

## DEEP LEARNING APLICADO NA DETECÇÃO DE CORPOS D'ÁGUA EM IMAGENS DE VANT DO PANTANAL BRASILEIRO

**Lucas Yuri Dutra de Oliveira**

Universidade Federal de Mato Grosso do Sul, Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia,  
Campo Grande, MS  
[lucas.oliveira@ufms.br](mailto:lucas.oliveira@ufms.br)

**João Lucas Aparecido Rocha Paes**

Universidade Federal de Mato Grosso do Sul, Faculdade de Computação, Campo Grande, MS, Brasil  
[joao.paes@ufms.br](mailto:joao.paes@ufms.br)

**Maximilian Jaderson de Melo**

Universidade Federal de Mato Grosso do Sul, Faculdade de Computação, Campo Grande, MS, Brasil  
Instituto Federal de Mato Grosso do Sul, Naviraí, MS, Brasil  
[maximilian.melo@ifms.edu.br](mailto:maximilian.melo@ifms.edu.br)

**Maxwell da Rosa Oliveira**

Universidade Federal de Mato Grosso do Sul, Instituto de Biociências, Campo Grande, MS, Brasil  
Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Belo Horizonte, MG, Brasil  
[max.oliveira2102@gmail.com](mailto:max.oliveira2102@gmail.com)

**Eveline Terra Bezerra**

Universidade Federal de Mato Grosso do Sul, Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia,  
Campo Grande, MS  
[eveline.terra@ufms.br](mailto:eveline.terra@ufms.br)

**Ana Paula Marques Ramos**

Universidade Estadual Paulista, Departamento de Cartografia, Presidente Prudente, SP, Brasil  
[marques.amos@unesp.br](mailto:marques.amos@unesp.br)

**Jonathan Li**

Universidade de Waterloo, Departamento de Geografia e Gestão Ambiental, Waterloo, Ontário, Canadá  
[junli@uwaterloo.ca](mailto:junli@uwaterloo.ca)

**Geraldo Alves Damasceno Júnior**

Universidade Federal de Mato Grosso do Sul, Instituto de Biociências, Campo Grande, MS, Brasil  
[geraldodamasceno@gmail.com](mailto:geraldodamasceno@gmail.com)

**Wesley Nunes Gonçalves**

Universidade Federal de Mato Grosso do Sul, Faculdade de Computação, Campo Grande, MS, Brasil  
[wesley.goncalves@ufms.br](mailto:wesley.goncalves@ufms.br)

**José Marcato Júnior**

Universidade Federal de Mato Grosso do Sul, Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia,  
Campo Grande, MS  
[jose.marcato@ufms.br](mailto:jose.marcato@ufms.br)

### RESUMO

O Pantanal, a maior planície alagada contínua do mundo, enfrenta desafios de preservação devido ao ciclo sazonal de inundações e intervenções humanas. Para melhor compreender e preservar esse bioma, sistemas de monitoramento são essenciais, e o uso de técnicas de sensoriamento remoto combinadas com machine learning avançado surge como uma estratégia promissora. Este estudo investigou modelos de deep learning para classificação de corpos d'água em imagens de UAV (veículo aéreo não tripulado) do Pantanal. As imagens foram capturadas com a câmera MAVIC 2 Air, com resolução espacial de 3 cm. Foram comparados modelos de deep learning, como InterImage, DeepLabv3+ e SegFormer, para avaliar suas capacidades de classificação. Um protocolo foi estabelecido para a avaliação, considerando métricas como Intersection over Union (IoU) e Dice. O SegFormer apresentou os melhores resultados, com IoU de 96,16%, Recall de 97,85%, Precisão de 99,46% e F1 Score de 98,04%. Embora apresentassem métricas menores, DeepLabv3+ e InterImage também mostraram desempenho robusto. Todos os modelos apresentaram resultados satisfatórios, mas algumas dificuldades foram observadas na identificação precisa dos corpos d'água.

---

**Palavras-chave:** Sensoriamento Remoto. Inteligência Artificial. Classificação de Imagem.

## DEEP LEARNING FOR DETECTING WATER BODIES IN UAV IMAGERY OF THE BRAZILIAN PANTANAL

### ABSTRACT

The Pantanal, the largest continuous flooded plain in the world, faces preservation challenges due to the seasonal flooding cycle and human interventions. To better understand and preserve this biome, monitoring systems are essential, and the use of remote sensing techniques combined with advanced machine learning emerges as a promising strategy. This study investigated deep learning models for water body segmentation in UAV (unmanned aerial vehicle) images of the Pantanal. The images were captured using the MAVIC 2 Air camera, with a spatial resolution of 3 cm. Deep learning models such as InterImage, DeepLabv3+, and SegFormer were compared to evaluate their segmentation capabilities. A protocol was established for evaluation, considering metrics such as Intersection over Union (IoU) and Dice. SegFormer showed the best results, with an IoU of 96.16%, Recall of 97.85%, Precision of 99.46%, and an F1 Score of 98.04%. Although DeepLabv3+ and InterImage presented lower metrics, they also demonstrated robust performance. All models produced satisfactory results, but some difficulties were observed in accurately identifying water bodies.

**Keywords:** Remote Sensing. Artificial Intelligence. Image Classification.

---

### INTRODUÇÃO

O Pantanal, reconhecido como a maior planície alagada contínua do mundo, abrange aproximadamente 138.183 km<sup>2</sup> (Silva; Abdon, 1998). Localizado principalmente no estado de Mato Grosso do Sul, que abriga 65% de sua área, e também em 35% do estado de Mato Grosso, essa região apresenta um ciclo sazonal de inundações que influencia significativamente sua cobertura vegetal. É importante ressaltar que as zonas úmidas são um dos ecossistemas com maior biodiversidade, desempenhando funções essenciais de serviços ecológicos e de armazenamento de carbono (SUN et al., 2024). Nesse contexto, a implementação de sistemas de monitoramento é fundamental para a preservação e a compreensão do bioma.

A sazonalidade dos eventos de inundação no bioma do Pantanal é um fenômeno intrínseco a suas dinâmicas naturais. No entanto, é pertinente enfatizar que as intervenções antrópicas e as crescentes mudanças climáticas desempenham um papel significativo na modificação do ciclo hidrológico desse ecossistema. A implementação de sistemas de monitoramento é imperativa para compreender as complexas dinâmicas hidrológicas do Pantanal e registrar e analisar as transformações ocorridas ao longo do tempo, fornecendo subsídios para a formulação de estratégias de conservação e gestão adequadas desse importante bioma.

Dada a vasta extensão do Pantanal, a adoção de técnicas de monitoramento por sensoriamento remoto surge como uma estratégia ideal. Esse método permite cobrir áreas extensas de maneira eficiente, reduzindo os custos e o tempo associados à coleta de dados em campo. A integração de dados obtidos por meio de sensoriamento remoto com técnicas avançadas de aprendizado de máquina, como o deep learning, para a classificação de imagens, demonstrou ser uma estratégia promissora e confiável. Essa sinergia entre tecnologias permite obter resultados robustos e precisos, proporcionando uma compreensão mais detalhada da situação ambiental e dos processos em andamento na região.

A obtenção de dados por meio de missões espaciais é um método amplamente utilizado, resultando em um vasto banco de dados composto por imagens orbitais com uma ampla gama de aplicações em pesquisas e monitoramento ambiental. Diversos estudos na literatura acadêmica exploraram a integração de técnicas de sensoriamento remoto com metodologias de aprendizado profundo para o

mapeamento de corpos d'água. Por exemplo, Qi et al. (2024) utilizaram uma abordagem híbrida para extrair lagos, aproveitando o conjunto de dados do Planalto Tibetano, que inclui imagens do espectro visível adquiridas por meio do Google Earth. Da mesma forma, Yang et al. (2020) implementaram um modelo baseado no Mask R-CNN para automatizar a detecção e a classificação de corpos d'água, utilizando também imagens do Google Earth. No contexto do mapeamento de áreas alagadas no Pantanal, Milien et al. (2023) exploraram o potencial de imagens ópticas do satélite Sentinel-2 para determinar a extensão e os padrões sazonais de inundação na região norte do Pantanal. Da mesma forma, Evans et al. (2010) utilizaram dados multitemporais do ALOS/PALSAR (banda L) e do RADARSAT-2 (banda C) para mapeamento do Pantanal, destacando a utilidade desses dados para o mapeamento de uso e cobertura do solo, bem como para a análise de padrões de inundação.

No entanto, apesar dos benefícios proporcionados por essas imagens, é importante notar que sua resolução espacial pode ser limitada, especialmente ao detalhar ambientes complexos. Como alternativa viável, as imagens obtidas por veículos aéreos não tripulados (VANT) surgem como solução para esse problema. O uso de VANTs para a aquisição de dados tem se destacado cada vez mais na pesquisa ambiental contemporânea. Devido ao seu baixo custo operacional, esses dispositivos têm sido reconhecidos como uma ferramenta eficaz e versátil, capaz de fornecer imagens de alta resolução aplicáveis a diversas finalidades ambientais. A capacidade dos VANTs de sobrevoar áreas inacessíveis ou monitorar regiões extensas com detalhes os torna especialmente úteis para a obtenção de dados precisos e abrangentes em ecossistemas como o Pantanal.

No campo das imagens de VANT, Li et al. (2021) adotaram uma solução inovadora ao aprimorar o modelo U-Net tradicional. Eles adicionaram conexões circulares em forma de "S" para aumentar as conexões entre as camadas intermediárias da rede, ajudando a reter mais características da imagem e a melhorar a precisão na identificação. O estudo combinou redes neurais profundas com características visuais e explorou como as transformadas wavelet e as matrizes de ocorrência de níveis de cinza (GLCM) impactam a extração de corpos d'água. Essa combinação, quando utilizada com o modelo SU-Net aprimorado, melhorou significativamente a capacidade de distinguir e identificar objetos mistos e escuros, como sombras e vegetação densa. Wang et al. (2021) concentraram-se na extração de corpos d'água com o MobileNetV2, um modelo leve aplicado a imagens de sensoriamento remoto de alta resolução obtidas por vários sensores, incluindo os dos satélites GF-2 e WorldView-2, bem como por meio de imagens ortofotográficas de VANT. Eles compararam o desempenho desse modelo com o de métodos tradicionais, como a máquina de vetores de suporte (SVM), o Random Forest (RF) e o U-Net, para avaliar sua eficácia na extração de corpos d'água.

A comparação entre imagens obtidas por missões espaciais e aquelas capturadas por VANT mostra que estas últimas têm uma resolução espacial significativamente maior, oferecendo um nível substancialmente maior de detalhe e precisão. Por sua vez, Khaliq et al. (2019) investigaram o desempenho de imagens de satélite, especificamente as do Sentinel-2, em comparação com as obtidas por VANT na agricultura. Os resultados da pesquisa indicaram que as imagens de VANT são mais confiáveis para o mapeamento, devido ao seu maior nível de detalhamento em relação às imagens de satélite.

No mesmo contexto, Messina et al. (2020) utilizaram uma lógica semelhante, empregando imagens de VANT em conjunto com imagens orbitais — incluindo as do PlanetScope e do Sentinel-2 — para monitorar culturas agrícolas. Os resultados mostraram que, embora a combinação de imagens orbitais com imagens de VANT seja útil para a análise, a menor resolução espacial das imagens orbitais limita sua capacidade de compreender áreas mais detalhadas, perceptíveis nas imagens de VANT.

A literatura apresenta trabalhos que abordam a aplicação de deep learning em imagens de veículos aéreos não tripulados (VANTs) para o monitoramento de recursos hídricos. Munawar et al. (2021), por exemplo, utilizaram redes neurais convulsionais (CNNs) em imagens de drones para detectar inundações em uma região ao norte do Paquistão. Samaranayake et al. (2023) propõem a identificação de corpos d'água a partir de imagens de VANT. A abordagem para câmeras frontais foi baseada em uma arquitetura UNet modificada, que incorporou Módulos de Atenção à Reflexão e foi denominada UNet-8-RAU. Essa arquitetura segmenta a água com base em suas propriedades refletivas. Para

validar o método, os autores utilizaram o DeepLabV3+ como um dos modelos de referência, e o UNet-8-RAU demonstrou desempenho superior nos conjuntos de dados avaliados.

No contexto do Pantanal, há uma lacuna significativa na literatura sobre estudos que combinam o uso de técnicas de aprendizado profundo com imagens de VANT para o mapeamento de corpos d'água. A integração dessas abordagens representa uma área promissora e ainda inexplorada, cuja implementação pode proporcionar avanços substanciais na precisão e na eficiência do monitoramento ambiental na região. A aplicação do aprendizado profundo a imagens de alta resolução obtidas por VANTs tem o potencial de superar as limitações inerentes às imagens orbitais tradicionais, oferecendo maior detalhamento e melhor capacidade de detectar características hidrológicas complexas.

Ao enfrentar desafios relacionados à classificação de imagens, é imperativo adotar modelos capazes de realizar duas tarefas distintas: extração de características da imagem e interpretação/geração subsequentes. Portanto, é crucial selecionar modelos especializados, adotando uma abordagem mais generalista na fase de codificação (encoder), responsável pela extração de características, e um decodificador especializado na resolução específica do problema em questão. Entre os principais modelos de cada ramo do deep learning, destacam-se as redes neurais convolucionais, os transformadores e as convoluções deformáveis.

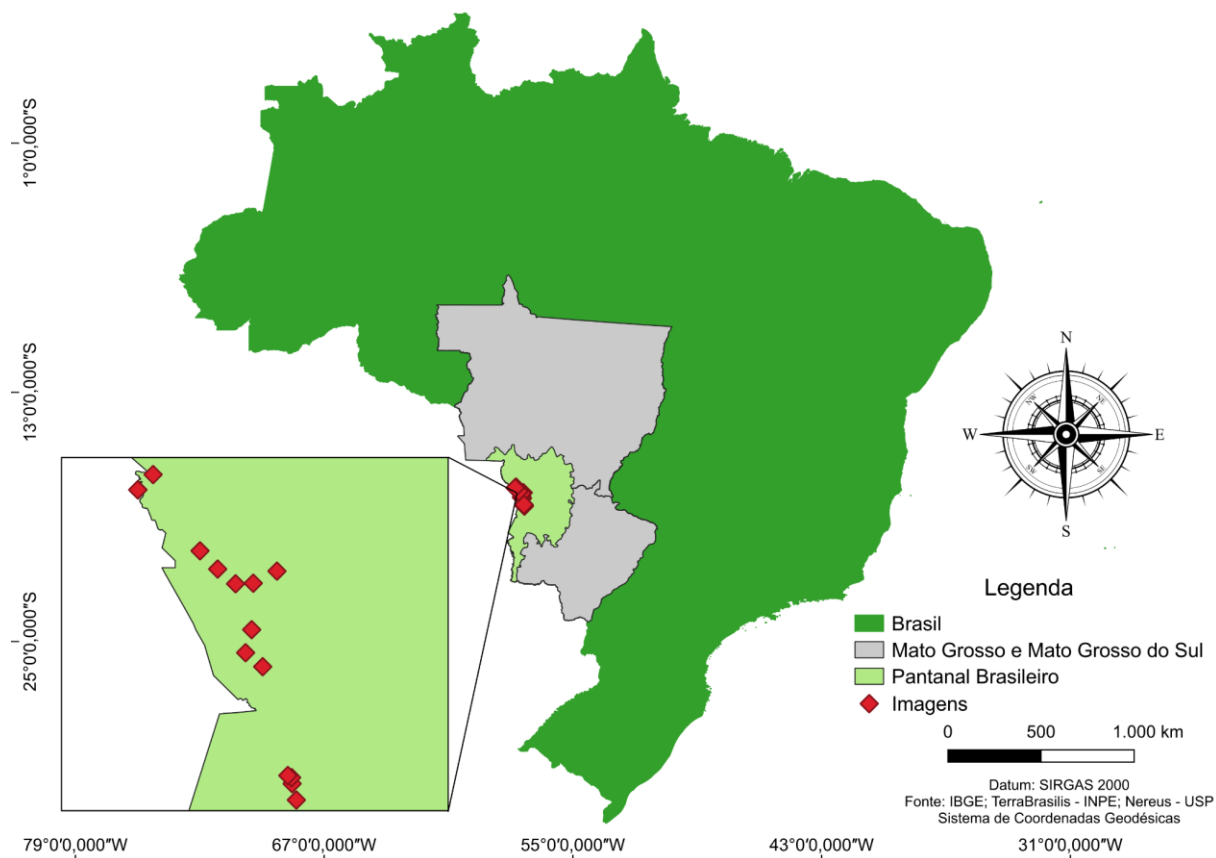
Este trabalho tem como objetivo investigar a viabilidade e a eficácia das técnicas de deep learning, quando aplicadas em conjunto com dados de sensoriamento remoto de VANT, para a classificação precisa de corpos d'água no Pantanal. O monitoramento remoto de áreas alagadas enfrenta diversos desafios, como a presença de vegetação sobre os corpos d'água, a resposta espectral variável dessas áreas e a limitação da resolução espacial das imagens orbitais, que frequentemente não conseguem detectar corpos d'água de pequeno porte (GALLANT, 2015). O conjunto de dados resultante deste estudo será disponibilizado publicamente para facilitar o avanço das pesquisas nessa área. Ao tornar o conjunto de dados acessível, espera-se que os desafios enfrentados durante o mapeamento de corpos d'água possa ser abordados de maneira mais eficaz por pesquisadores e profissionais interessados. Essa iniciativa tem o potencial de estimular o desenvolvimento de novas metodologias e abordagens para a segmentação e a classificação de corpos d'água, contribuindo, assim, para uma melhor compreensão e gestão dos recursos hídricos em diferentes ambientes.

## MATERIAIS E MÉTODOS

### *Área de Estudo*

A planície alagada do Pantanal, reconhecida como a maior do mundo, constitui um elemento distintivo da paisagem geográfica brasileira. Este ecossistema único abrange uma vasta extensão territorial nos estados de Mato Grosso e Mato Grosso do Sul, ultrapassando as fronteiras nacionais e englobando áreas em países vizinhos como Bolívia e Paraguai. Para adquirir informações relevantes para fins analíticos, foi gerada uma coleção de quinze ortoimagens de VANT. As localizações das imagens capturadas estão representadas na Figura 1.

Figura 1 - Pantanal brasileiro: Localização das imagens, 2024



Fonte: IBGE; TerraBrasilis - INPE; Nereus - USP. Organização: Autores, 2024.

### Imagens

Para a aquisição de imagens, utilizamos um VANT DJI MAVIC 2 Air, equipado com uma câmera RGB de resolução 4K. As imagens foram capturadas em julho de 2024. Os voos foram planejados e executados de maneira programada para garantir cobertura abrangente da área-alvo. Adotamos uma sobreposição de 80% (lateral e longitudinal) entre as imagens capturadas para assegurar a integração adequada dos dados e a cobertura completa da área de interesse. O voo foi realizado a uma altitude de 80 metros, garantindo uma resolução espacial apropriada para a análise e a interpretação das imagens obtidas. Esses parâmetros foram selecionados para maximizar a qualidade e a utilidade dos dados coletados para os fins específicos do estudo. Os ortomosaicos foram gerados no software Agisoft Metashape (AGISOFT LLC, 2024), com uma resolução espacial de 3 cm.

### Dataset

Com o uso de ortomosaicos de alta resolução obtidos por VANT sobre um grande rio coberto por regiões de vegetação de diferentes dimensões, uma das primeiras abordagens foi a separação dos conjuntos de treinamento, teste e validação. Em outras palavras, se uma imagem era cortada para treinamento, ela era usada inteiramente para esse fim, e o mesmo se aplicava a teste e validação. Foram obtidos 915 blocos de imagem, posteriormente divididos em conjuntos de treinamento, validação e teste. A divisão seguiu um protocolo padrão: aproximadamente 70% dos blocos foram alocados para o conjunto de treinamento, 15% para validação e os restantes, para teste.

Para determinar a estratégia de processamento mais eficaz e otimizada, foi selecionado um tamanho de corte de 512 x 512 pixels. Nenhuma combinação de camadas foi aplicada, pois apenas as bandas espectrais vermelha, verde e azul (RGB) estavam disponíveis para análise.

Além disso, foram aplicadas técnicas tradicionais de pré-processamento, como inversões aleatórias horizontais e verticais, para aumentar os dados durante a fase de treinamento, e técnicas de pós-processamento, como fechamento, para melhorar a precisão. As técnicas de pós-processamento foram empregadas para aumentar a acurácia e reduzir falsos positivos e negativos em áreas dominadas por determinadas classes.

## Modelos

Este estudo adota uma metodologia experimental, empregando três modelos distintos: DeepLabV3+ (CHEN et al., 2018), Segformer (XIE et al., 2021) e InternImage (WANG et al., 2023), por meio da plataforma MMSegmentation (CONTRIBUTORS, 2020). DeepLabV3+ (CHEN et al., 2018), Segformer (XIE et al., 2021) e InternImage (WANG et al., 2023), por meio da plataforma MMSegmentation (CONTRIBUTORS, 2020). Esses modelos foram selecionados com base em sua reputação e desempenho em tarefas análogas de sensoriamento remoto (OSCO et al., 2021; DU et al., 2021; LIN et al., 2023; WANG et al., 2024), abrangendo uma variedade de arquiteturas e técnicas de aprendizado profundo para explorar, de forma eficaz, a classificação de imagens no contexto específico deste estudo.

O modelo InternImage representa o estado da arte em muitos testes de benchmarking de visão computacional. Ele combina três tipos de blocos de camadas, principalmente convolucionais. Seu bloco inicial, denominado Stem, consiste em uma convolução simples 3x3, seguida por normalização linear. Esse bloco é replicado duas vezes, intercalado por uma camada de ativação GELU. O próximo bloco, chamado "Stage", é composto por dois blocos: o primeiro utiliza uma convolução deformável como entrada, e o segundo, um perceptron multicamadas. Em seguida, os dados passam por normalização linear e por conexões de salto em relação à entrada de cada bloco.

No entanto, apenas o bloco "Stem" realiza o *downsampling* dos dados. Para garantir a otimização do modelo, é essencial incluir um bloco de *downsampling*, composto por uma convolução 3x3 e normalização linear. Isso permite a propagação de múltiplas escalas de extração de dados para o decodificador posteriormente, como  $\{\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$ . InternImage é um modelo que descreve apenas um codificador, sendo necessário um modelo adicional de decodificador. Para isso, é utilizado o *MaskFormerHead*, que é responsável por gerar os patches de classificação para cada classe.

O modelo SegFormer, também de ponta, consiste em um codificador baseado em *Transformer* com um decodificador leve de perceptron multicamadas. A ideia do modelo é que, se seu *backbone* for suficientemente capaz de extrair representações suficientemente robustas dos dados, o trabalho da próxima camada será significativamente mais simples. Além disso, o modelo trabalha com uma composição de múltiplas escalas, na qual, através de cada bloco de atenção, os dados entram e saem de uma função denominada *Mixed Feed Forward*, cuja tarefa é substituir a codificação posicional.

Por fim, o DeepLabv3+ representa a evolução dos modelos DeepLab anteriores (v1, v2 e v3). Sua robusta capacidade de capturar características em múltiplas escalas por meio de um modelo totalmente convolucional contribuiu para que se tornasse um modelo extremamente popular na literatura, sendo utilizado em uma ampla gama de aplicações em vários contextos, desde problemas de benchmark tradicionais (CHEN et al., 2018) até sensoriamento remoto (OSCO et al., 2021). O codificador do DeepLabv3+ consiste na aplicação de uma série de convoluções dilatadas com diferentes taxas, compondo um módulo de agrupamento chamado pooling piramidal espacial atrous (ASPP). Na versão mais recente do modelo, o decodificador foi remodelado para que cada bloco seja composto por uma convolução 1x1, seguida pela concatenação com a respectiva saída do codificador. Em seguida, os dados passam por filtros 3x3 e um upsample, de modo que o resultado final tenha a mesma dimensão



dos dados de entrada originais. Tanto o codificador quanto o decodificador são modelados usando convoluções separáveis, a fim de minimizar o número de parâmetros exigidos pelo modelo.

### **Protocolo para Experimentos e Métricas**

Inicializar o modelo com pesos pré-treinados tornou-se uma prática comumente adotada em deep learning. Esses pesos pré-treinados funcionam como um ponto de partida eficaz, principalmente em cenários nos quais o conjunto de dados-alvo é pequeno ou específico de um domínio. Ajustar esses modelos permite a transferência da extração de características de baixo nível e possibilita que o modelo se adapte às exigências específicas da tarefa. Essa solução reduz o custo computacional e o tempo de treinamento desde o início e melhora o desempenho do modelo, pois aproveita as ricas hierarquias de características aprendidas a partir de dados visuais diversificados.

Neste estudo, todos os modelos foram inicializados com pesos pré-treinados. O modelo SegFormer utilizou o MiT-B5 como backbone, enquanto o DeepLabv3+ utilizou o ResNet-50. Tanto o DeepLabv3+ quanto o SegFormer se beneficiaram dos pesos pré-treinados no conjunto de dados ImageNet. O modelo InternImage foi treinado com o backbone UpperNet e pesos pré-treinados nos conjuntos COCO-Stuff e ADE20K.

Para avaliar os modelos de maneira concisa, foi estabelecido um protocolo padrão, de modo que os resultados expressassem os ganhos reais de uma abordagem em relação à outra. Quanto aos dados, além dos recortes, foi calculado um peso relativo para as classes, com o objetivo de suavizar o desequilíbrio entre elas. O cálculo é dado pela equação 1: para os pixels da classe rio, o peso foi de 2,5038, e para o fundo, de 0,6247.

$$\text{Peso da Classe} = \frac{\text{Pixels Totais}}{\text{Quantidade de Classes} * \text{Pixels da Classe}} \quad (1)$$

Para o treinamento dos modelos, foi utilizado o otimizador AdamW (LOSHCHILOV; HUTTER, 2017), com uma taxa de aprendizado de 0,0006. O modelo foi treinado por 80 mil iterações. O esquema de treinamento utilizou um sistema de "aquecimento" de até 20% das iterações. Nos últimos 80%, a taxa de aprendizado foi reduzida para 0,999. Também foi utilizado o amostrador Online Hard Example Mining (SHRIVASTAVA et al., 2016). Por fim, a função de perda utilizada nos experimentos foi a função de entropia cruzada binária.

Para o treinamento, foi utilizado um processador AMD Ryzen™ 7 5800X CPU @ 3.8 GHz x 8, com 32 GB de memória RAM DDR4, uma placa de vídeo NVIDIA GeForce RTX 3080 TI de 12 GB de VRAM e o sistema operacional Linux Ubuntu 20.04, com driver NVIDIA 555 e CUDA 11.8.

O framework de desenvolvimento utilizado foi o MMSegmentation 0.28.0 (CONTRIBUTORS, 2020), com Pytorch 2.0.1, MMCV 1.6.2, cuDNN 8.7.0 e Python 3.10.

As métricas de comparação quantitativa foram empregadas para determinar qual modelo foi mais vantajoso para a classificação precisa de corpos d'água no Pantanal. As métricas mais utilizadas na literatura são a Interseção sobre União (IoU) e o Dice (GARCIA-GARCIA et al., 2017; LI et al., 2020). As métricas IoU e Dice são apresentadas, respectivamente, nas Equações 2 e 3. Nelas, Verdadeiro Positivo (TP) representa os pixels em que tanto a predição quanto a máscara real apresentaram as mesmas classes. Falsos positivos (FP) são os pixels em que o modelo previu uma classe positiva, mas a máscara real foi marcada como uma classe negativa. Por fim, os falsos negativos (FN) são os pixels em que o modelo previu uma classe negativa, mas a máscara real foi marcada como uma classe positiva.

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

$$Dice = \frac{2 * TP}{2 * TP + FP + FN} \quad (3)$$

A métrica IoU consegue obter a interseção por meio do uso de TP, e, utilizando as características de falsos negativos e falsos positivos, é possível obter uma razão de união das figuras. De forma demonstrativa, o índice Dice é menos punitivo em relação aos falsos positivos, comportando-se assim mais como um índice de similaridade.

## RESULTADOS E DISCUSSÃO

A utilização de VANTs proporciona uma resolução espacial excepcional, permitindo detectar detalhes impossíveis de serem vistos pela maioria dos satélites. No entanto, essa alta resolução representa um desafio de escalabilidade para o monitoramento contínuo de uma área como a do Pantanal. O desempenho de arquiteturas de segmentação, altamente sensíveis a texturas, padrões espaciais e características espectrais, está intrinsecamente ligado à paisagem na qual foram treinadas. O Pantanal é um bioma heterogêneo, um vasto mosaico de sub-regiões com distintas fitofisnomias e regimes hídricos. Dessa forma, a aplicação direta dos modelos treinados em outras áreas, como o Pantanal do Paiaguás ou do Poconé, pode resultar em queda de desempenho, sendo necessário treiná-los novamente com um conjunto de dados mais robusto e diverso, de modo que os aspectos singulares de cada região sejam englobados.

Outro aspecto que impacta o desempenho do modelo é a sazonalidade do Pantanal. Um modelo treinado principalmente com dados da estação seca, com lâminas d'água abertas, pode falhar ao reconhecer a mesma área durante a cheia, quando ela pode estar coberta por densos bancos de macrófitas aquáticas (por exemplo, *Eichhornia crassipes*), o que leva a uma subestimação da área inundada (falsos negativos). Dessa maneira, a criação de um conjunto de dados multitemporal, que inclua a variabilidade de um ciclo hidrológico completo, é um passo essencial para o desenvolvimento de um sistema de monitoramento resiliente e adaptado às dinâmicas pantaneiras.

### Análise Quantitativa

Esta seção fornecerá orientações sobre os resultados das métricas e suas inferências nas previsões visuais dos modelos, permitindo assim comparar e entender todas as diferenças entre os modelos. A Tabela 1 apresenta os resultados da IoU, *Recall*, Precisão e Índice F1 para cada modelo.

Tabela 1 - Resultados quantitativos dos modelos

MODELO	IOU	RECALL	PRECISÃO	ÍNDICE F1	Parâmetros (milhões)
SegFormer	96.16%	97.85%	99.46%	98.04%	81.97
DeepLabv3+	95.39%	98.31%	99.03%	97.61%	62.57
InterImage	92.21%	98.82%	99.03%	95.95%	128.27

Fonte: Autores, 2024.



O modelo SegFormer obteve desempenho superior nas métricas de classificação, alcançando um índice de IoU de 96,16%. Esse alto IoU reflete uma sobreposição extremamente precisa entre a classificação prevista e a real, o que é crucial para aplicações que exigem alta fidelidade na delimitação de objetos. Estudos recentes, como o de Tripathy e Mishra (2024), demonstram a robustez e a precisão das arquiteturas Transformer, incluindo o SegFormer, em análises detalhadas de corpos d'água e de vegetação aquática, nas quais podem ser necessários ajustes específicos devido a sobreposições espectrais. Taye (2023) também discute as vantagens dos modelos Transformer sobre as CNNs, como o DeepLabv3+, na redução de interferências de ruído, o que apoia a escolha do SegFormer para o presente estudo.

As redes DeepLab introduziram camadas de convolução dilatadas em sua arquitetura inicial (Yu; Koltun, 2015) e sua versão mais recente, a DeepLabv3+ (Chen et al., 2018), utiliza essas camadas para ampliar o campo receptivo sem perda de resolução, o que é vantajoso em cenários com alta variabilidade espectral. Neste estudo, o DeepLabv3+ obteve um IoU de 95,39%, comprovando sua eficácia em segmentação semântica, embora tenha se mostrado ligeiramente inferior ao SegFormer. Por outro lado, o modelo InterImage apresentou um IoU de 92,21%, o mais baixo entre os três, mas ainda assim um resultado respeitável no contexto das tecnologias atuais de classificação.

O recall, métrica útil para avaliar a capacidade de identificar corretamente as instâncias positivas, também foi analisado. Nesse aspecto, o InterImage se destacou, apresentando um recall de 98,82%. Isso sugere que ele é capaz de detectar quase todas as instâncias positivas, o que é uma característica valiosa em aplicações que priorizam a detecção completa, como em análises médicas e vigilância. O DeepLabv3+ obteve um recall próximo, de 98,31%, e o SegFormer, com 97,85%, também demonstrou boa capacidade de identificação de instâncias, embora um pouco inferior à dos outros modelos.

Na métrica de precisão, que mede a proporção de verdadeiros positivos em relação ao total de positivos preditivos, todos os modelos obtiveram resultados notáveis. O SegFormer obteve a maior precisão (99,46%), seguido de perto pelo DeepLabv3+ e pelo InterImage, com 99,03%. Esses resultados destacam a robustez dos modelos em tarefas de classificação.

O índice F1, que equilibra precisão e recall, indicou que o SegFormer possui o melhor desempenho geral, com valor de 98,04%, mostrando-se tanto preciso quanto sensível, sendo ideal para a maioria das aplicações. O DeepLabv3+ obteve um índice F1 de 97,61%, demonstrando bom equilíbrio. Por sua vez, o InterImage, apesar de suas altas taxas de precisão e recall, obteve um índice F1 mais baixo (95,95%), possivelmente devido ao seu IoU inferior.

A DeepLabv3+ tem a menor quantidade de parâmetros entre os modelos avaliados, o que indica um menor custo computacional durante o treinamento. No entanto, a diferença de aproximadamente 20 milhões de parâmetros em relação ao SegFormer não é suficiente para justificar sua escolha, pois ele apresenta um desempenho quantitativo inferior. Por sua vez, o InterImage possui mais do que o dobro de parâmetros da DeepLabv3+, mas, mesmo assim, apresenta resultados quantitativamente inferiores, o que indica um comprometimento da eficiência sem ganhos significativos em desempenho.

Em suma, o SegFormer se destacou como a melhor escolha entre os três modelos analisados, especialmente em termos de IoU e índice F1. Trabalhos como o de Lin et al. (2023), que aplicaram o SegFormer na segmentação de áreas úmidas costeiras com imagens Sentinel-2, confirmam sua superioridade em precisão, recall e índice F1 em relação ao DeepLabv3+, U-Net e Segnext.

O DeepLabv3+ surge como uma alternativa sólida, com alto recall e precisão, enquanto o InterImage, apesar do IoU inferior, se destaca no recall. Em comparação com outros modelos, como a U-Net, o DeepLabv3+ demonstrou maior precisão na classificação da cobertura do solo (HERLAWATI; HANDAYANTO, 2024). Para aplicações em recursos hídricos, Li et al. (2019) observaram que o DeepLabv3+ possui elevada acurácia na identificação de corpos hídricos em imagens de alta resolução.

A escolha do modelo ideal depende das exigências específicas de cada aplicação, como a precisão na detecção ou a capacidade de identificar todas as instâncias positivas. Isso evidencia a importância de

uma análise cuidadosa das métricas de desempenho para a seleção do modelo mais adequado em cenários práticos.

É importante destacar que a comparação dos resultados deste estudo com outros trabalhos na literatura apresenta limitações significativas, pois há uma escassez de estudos que utilizem uma metodologia similar à nossa, tanto na avaliação comparativa entre os modelos analisados quanto no uso do mesmo conjunto de dados. A ausência de estudos anteriores com o mesmo enfoque, especialmente no contexto de segmentação de corpos d'água e vegetação aquática utilizando arquiteturas Transformer, como o SegFormer, dificulta a criação de uma base de comparação robusta que corrobore ou contrarie nossos achados. A ausência de trabalhos comparáveis ressalta a originalidade e a relevância deste estudo, mas também impõe desafios, pois limita a possibilidade de validar os resultados com referências preexistentes. Futuras pesquisas devem considerar abordagens metodológicas e conjuntos de dados similares para ampliar a base de conhecimento na área.

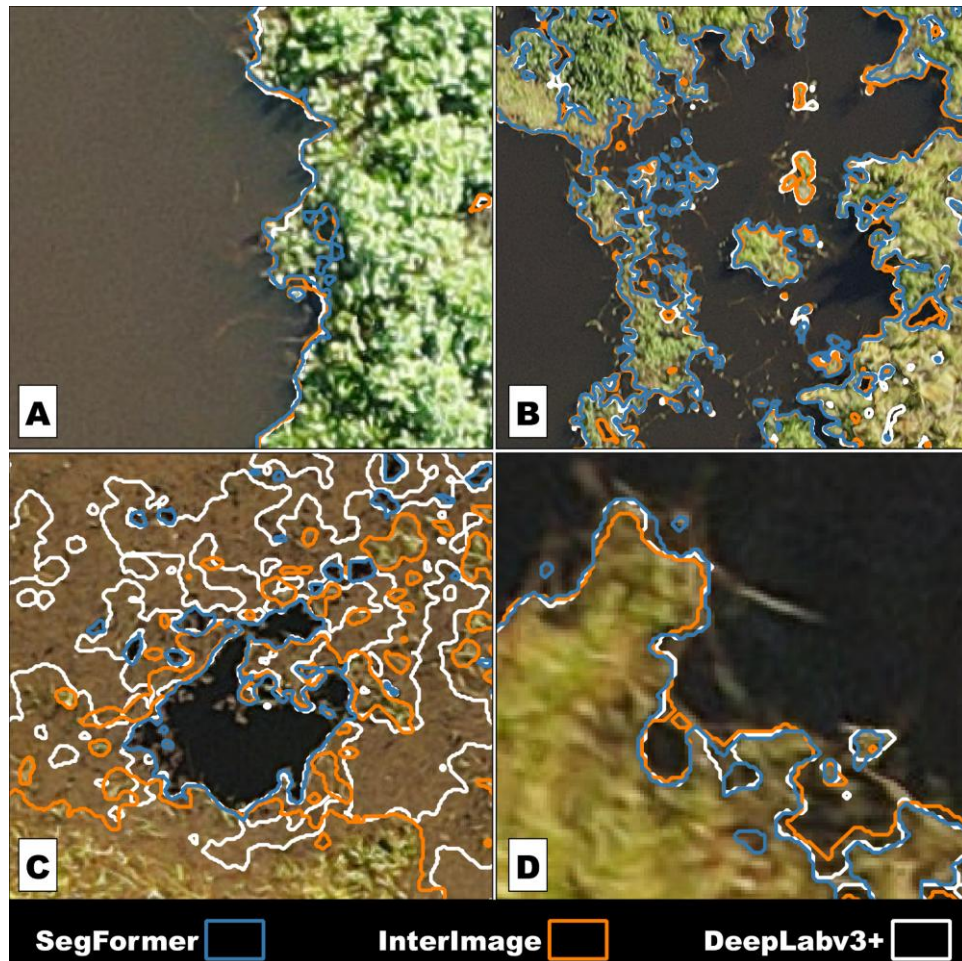
### **Análise Qualitativa**

Conforme ilustrado na Figura 2, os resultados qualitativos oferecem uma visão complementar aos resultados quantitativos, que revelam o desempenho superior do modelo SegFormer em diversas métricas de avaliação. No entanto, uma análise qualitativa mais aprofundada sugere que, em termos de qualidade visual e de detalhes discerníveis nas classificações, os modelos DeepLabv3+ e InterImage superam o SegFormer. Esse contraste entre os resultados quantitativos e qualitativos evidencia uma limitação fundamental das métricas tradicionais de avaliação, que frequentemente não captam as nuances e complexidades das abordagens de classificação, principalmente em contextos nos quais a precisão visual é essencial.

Na análise da detecção de corpos d'água, todos os modelos apresentaram desempenho satisfatório, embora com variações significativas em diferentes áreas de amostragem. Na Área 1, que inclui as amostras A e B, o SegFormer alcançou os melhores resultados quantitativos, mas falhou em capturar detalhes importantes identificados pelo InterImage e pelo DeepLabv3+. Na amostra A, o SegFormer demonstrou uma tendência preocupante de classificar elementos não aquáticos como água, interpretando sombras de vegetação como áreas hídricas. Essa limitação pode ser atribuída às respostas espectrais semelhantes entre a vegetação e a água, o que revela uma vulnerabilidade do modelo em cenários nos quais as características espectrais se sobrepõem.

Em contrapartida, na amostra B, o SegFormer apresentou a menor tendência entre os modelos de identificar erroneamente objetos menores não aquáticos nas áreas categorizadas como aquáticas. Isso sugere uma força do modelo em evitar a inclusão de detecções falsas positivas, mesmo que tenha havido um sacrifício na captura de detalhes visuais em certos contextos. Por outro lado, tanto o InterImage quanto o DeepLabv3+ falharam em distinguir áreas sem água, mesmo diante de respostas espectrais diferentes, o que evidencia uma sensibilidade às características visuais que os modelos podem ter explorado mais efetivamente.

Figura 2 - Amostras das imagens com a delimitação da detecção de cada modelo



Fonte: Autores, 2024.

Na análise da detecção de corpos d'água, todos os modelos apresentaram desempenho satisfatório, embora com variações significativas em diferentes áreas de amostragem. Na Área 1, que inclui as amostras A e B, o SegFormer alcançou os melhores resultados quantitativos, mas falhou em capturar detalhes importantes identificados pelo InterImage e pelo DeepLabv3+. Na amostra A, o SegFormer demonstrou uma tendência preocupante de classificar elementos não aquáticos como água, interpretando sombras de vegetação como áreas hídras. Essa limitação pode ser atribuída às respostas espectrais semelhantes entre a vegetação e a água, o que revela uma vulnerabilidade do modelo em cenários nos quais as características espectrais se sobrepõem.

Em contrapartida, na amostra B, o SegFormer apresentou a menor tendência entre os modelos de identificar erroneamente objetos menores não aquáticos nas áreas categorizadas como aquáticas. Isso sugere uma força do modelo em evitar a inclusão de detecções falsas positivas, mesmo que tenha havido um sacrifício na captura de detalhes visuais em certos contextos. Por outro lado, tanto o InterImage quanto o DeepLabv3+ falharam em distinguir áreas sem água, mesmo diante de respostas espectrais diferentes, o que evidencia uma sensibilidade às características visuais que os modelos podem ter explorado mais efetivamente.

Na área 2, que inclui as amostras C e D, o desempenho visual dos modelos InterImage e DeepLabv3+ foi inferior ao do SegFormer. Na amostra D, todos os modelos apresentaram desempenhos

relativamente próximos e satisfatórios, mas o DeepLabv3+ se destacou pela precisão visual. A incapacidade do InterImage de identificar certos detalhes e a tendência do SegFormer de classificar incorretamente áreas não aquáticas como água ilustram a complexidade da tarefa de classificação em ambientes aquáticos. Entretanto, na amostra C, o SegFormer superou significativamente o DeepLabv3+ e o InterImage ao identificar corretamente as áreas de água, enquanto os outros dois modelos as consideraram incorretamente como aquáticas.

Apesar de apresentar uma maior taxa de erros visuais, o SegFormer conseguiu manter um desempenho médio geral mais alto, indicando que, em um contexto de avaliação de desempenho, a precisão geral pode ser considerada superior. Por outro lado, o InterImage e o DeepLabv3+, embora tenham cometido menos erros, resultaram em equívocos mais graves, que culminaram em um desempenho médio inferior ao do SegFormer. Esse cenário levanta questões importantes sobre a definição do modelo ideal, que deve considerar as necessidades específicas da aplicação em questão.

A escolha do modelo ideal é complexa e depende de fatores como a priorização entre a alta precisão e a capacidade de detectar todas as instâncias positivas. Por exemplo, em aplicações onde a detecção precisa de corpos d'água é crítica, pode-se preferir um modelo que mostre um desempenho qualitativo robusto, mesmo que quantitativamente inferior. Em contrapartida, em cenários onde a precisão global é priorizada, o *SegFormer* pode ser a escolha preferida, dado seu desempenho superior nas métricas de avaliação. Assim, a integração de análises qualitativas e quantitativas é fundamental para uma compreensão mais abrangente do desempenho dos modelos de classificação, permitindo uma escolha mais informada e alinhada às demandas específicas de cada aplicação.

## CONCLUSÃO

Monitorar ecossistemas como o Pantanal é essencial para compreender e conservar esse ambiente único. Nesse contexto, a combinação de técnicas de sensoriamento remoto com machine learning avançado surge como uma alternativa prática e eficaz para a aquisição de dados abrangentes e de alta qualidade em grande escala.

A comparação entre diferentes modelos de deep learning, como InterImage, DeepLabV3+ e SegFormer, revelou suas capacidades e limitações respectivas na classificação de corpos d'água. Embora todos os modelos tenham apresentado resultados satisfatórios, foram observadas algumas dificuldades ocasionais em delinear corpos d'água com bordas mais detalhadas e uma tendência à superestimativa da classificação de objetos como corpos d'água, mesmo quando esses não correspondiam à categoria específica.

O monitoramento de áreas alagadas por meio de imagens de VANT, devido à sua resolução espacial de 3 cm e ao alto nível de detalhamento, pode resolver dois desafios principais de forma eficaz: a variabilidade na resposta espectral e o tamanho reduzido de alguns corpos d'água, que não são identificáveis em imagens orbitais de resolução espacial mais baixa. O uso de VANT permite a captura de dados com precisão e detalhes que superam as limitações das imagens orbitais. Com isso, é possível identificar e analisar corpos d'água menores, bem como obter informações mais precisas sobre a cobertura vegetal e as características hidrológicas da região monitorada.

## REFERÊNCIAS

- AGISOFT LLC. **Metashape 2.1.3**. 2024. Disponível em: <https://www.agisoft.com/downloads/installer/>. Acesso em: 3 out. 2024.
- ALHO, C. J.; SILVA, J. S. Effects of severe floods and droughts on wildlife of the pantanal wetland (Brazil) - a review. **Animals**, Basel, v. 2, p. 591-610, 2012. DOI: <https://doi.org/10.3390/ani2040591>.
- CHEN, L. C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2018. **Proceedings of the European conference on computer**



**vision (ECCV)**. Cham: Springer, 2018. p. 801-818. DOI: [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49).

CONTRIBUTORS, M. **MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark**. 2020. Disponível em: <https://github.com/open-mmlab/mmdetection>. Acesso em: 3 out. 2024.

DU, S.; LIU, B.; ZHANG, X. Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high-resolution remote sensing images. **International Journal of Digital Earth**, Abingdon, v. 14, n. 3, p. 357-378, 2021. DOI: <https://doi.org/10.1080/17538947.2020.1831087>.

EVANS, T. L.; COSTA, M.; TELMER, K.; SILVA, T. S. Using ALOS/PALSAR and RADARSAT-2 to map land cover and seasonal inundation in the Brazilian Pantanal. **IEEE Journal of Selected Topics in Applied Earth Observations**, 2010. DOI: <https://doi.org/10.1109/JSTARS.2010.2089042>.

GALLANT, A. L. The challenges of remote monitoring of wetlands. **Remote Sensing**, Basel, v. 7, p. 10938-10950, 2015. DOI: <https://doi.org/10.3390/rs70810938>.

GARCIA-GARCIA, A.; ORTS-ESCOLANO, S.; OPREA, S.; VILLENA-MARTINEZ, V.; GARCIA-RODRIGUEZ, J. L. A review on deep learning techniques applied to semantic segmentation. **arXiv preprint arXiv:1704.06857**, 2017. Disponível em: <https://arxiv.org/abs/1704.06857>. Acesso em: 3 out. 2024.

HERLAWATI; HANDAYANTO, R. T. Land Cover Segmentation of Multispectral Images Using U-Net and DeeplabV3+ Architecture. **Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)**, v. 17, n. 1, 2024. DOI: <https://doi.org/10.21609/jiki.v17i1.1206>.

JEAN MILIEN, E.; NUNES, G. M.; PIERRE, G.; HAMILTON, S. K.; DA CUNHA, C. N. Hydrological dynamics of the Pantanal, a large tropical floodplain in Brazil, revealed by analysis of Sentinel-2 satellite imagery. **Water**, Basel, v. 15, n. 12, p. 2180, 2023. DOI: <https://doi.org/10.3390/w15122180>.

KHALIQ, A.; COMBA, L.; BIGLIA, A.; RICAUDA AIMONINO, D.; CHIABERGE, M.; GAY, P. Comparison of satellite and UAV-based multispectral imagery for vineyard variability assessment. **Remote Sensing**, Basel, v. 11, n. 4, p. 436, 2019. DOI: <https://doi.org/10.3390/rs11040436>.

LI, W.; LI, Y.; GONG, J.; FENG, Q.; ZHOU, J.; SUN, J.; HU, W. Urban water extraction with UAV high-resolution remote sensing data based on an improved U-Net model. **Remote Sensing**, Basel, v. 13, n. 16, p. 3165, 2021. DOI: <https://doi.org/10.3390/rs13163165>.

LI, X.; SUN, X.; MENG, Y.; LIANG, J.; WU, F.; LI, J. Dice loss for data-imbalanced NLP tasks. **arXiv preprint arXiv:1911.02855**, 2019. Disponível em: <https://arxiv.org/abs/1911.02855>. Acesso em: 3 out. 2024.

LI, Z.; WANG, R.; ZHANG, W.; HU, F.; MENG, L. Multiscale features supported DeepLabV3+ optimization scheme for accurate water semantic segmentation. **IEEE Access**, v. 7, p. 155787-155804, 2019. DOI: 10.1109/ACCESS.2019.2949635.

LIN, X.; CHENG, Y.; CHEN, G.; CHEN, W.; CHEN, R.; GAO, D.; WU, Y. Semantic segmentation of China's coastal wetlands based on Sentinel-2 and Segformer. **Remote Sensing**, Basel, v. 15, n. 15, p. 3714, 2023. DOI: <https://doi.org/10.3390/rs15153714>.

LOSCHIOLOV, I.; HUTTER, F. Decoupled weight decay regularization. **arXiv preprint arXiv:1711.05101**, 2017. Disponível em: <https://doi.org/10.48550/arXiv.1711.05101>. Acesso em: 3 out. 2024.

MESSINA, G.; PEÑA, J. M.; VIZZARI, M.; MODICA, G. A comparison of UAV and satellites multispectral imagery in monitoring onion crop. An application in the 'Cipolla Rossa di Tropea' (Italy). **Remote Sensing**, Basel, v. 12, n. 20, p. 3424, 2020. DOI: <https://doi.org/10.3390/rs12203424>.

MUNAWAR, H. S., ULLAH, F., QAYYUM, S., & HERAVI, A. Application of deep learning on uav-based aerial images for flood detection. **Smart Cities**, v. 4, n. 3, p. 1220-1242. 2021. DOI: <https://doi.org/10.3390/smartcities4030065>.

OSCO, L. P.; JUNIOR, J. M.; RAMOS, A. P. M.; DE CASTRO JORGE, L. A.; FATHOLAH, S. N.; DE ANDRADE SILVA, J.; LI, J. A review on deep learning in UAV remote sensing. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 102, p. 102456, 2021. DOI: <https://doi.org/10.1016/j.jag.2021.102456>.

QI, H.; KONG, X.; CHENG, L.; HU, J.; GU, J. Addressing fine-grained lake water body extraction: A hybrid approach combining vision transformer and geodesic active contour. **IEEE Transactions on Geoscience and Remote Sensing**, 2024. DOI: <https://doi.org/10.1109/TGRS.2024.3379506>.

SAMARANAYAKE, H., MUDANNAYAKE, O., PERERA, D., KUMARASINGHE, P., SUDUWELLA, C., DE ZOYSA, K., & WIMALARATNE, P. Detecting water in visual image streams from UAV with flight constraints. **Journal of Visual Communication and Image Representation**, v. 96, p. 103933. 2023. DOI: <https://doi.org/10.1016/j.jvcir.2023.103933>.

SHRIVASTAVA, A.; GUPTA, A.; GIRSHICK, R. Training region-based object detectors with online hard example mining. *In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION*, 2016. **Proceedings of the IEEE conference on computer vision and pattern recognition**. Piscataway: IEEE, 2016. p. 761-769. DOI: <https://doi.org/10.1109/CVPR.2016.89>.

SILVA, J. D. S. V.; MOURA ABDON, M. Delimitação do Pantanal brasileiro e suas sub-regiões. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 33, n. 13, p. 1703-1711, 1998.

SILVA, V. O.; DE MELLO, C. R.; CHOU, S. C.; GUO, L.; VIOLA, M. R. Characteristics of extreme meteorological droughts over the Brazilian Pantanal throughout the 21st century. **Frontiers in Water**, Lausanne, v. 6, p. 1385077, 2024. DOI: <https://doi.org/10.3389/frwa.2024.1385077>.

SUN, W.; CHEN, D.; LI, Z.; LI, S.; CHENG, S.; NIU, X.; YANG, X. Monitoring wetland plant diversity from space: progress and perspective. **International Journal of Applied Earth Observation and Geoinformation**, Amsterdam, v. 130, p. 103943, 2024. DOI: <https://doi.org/10.1016/j.jag.2024.103943>.

TAYE, M. M. Understanding of machine learning with deep learning: architectures, workflow, applications, and future directions. **Computers**, v. 12, n. 5, p. 91, 2023. DOI: <https://doi.org/10.3390/computers12050091>.

TRIPATHY, K. P.; MISHRA, A. K. Deep learning in hydrology and water resources disciplines: Concepts, methods, applications, and research directions. **Journal of Hydrology**, v. 625, p. 130458, 2023. DOI: <https://doi.org/10.1016/j.jhydrol.2023.130458>.

WANG, Q.; HU, C.; WANG, H.; WANG, R.; XIE, Y.; ZHAO, Y. Semantic segmentation of urban land classes using a multi-scale dataset. **International Journal of Remote Sensing**, Abingdon, v. 45, n. 2, p. 653-675, 2024. DOI: <https://doi.org/10.1080/01431161.2024.2302351>.

WANG, W.; DAI, J.; CHEN, Z.; HUANG, Z.; LI, Z.; ZHU, X.; QIAO, Y. Internimage: Exploring large-scale vision foundation models with deformable convolutions. *In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Piscataway: IEEE, 2023. p. 14408-14419. DOI: <https://doi.org/10.1109/CVPR52729.2023.01385>.

WANG, Y.; LI, S.; LIN, Y.; WANG, M. Lightweight deep neural network method for water body extraction from high-resolution remote sensing images with multisensors. **Sensors**, Basel, v. 21, n. 21, p. 7397, 2021. DOI: <https://doi.org/10.3390/s21217397>.

XIE, E.; WANG, W.; YU, Z.; ANANDKUMAR, A.; ALVAREZ, J. M.; LUO, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. **Advances in Neural Information Processing Systems**, v. 34, 2021. Disponível em:



Deep learning aplicado na detecção de corpos  
d'água em imagens de vant do Pantanal Brasileiro

Lucas Yuri Dutra de Oliveira  
João Lucas Aparecido Rocha Paes  
Maximilian Jaderson de Melo  
Maxwell da Rosa Oliveira  
Eveline Terra Bezerra  
Ana Paula Marques Ramos  
Jonathan Li  
Geraldo Alves Damasceno Júnior  
Wesley Nunes Gonçalves  
José Marcato Júnior

---

<https://proceedings.neurips.cc/paper/2021/hash/64f1f27bf1b4ec22924167ba9406af79-Abstract.html>.

Acesso em: 3 out. 2024.

YANG, F.; FENG, T.; XU, G.; CHEN, Y. Applied method for water-body segmentation based on mask R-CNN. **Journal of Applied Remote Sensing**, Bellingham, v. 14, n. 1, p. 014502, 2020. DOI: <https://doi.org/10.1117/1.JRS.14.014502>.

YU, F.; KOLTUN, V. Multi-scale context aggregation by dilated convolutions. **arXiv preprint arXiv:1511.07122**, 2015. DOI: <https://doi.org/10.48550/arXiv.1511.07122>.

---

Recebido em: 14/11/2024

Aceito para publicação em: 21/07/2025