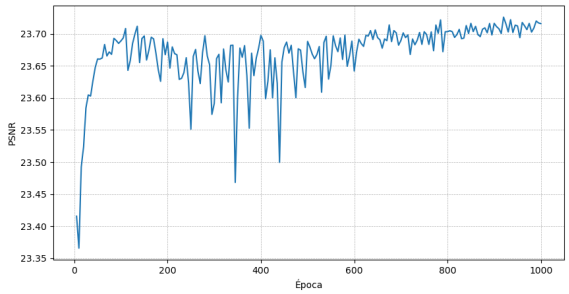


(d) MAE

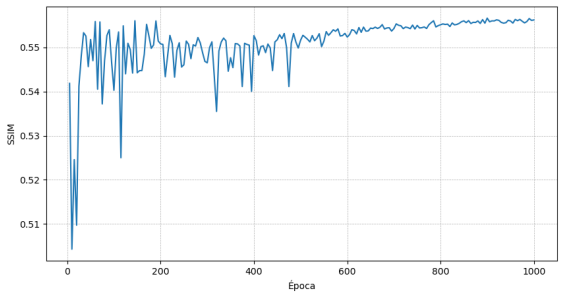
Fig. 5: Resultados do treinamento (1)



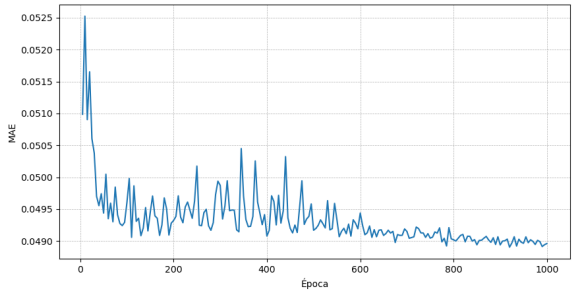
(a) Perda (loss)



(b) PSNR

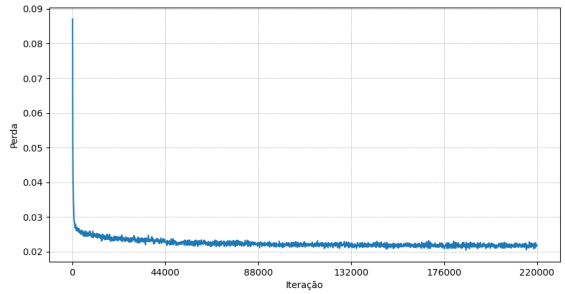


(c) SSIM

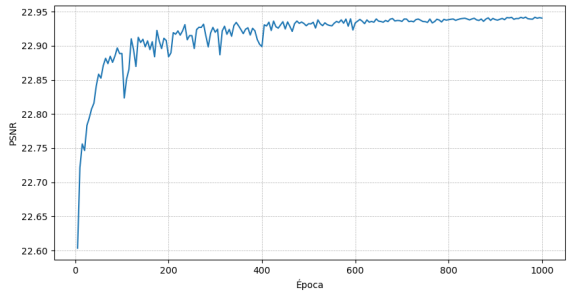


(d) MAE

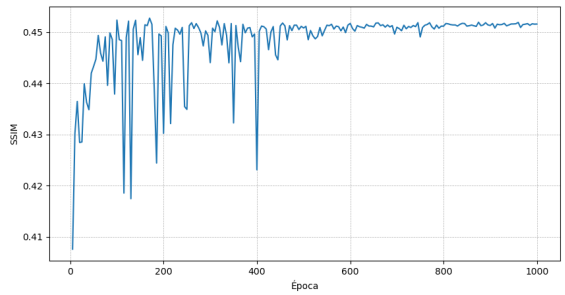
Fig. 6: Resultados do treinamento (2)



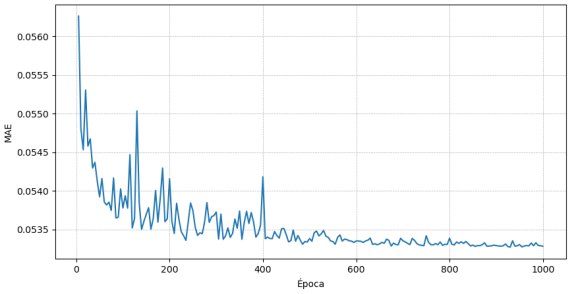
(a) Perda (loss)



(b) PSNR

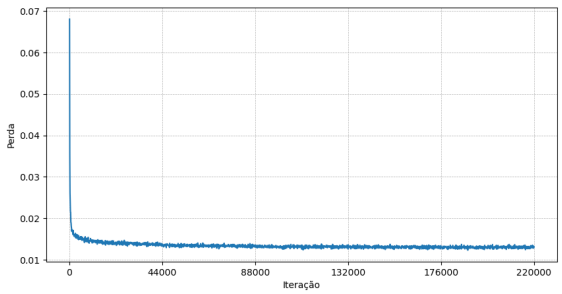


(c) SSIM

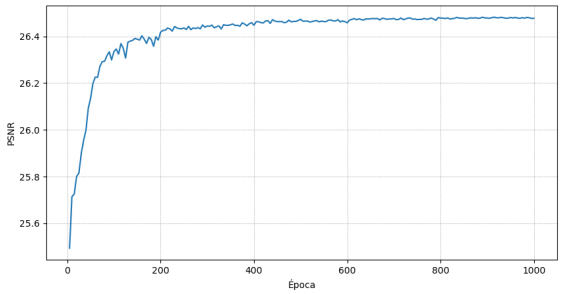


(d) MAE

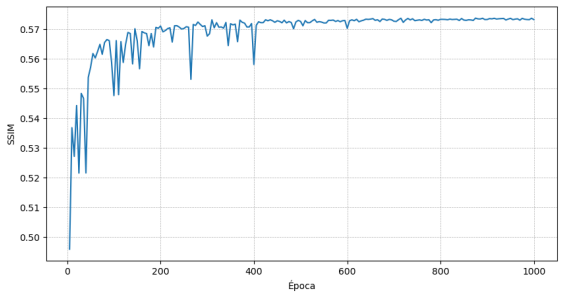
Fig. 7: Resultados do treinamento (3)



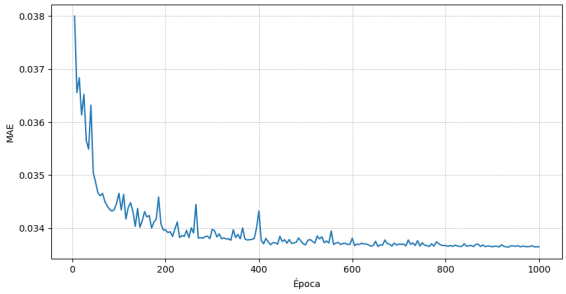
(a) Perda (loss)



(b) PSNR



(c) SSIM



(d) MAE

Fig. 8: Resultados do treinamento (4)

V. Conclusão

No quesito quantitativo, o sucesso do treinamento do modelo foi evidenciado pelos resultados superiores em comparação ao algoritmo *resized* do OpenCV, no qual as métricas PSNR, SSIM e MAE do modelo treinado demonstraram que o mesmo é mais preciso do que o algoritmo tradicional. No quesito qualitativo, o modelo claramente obteve melhores resultados de Super-Resolução quando comparado ao OpenCV. Esta melhoria pode tornar imagens de baixa qualidade em imagens propensas a serem utilizadas por datasets em outros projetos de visão computacional. Estes resultados confirmam tanto a eficácia no processo de treinamento quanto a capacidade do modelo em gerar imagens de maior qualidade e precisão.

References

[1] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[2] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8628–8638, 2021.

[3] Guilherme Defalque, Pedro Arfux, Marcio Pache, Gumercindo Franco, and Ricardo Santos. A dataset for pasture parameter estimation based on satellite remote sensing and weather variables. *Data in Brief*, 53:110206, 2024.

[4] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.

[5] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[7] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.

[8] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1637–1645, 2016.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017.

[11] MMagic Contributors. MMagic: OpenMMLab multimodal advanced, generative, and intelligent creation toolbox. https://github.com/open-mmlab/mmagic, 2023.

[12] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

[13] César Sousa Santos, Cássio Marcelo Silva Castro, and Tiago Ramos Ribeiro. Aplicações de imagens de satélite de alta resolução no planejamento urbano: o caso do cadastro técnico multifinalitário de mata de são joão, bahia. *XV Simpósio Brasileiro de Sensoriamento Remoto(SBSR)*, 5 2011.

[14] X. Wang, J. Yi, J. Guo, Y. Song, J. Lyu, J. Xu, W. Yan, J. Zhao, Q. Cai, and H. Min. A review of image super-resolution approaches based on deep learning and applications in remote sensing. *Remote Sensing*, 14:5423, 2022.

Declaração de uso de IA generativa e tecnologias assistidas por IA no processo de escrita

Durante a preparação deste trabalho, os autores utilizaram o ChatGPT para auxílio na coesão do texto. Após o uso dessa ferramenta, os autores revisaram e editaram o conteúdo conforme necessário e têm total responsabilidade pelo conteúdo da publicação.