

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
BACHARELADO EM CIÊNCIAS BIOLÓGICAS

Guilherme Rodrigues Machado

Trabalho de Conclusão de Curso

**Deteção de cervos-do-pantanal em imagens aéreas com um modelo de
aprendizado profundo de máquina**

Porto Alegre

2024

Guilherme Rodrigues Machado

**Detecção de cervos-do-pantanal em imagens aéreas com um modelo de
aprendizado profundo de máquina**

Trabalho de Conclusão do Curso de Bacharelado em Ciências Biológicas do Instituto de Biociências, da Universidade Federal do Rio Grande do Sul, como requisito para a obtenção do Título de Bacharelado(a) em Ciências Biológicas.

Orientador: Prof. Dr. Andreas Kindel

Coorientador: Dr. Ismael Verrastro Brack

Porto Alegre

2024

CIP - Catalogação na Publicação

Rodrigues Machado, Guilherme
Detecção de cervos-do-pantanal em imagens aéreas
com um modelo de aprendizado profundo de máquina /
Guilherme Rodrigues Machado. -- 2024.
37 f.
Orientador: Andreas Kindel.

Coorientador: Ismael Verrastro Brack.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Biociências, Bacharelado em Ciências Biológicas,
Porto Alegre, BR-RS, 2024.

1. Inteligência Artificial. 2. Aprendizado de
Máquina. 3. Detecção de Objeto. 4. Drones. 5.
Monitoramento Populacional. I. Kindel, Andreas,
orient. II. Verrastro Brack, Ismael, coorient. III.
Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os
dados fornecidos pelo(a) autor(a).

Guilherme Rodrigues Machado

**Detecção de cervos-do-pantanal em imagens aéreas com um modelo de
aprendizado profundo de máquina**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Biociências da
Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título
de Bacharel em Ciências Biológicas.

Aprovado em: 13 de agosto de 2024

BANCA EXAMINADORA

Cintia Fernanda da Costa - UFRGS

Claudio Rosito Jung - UFRGS

Andreas Kindel - UFRGS (orientador)

Ismael Verrastro Brack – University of Florida (coorientador)

RESUMO

Modelos de detecção de objeto têm evoluído muito e se mostrado uma ferramenta interessante para aumentar a eficiência na amostragem de animais silvestres. Neste trabalho, treinamos um modelo do tipo YOLOv8 usando um conjunto de imagens rotuladas de cervos-do-pantanal na região do Pantanal (MT), registradas em 2017 e 2018, e avaliamos sua performance em um novo conjunto de imagens coletadas em 2022. Embora o modelo demonstre uma alta capacidade de detecção de verdadeiros positivos, o elevado número de falsos positivos indica a necessidade de mais desenvolvimento para alcançar um desempenho ideal. Apesar disso, o algoritmo desenvolvido pode ser útil para protocolos de revisão semiautomática, potencialmente reduzindo o esforço necessário para o processamento das imagens. Direções futuras de pesquisa podem incluir a exploração de novas técnicas de aumento de dados, ajuste de hiperparâmetros ou a incorporação de arquiteturas de modelos alternativos. Essas abordagens podem ajudar a enfrentar melhor os desafios apresentados pelo plano de fundo complexo onde os indivíduos aparecem e a melhorar a capacidade do modelo em distinguir cervos de outros objetos de confusão, como vacas, cavalos, troncos de árvore e vegetação seca.

Palavras-chave: Inteligência Artificial, Aprendizado de Máquina, Detecção de Objeto, Drones, Monitoramento Populacional.

Abstract

Object detection models have evolved a lot and have proven to be an interesting tool for increasing efficiency in wildlife sampling. In this work, we trained a YOLOv8-type model using a set of labeled images of marsh deer in the Pantanal region (MT), recorded in 2017 and 2018, and evaluated its performance on a new set of images collected in 2022. Although the model shows a high capacity for detecting true positives, the high number of false positives indicates the need for further development to achieve optimal performance. Despite this, the algorithm developed could be useful for semi-automated review protocols, potentially reducing the effort required to process the images. Future research directions may include exploring new data augmentation techniques, adjusting hyperparameters or incorporating alternative model architectures. These approaches may help to better address the challenges presented by the complex background in which individuals appear and improve the model's ability to distinguish deer from other confounding objects such as cows, horses, tree trunks and dry vegetation.

Keywords: Artificial Intelligence, Machine Learning, Drones, Object Detection, Population Monitoring.

SUMÁRIO

1	INTRODUÇÃO	8
2	MATERIAIS E MÉTODOS	11
2.1	ÁREA DE ESTUDO E ESPÉCIE ESTUDADA	11
2.2	BASES DE DADOS E TREINAMENTO DO MODELO	11
2.3	RECURSOS COMPUTACIONAIS	14
2.4	PRÉ-PROCESSAMENTO	14
2.5	TREINAMENTO E VALIDAÇÃO DO ALGORITMO	15
2.6	AVALIAÇÃO DA PRECISÃO E REVOCAÇÃO	17
3	RESULTADOS	18
4	DISCUSSÃO	27
5	CONCLUSÃO	30
	REFERÊNCIAS	31

Introdução

O monitoramento da biodiversidade é essencial para avaliar o estado de conservação de espécies e ecossistemas, entender as respostas desses organismos a distúrbios antrópicos e avaliar a efetividade dos esforços de conservação e manejo (Navarro *et al.*, 2017; Ceballos *et al.*, 2017; Mawdsley *et al.*, 2009). O monitoramento a longo-prazo da distribuição ou abundância de populações ameaçadas, e em múltiplas escalas (extensões e resoluções), é essencial para prover informações de alta qualidade em tempo hábil para projetos de conservação (Leung *et al.*, 2024).

Amostragens aéreas são frequentemente utilizadas para censos de biodiversidade, permitindo a coleta de dados sobre a distribuição e abundância de espécies em diversos habitats, o que facilita o monitoramento de ecossistemas e iniciativas de conservação (Oleksyn *et al.*, 2021; Ivanova *et al.*, 2022; Povlsen *et al.*, 2023). Tradicionalmente, os métodos de amostragem aérea eram realizados com o uso de aviões ou helicópteros (Mourão, G. *et al.*, 2000). Para isso, é necessário contar com pilotos altamente capacitados para a realização das pesquisas, o que implica em custos elevados tanto para o treinamento especializado dos profissionais quanto para a aquisição e manutenção dos equipamentos, que são de difícil acesso (Milligan *et al.*, 2021; Gentle *et al.*, 2018). Também, as altas altitudes necessárias para esse tipo de voo podem prejudicar a percepção do observador (Franke *et al.*, 2012; Habib *et al.*, 2012). Por isso, há uma necessidade de desenvolver e testar novas técnicas de amostragem aéreas que sejam mais econômicas, seguras e eficazes para estimar e monitorar a biodiversidade.

Uma alternativa promissora para amostragens populacionais é o uso de drones. Esses dispositivos, que são controlados remotamente, funcionam como plataformas aéreas que podem ser equipadas com uma ampla variedade de sensores, como câmeras de alta resolução e câmeras termais (Povlsen *et al.*, 2023; Schad *et al.*, 2022). Eles podem operar em vários tipos de ambientes abertos, permitindo que cientistas conduzam pesquisas e obtenham imagens de alta resolução e cobertura espacial em áreas remotas e de difícil acesso (Hodgson *et al.*, 2018). Os drones têm sido uma ferramenta útil na coleta de dados populacionais, permitindo uma avaliação mais precisa (Koger *et al.* 2023, Koh *et al.*, 2012; Mulero-Pázmány *et al.*, 2014; Barasona, *et al.*, 2014; Chrétien *et al.*, 2016; Gentle, *et al.*, 2018; Hodgson, *et al.* 2018), e oferecem uma solução para o censo de biodiversidade mais acessível que os métodos tradicionais. As imagens capturadas com drones podem fornecer informações valiosas sobre a distribuição dos animais, bem como sobre as características do habitat (Hodgson, *et al.* 2018). Drones também podem voar em menor altitude que aviões tripulados e causam menos distúrbio aos animais (Schad *et al.*, 2023). Porém, o uso de drones também apresenta desafios. As amostragens com drones geram grandes conjuntos de imagens e a revisão manual é muito trabalhosa, podendo até inviabilizar a aplicação da metodologia (Witmer, G., 2005). Além disso, no processo de revisão de imagens, erros de amostragem, como não

detectar um indivíduo presente na área ou contar o mesmo indivíduo duas vezes, podem produzir vieses nas estimativas de abundância, se não forem devidamente tratados (Brack *et al*, 2018). As limitações físicas dos seres humanos, o excesso de atributos observáveis que podem estar presentes nas imagens, e a dimensão da área que pode ser efetivamente monitorada inevitavelmente restringem o número de indivíduos que podem ser observados (Tuia *et al.*, 2022; Brack *et al*, 2018). Consequentemente, os métodos de monitoramento necessitam prover avaliações o mais rápidas e acuradas possíveis, fornecendo informações populacionais confiáveis para amplas áreas. Entretanto, a falta de acurácia nos dados de monitoramento, ou a demora na sua obtenção, pode acarretar conclusões equivocadas sobre as dinâmicas populacionais ou não perceber declínios em tempo de evitar a extinção da espécie-alvo. (Ferreira & Aarde, 2009; Fritsch & Downs, 2020).

Os métodos de visão computacional podem auxiliar na detecção dos animais em imagens aéreas, reduzindo significativamente o esforço humano necessário. Tentativas de automatização do processo de revisão de imagens têm sido feitas usando algoritmos de classificação com aprendizado de máquina, como janelas deslizantes e *random forest*, e já se demonstram aplicações viáveis para detectar, contar ou identificar espécies silvestres em imagens (Brodrick *et al.* 2019, ;Chabot *et al*, 2016; Corcoran *et al.*, 2021; Brack, 2022; Hollings *et al*, 2018; Patterson *et al*, 2016; Lenzi *et al*, 2023; Groom *et al*, 2001). Métodos frequentemente usados para detectar animais em imagens digitais incluem correspondência de modelos (Ollier *et al*, 2023), OBIA (*object-based image analysis*) (Chabot *et al* 2018), redes neurais convolucionais (Brack, I.V., 2022), e são normalmente desenhados para diferenciar entre classes de interesse (por exemplo, espécies diferentes ou fundo versus animal) (Brack, I.V., 2022; Lenzi *et al*, 2023). Entretanto, esses métodos podem encontrar dificuldade em atingir alta precisão e podem falhar ou gerar classificações imprecisas se as condições necessárias (eg. oferta insuficiente de dados de treinamento) para o treinamento não forem atingidas (Eikelboom *et al*, 2029). Além disso, demandam recursos computacionais de alto desempenho (Chabot *et al* 2016; Corcoran *et al*, 2021; Hollings *et al*, 2018).

Para superar esses desafios, novos algoritmos para detecção de objeto vêm sendo desenvolvidos e atualizados (Patterson *et al*, 2016; Lenzi *et al*, 2023). Esses algoritmos, em geral, funcionam processando grandes quantidades de dados usando técnicas de aprendizado profundo para fazer previsões ou decisões (Liu *et al*, 2021). Esse processamento, chamado de treinamento, envolve alimentar o algoritmo com dados rotulados, onde o algoritmo aprende a identificar padrões nos dados que se correlacionam com resultados específicos (Xu *et al*, 2024). Depois de treinado, o algoritmo pode fazer previsões ou decisões sobre imagens novas e não rotuladas. Os novos métodos de detecção de objeto com aprendizado profundo se destacam por sua capacidade de aprender características relevantes das imagens sem a necessidade de extração manual de atributos (Chabot *et al*, 2016; Corcoran *et al*, 2021; Hollings *et al*, 2018). Todavia, como esses modelos exigem grandes quantidades de dados rotulados para seu treinamento, sua performance pode ser

comprometida. Em cenários reais, nem sempre é possível adquirir a quantidade necessária estipulada para esses diferentes modelos, especialmente quando falamos em espécies ameaçadas e com populações reduzidas. Além disso, eles podem apresentar problemas de generalização para novas condições de imagem, como variações de iluminação, escala, ângulo e fundo (Schad *et al.*, 2023).

Neste trabalho, apresentamos um modelo de aprendizado profundo para a detecção do cervo-do-pantanal, um cervídeo de grande porte ameaçado de extinção (IUCN; Duarte *et al.*, 2016), em imagens aéreas obtidas por drones. O modelo utiliza uma rede neural convolucional pré-treinada em um grande conjunto de dados, adaptada ao problema específico por técnicas de transferência de aprendizado e aumento de dados. Treinada com imagens aéreas capturadas por drones em 2017 e 2018 em uma reserva no Pantanal Norte (MT), a rede visa localizar e classificar os cervos-do-pantanal com alta precisão e acurácia. Nossa abordagem busca reduzir o tempo e o esforço necessários para revisar as imagens em campanhas de avaliação populacional da espécie. Dada a ameaça de extinção, desenvolver métodos eficientes, confiáveis e pouco invasivos para amostragem e monitoramento é crucial para planejar a proteção e avaliar o sucesso das ações de conservação, considerando as características comportamentais e morfológicas do cervo-do-pantanal e seu habitat.

Materiais e Métodos

Espécie estudada e área de estudo

O Pantanal está localizado na região oeste do Brasil, nos estados do Mato Grosso e Mato Grosso do Sul. Este ecossistema de savana sazonalmente alagável possui uma grande diversidade de flora e fauna. Uma das espécies icônicas desse ecossistema é o cervo-do-pantanal (*Blastocerus dichotomus*), um cervídeo de grande porte, com até 150kg, associado a áreas úmidas e considerado globalmente ameaçado de extinção na categoria vulnerável (IUCN; Duarte *et al.*, 2016). O cervo-do-pantanal é o maior cervídeo da América do Sul e enfrenta sérios desafios devido à perda e fragmentação de seu habitat, alterações causadas por atividades antrópicas e a caça ilegal. As maiores populações da espécie são encontradas no Pantanal, enquanto a maioria das outras populações são mais isoladas e menos numerosas (Costa *et al.*, 2006; Marin *et al.*, 2020).

A área de estudo é a Reserva Particular do Patrimônio Natural Sesc Pantanal (16° 45' S e 56° 15' W; 108.000ha). A área é composta por uma vegetação heterogênea (campos, cerrado, arbustivos e florestas ciliares, inundadas e sazonalmente secas) estruturada principalmente por diferenças de microrrelevo que recebem diferentes regimes de inundação (Oliveira *et al.*, 2013). A população do cervo-do-pantanal vem sendo monitorada na reserva desde 2017 usando drones (Brack *et al.* 2022).

Bases de dados e treinamento do modelo

Os voos foram realizados em sete rotas de voo pré-planejadas usando um drone de asa fixa (Echar 20B, XMobots) equipado com uma câmera Sony A7R (7392x4920 px)

ou Sony Alpha 500 (6000x4000 px) orientadas para baixo (perpendicular ao solo), resultando em um *Ground Sample Distance* (GSD), de 2cm/pixel (para maiores detalhes ver Brack *et al.*, 2023). Dividimos o conjunto de imagens disponíveis em três subconjuntos distintos: treinamento, validação e teste.

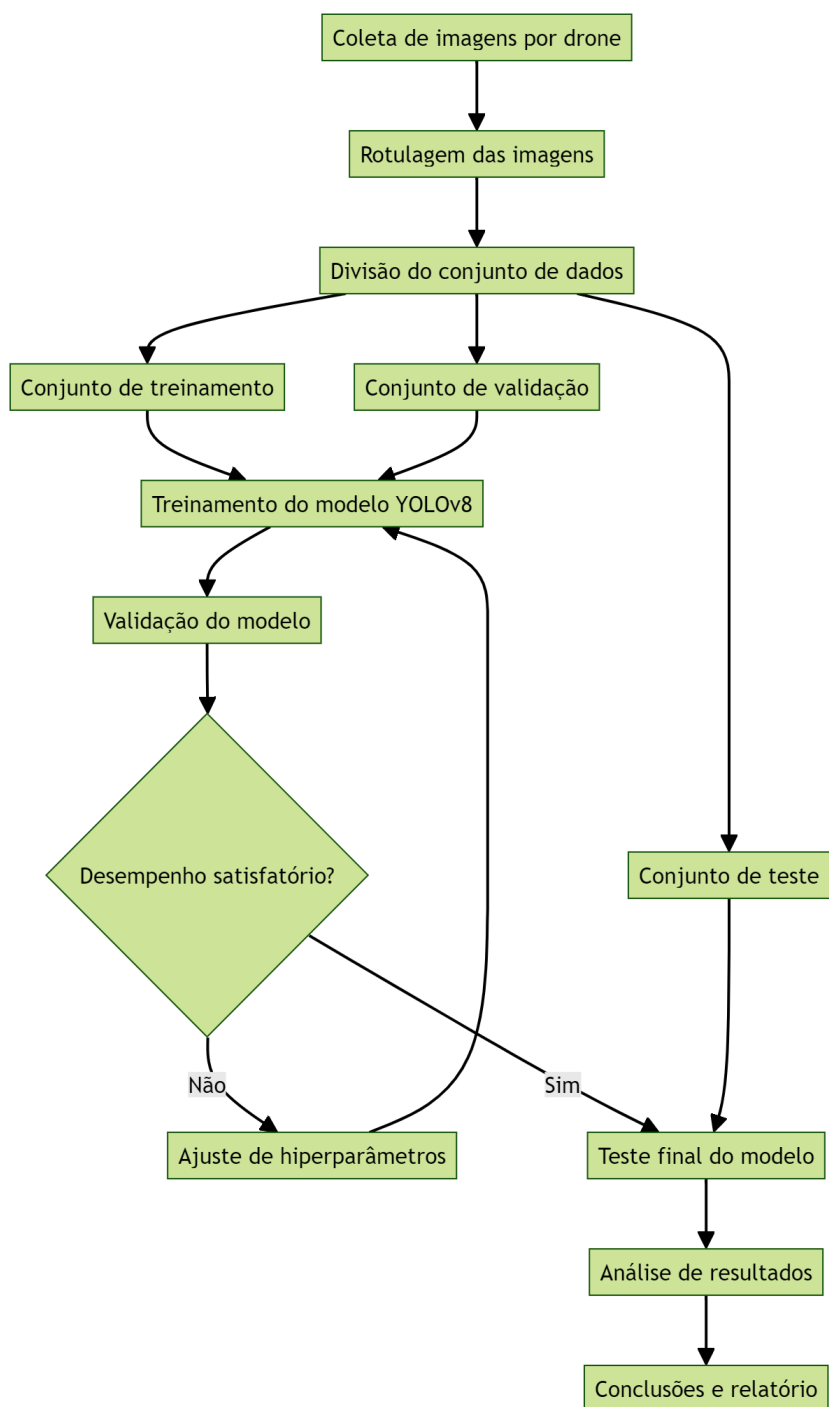
Para o treinamento e validação, utilizamos imagens de um total de 67 voos (25 em 2017 e 42 em 2018) Observadores humanos independentes revisaram manualmente as 75.613 imagens coletadas nos 67 voos, resultando em 573 amostras de cervídeos (encontradas em 457 imagens) de 183 registros independentes (excluindo imagens sobrepostas). Para cada cervo-do-pantanal detectado, os observadores marcaram o indivíduo com um ponto central usando o software *ClickPoints* (Gerum *et al.*, 2017). Esses registros foram rotulados manualmente com caixas delimitadoras (*bounding boxes*) para se adequar aos requerimentos de treinamento do modelo, que não se guia por um ponto central no objeto, e sim por uma caixa que delimita o objeto de interesse.

O conjunto de treinamento foi composto por 90% dessas imagens, totalizando 458 imagens de cervos-do-pantanal e uma parcela de amostras de fundo correspondentes a um pouco mais de 15% do volume do conjunto original (100 imagens). Essas imagens foram utilizadas para ajustar os parâmetros do modelo YOLOv8, que é pré-treinado em um grande conjunto de dados intitulado COCO (*Common Objects in Context*) por meio do processo de aprendizado supervisionado (Jocher *et al.*, 2023; Lin *et al.*, 2015). Durante o treinamento, as imagens passam por processos de ampliação dos dados (*Data Augmentation*), com a aplicação de transformações nas imagens como rotação e espelhamento, para aumentar a oferta de dados disponíveis para o modelo. O conjunto de validação, por sua vez, consistiu nos outros 10% dessas imagens dos voos de 2017 e 2018, totalizando 57 amostras de cervos e uma quantidade proporcional de amostras de fundo (n=14). Esse conjunto de validação foi utilizado durante o processo de treinamento para monitorar o desempenho do modelo em dados não vistos previamente e ajustar os hiperparâmetros, evitando o sobreajuste (*overfitting*). Em aprendizado de máquina, o sobreajuste é um problema que ocorre quando um modelo é treinado de forma tão específica para os dados de treinamento, que acaba capturando até mesmo ruídos ou variações aleatórias nos dados, em vez de aprender apenas os padrões subjacentes. Como resultado, o modelo tem um bom desempenho nos dados de treinamento, mas um desempenho ruim em novos dados ou dados de teste, pois não generaliza bem (Ying *et al.*, 2019).

No treinamento final do modelo, após os ajustes finais de hiperparâmetros, o conjunto de treinamento e validação foi unido em um único conjunto de treinamento. A divisão dos conjuntos de dados foi realizada de forma aleatória estratificada, garantindo que cada subconjunto mantivesse uma distribuição proporcional de amostras positivas (cervos) e negativas (fundo), e representasse todas as missões com a mesma proporção entre si. Essa abordagem visa minimizar possíveis vieses e fornecer uma avaliação mais confiável e representativa do desempenho do modelo.

Por fim, o conjunto de teste foi formado por 13.963 imagens, correspondendo a 10 voos realizados em 2022 nas mesmas missões dos anos anteriores, representando condições ambientais e temporais distintas das utilizadas para treinamento e validação (4-5 anos depois). Esse conjunto de dados independente foi utilizado para avaliar o desempenho final do modelo treinado em um cenário realista e não visto anteriormente. Esse processo foi simplificado em um fluxograma, conforme mostrado na Figura 1.

Figura 1 - Fluxo desde a coleta de imagens por drones na RPPN SESC-Pantanal, passando por divisões dos conjuntos de dados, até treinamento e teste do modelo YOLOv8.



Fonte: Autor

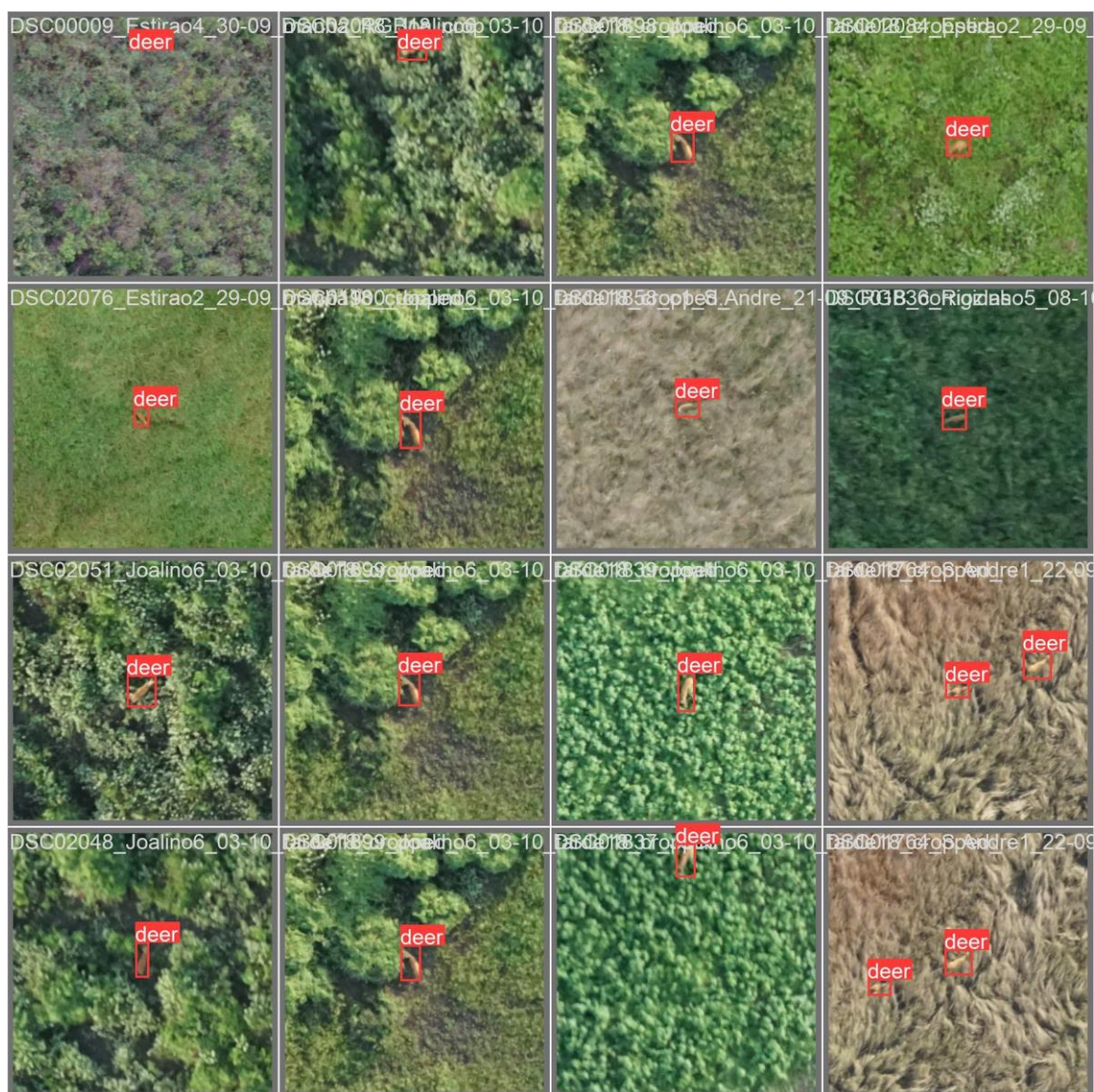
Recursos Computacionais

Para armazenamento dos dados e treinamento do algoritmo, nós utilizamos um computador com as seguintes especificações: NVIDIA RTX 2080 com 8192MiB de VRAM, Processador Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz, 16GB de Memória RAM e 9TB de memória para armazenamento HDD, com sistema operacional Microsoft Windows 10 Pro. Todo o código necessário para o desenvolvimento do projeto foi escrito em Python versão 3.11.6 utilizando as bibliotecas PyTorch 2.1.0, CUDA 12.1, Ultralytics v8.1.0, Sahi v0.11.15, OpenCV 4.9.0, Pandas 2.1.4, Jupyter Notebook 6.4.12 e Albumentations 1.0.3. Para garantir a replicabilidade da aplicação e isolar o desenvolvimento do resto do sistema, utilizamos a plataforma Docker versão 24.0.7 para criar um contêiner (container) onde a aplicação seria executada em um ambiente virtual Linux.

Pré-processamento

Para cada uma das imagens de cervos no conjunto de treinamento, recortamos blocos de 640 x 640 pixels em torno das coordenadas centrais anotadas durante a revisão manual, conforme pode ser observado na Figura 2. Esse tamanho permite plena visibilidade do cervo, e esta resolução nos permite treinar o modelo com recursos computacionais limitados. Para as amostras de fundo (“não-cervos”), geramos pontos aleatórios nas imagens do drone (exceto nas imagens dos voos separados para teste) para recortar janelas de 640 x 640 pixels, para auxiliar o algoritmo a distinguir características complexas do fundo do Pantanal e não as confundir com cervos. Esses recortes representavam cerca de 15% do conjunto total.

Figura 2 – Exemplos de imagens de cervos-do-pantanal obtidas por drone na RPPN SESC-Pantanal, recortadas (640 x 640 pixels) e rotuladas e representando múltiplas condições de fundo e obstrução da espécie alvo.



Fonte: Autor

Treinamento e validação do algoritmo

A escolha do modelo depende dos requisitos específicos da aplicação, como velocidade, precisão, otimização de recursos computacionais, e conformação aos dados. Exploramos diferentes modelos de aprendizado de máquina para detecção de objetos em imagens, conforme indicado na literatura, e optamos pela escolha do modelo YOLOv8 (Jocher *et al*, 2023), devido ao seu desempenho otimizado em recursos computacionais limitados e alta velocidade de treinamento e inferência, além da sua comprovada eficácia em outras aplicações de classificação de fauna em imagens capturadas com drones (Rančić *et al.*, 2023).

O modelo YOLOv8 utilizado neste trabalho é uma rede neural convolucional de detecção de objetos pré-treinada no conjunto de dados COCO (Common Objects in Context). O COCO é um grande conjunto de dados de imagens anotadas amplamente utilizado em tarefas de visão computacional, contendo mais de 300.000 imagens com mais de 2 milhões de instâncias rotuladas pertencentes a 80 categorias de objetos (Lin *et al.*, 2015). A arquitetura do YOLOv8 baseia-se nas versões anteriores dos algoritmos YOLO, e utiliza uma rede neural convolucional que pode ser dividida em duas partes principais: o *backbone* e a cabeça. No *backbone*, há 53 camadas convolucionais, e ele emprega conexões parciais entre estágios para melhorar o fluxo de informações entre as diferentes camadas. A cabeça do YOLOv8 consiste em várias camadas convolucionais seguidas por uma série de camadas totalmente conectadas. Essas camadas são responsáveis pela previsão de caixas delimitadoras, *score* de confiança, e probabilidades de classe para os objetos detectados em uma imagem.

O processo de treinamento de uma rede neural consiste em passar as amostras de treinamento com seus rótulos correspondentes pela rede para ajustar os pesos do modelo com o objetivo de reduzir os valores da função de perda. As amostras de treinamento são passadas em mini-lotes (por exemplo, 32 ou 64 amostras) até completar uma época (uma iteração de processamento do conjunto de treinamento). Durante o treinamento no COCO, o modelo YOLOv8 recebe como entrada imagens RGB e aprende a detectar e classificar objetos presentes nessas imagens por meio da otimização dos pesos de sua rede neural (Jocher *et al.*, 2023). A função de perda utilizada combina componentes de classificação (para determinar a classe do objeto) e regressão (para estimar as coordenadas da caixa delimitadora em torno do objeto). O treinamento é realizado utilizando aprendizado supervisionado profundo, com o algoritmo otimizando iterativamente os pesos da rede com base nas previsões e nos rótulos reais fornecidos pelo conjunto de dados. Técnicas como aumento de dados e regularização são aplicadas para melhorar a generalização e evitar sobreajuste. Após cada época, as métricas de desempenho são calculadas após a avaliação das previsões obtidas no conjunto de validação anotado. Esse processo é iterado por várias épocas até que as métricas calculadas para o conjunto de validação comecem a diminuir o desempenho (ou seja, apresentando sobreajuste no conjunto de treinamento). O objetivo desta etapa é minimizar as métricas da função de perda.

Após este treinamento inicial, o modelo adquire a capacidade de detectar e reconhecer uma ampla gama de objetos comuns em imagens genéricas. No entanto, para a tarefa específica de detecção de cervos-do-pantanal, é necessário um processo adicional de ajuste fino (*fine-tuning*) utilizando dados específicos deste domínio. Seu ajuste foi feito localmente com nosso conjunto de dados de imagens do cervo-do-pantanal. Durante o treinamento, as 10 primeiras camadas convolucionais no backbone do modelo tiveram seus pesos “congelados” (fixados) para tomar proveito das capacidades adquiridas pelo modelo em seu treinamento original e aplicar a técnica de transferência de aprendizado. A função para treinamento foi uma

classificação binária (cervo vs fundo), designando o único objeto de interesse como o cervo-do-pantanal.

Foi nesta etapa em que as 573 imagens dos cervos-do-pantanal foram apresentadas para o algoritmo, ao longo de 500 épocas, com o otimizador AdamW, taxa de aprendizado (*learning rate*) = 0.001 e tamanho de *batch* = 20. As demais configurações mantiveram-se como as pré-definidas pelos desenvolvedores. Durante esta etapa, os dados foram aumentados com aplicações de transformações aleatórias (rotação, espelhamento, mudança de alcance, clarear/escurecer e aumentar/diminuir o zoom) a cada época com a biblioteca *albumentations*, como pode ser observado na Figura 3. Essas transformações foram aplicadas em múltiplas combinações durante as épocas do treinamento, de forma que ao longo do treinamento, o modelo recebia diversas imagens “novas” construídas com elas, na tentativa de suprir a demanda de dados que esse treinamento exige. Para as amostras de fundo (“não-cervos”), geramos pontos aleatórios nas imagens dos voos (exceto nas imagens dos voos separados para teste) para recortar janelas de 640 x 640 pixels, para auxiliar o algoritmo no reconhecimento de recortes sem cervos na tentativa de diminuir o número de falsos-positivos encontrados.

Figura 3 – Exemplos de quatro imagens de treinamento contendo cervos rotulados com caixas delimitadoras (retângulos vermelhos) ou recortes de fundos com transformações aplicadas.



Legenda: a- Imagem de cervo rotulado com transformações de rotação e clareamento aplicadas b- Imagem de cervo rotulado com transformações de rotação, clareamento e aumento de *zoom* aplicadas. c- Imagem de recorte de fundo com transformações de rotação, mudança de alcance, escurecimento e diminuição de *zoom* aplicadas d- Imagem de cervo rotulado com transformações de rotação, espelhamento e escurecimento aplicadas. Fonte: Autor

Avaliação da precisão e revocação

Para garantir a consistência na resolução das imagens entre os estágios de treinamento e inferência, implementamos o *framework* Sahi (Akyon *et al.*, 2022) durante o estágio de previsão. Esse framework permitiu que as imagens de alta resolução capturadas pelas câmeras dos drones fossem “fatiadas” em janelas de 640 x 640 pixels, assegurando que a previsão fosse realizada no mesmo tamanho de imagem em que o modelo foi treinado. Cada imagem gerava 150 janelas, e a configuração de sobreposição entre janelas foi definida como 0.2 (20%).

Para avaliar o desempenho do modelo treinado na tarefa de detecção de cervos-do-pantanal, utilizamos as métricas de precisão e revocação (*recall*). Essas métricas são amplamente empregadas na avaliação de modelos de aprendizado de máquina para problemas de classificação e detecção de objetos. A precisão é quantificada pela proporção de detecções consideradas verdadeiros positivos em relação ao total de detecções positivas, sendo calculada pela seguinte equação:

$$\text{Precisão} = \text{Verdadeiros Positivos} / (\text{Verdadeiros Positivos} + \text{Falsos Positivos})$$

Já a revocação mede a proporção de exemplos positivos corretamente identificados pelo modelo em relação ao total de exemplos positivos reais. É calculada pela equação:

$$\text{Revocação} = \text{Verdadeiros Positivos} / (\text{Verdadeiros Positivos} + \text{Falsos Negativos})$$

Neste estudo um verdadeiro positivo representa uma detecção correta (anotado durante a revisão manual) de um cervo-do-pantanal pelo modelo. Um falso positivo ocorre quando o modelo identifica um objeto como sendo um cervo, quando na verdade não é. Por outro lado, um falso negativo refere-se a uma falha do modelo em detectar um cervo-do-pantanal presente nas imagens.

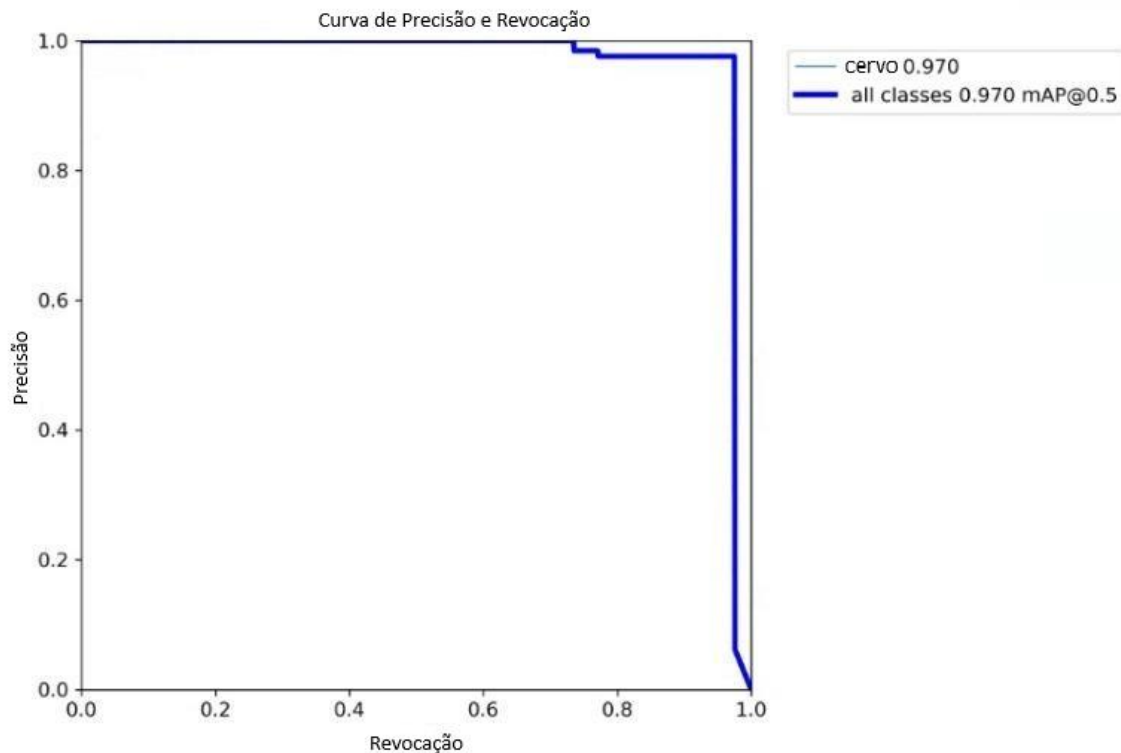
Nosso objetivo foi primeiramente buscar a maximização da revocação, pois se tratando de um estudo de conservação de um animal com uma população já limitada, entendemos que seria mais importante garantir a presença de todos os cervos disponíveis nas detecções. A precisão seria aprimorada em outro momento, por meio de iterações de testes e filtros na revisão destas detecções. Para compreender melhor os padrões de detecção do teste final, realizamos análises exploratórias dos tamanhos de caixas delimitadoras e *scores* de confiança das detecções, comparando falsos positivos e verdadeiros positivos. Como pré-definição, apenas as detecções com *score* de confiança igual ou maior que 0,60 foram consideradas válidas.

Resultados

Durante a fase de treinamento, observamos um desempenho promissor na validação do modelo YOLOv8 treinado em imagens de cervos-do-pantanal, como pode ser visto nas Figuras 4 e 5. A área da curva de precisão-revocação do conjunto de validação mostrou um valor de aproximadamente 0,970. Altas pontuações para ambos os indicadores mostram que o classificador está retornando resultados precisos e identificando a maioria dos resultados positivos.

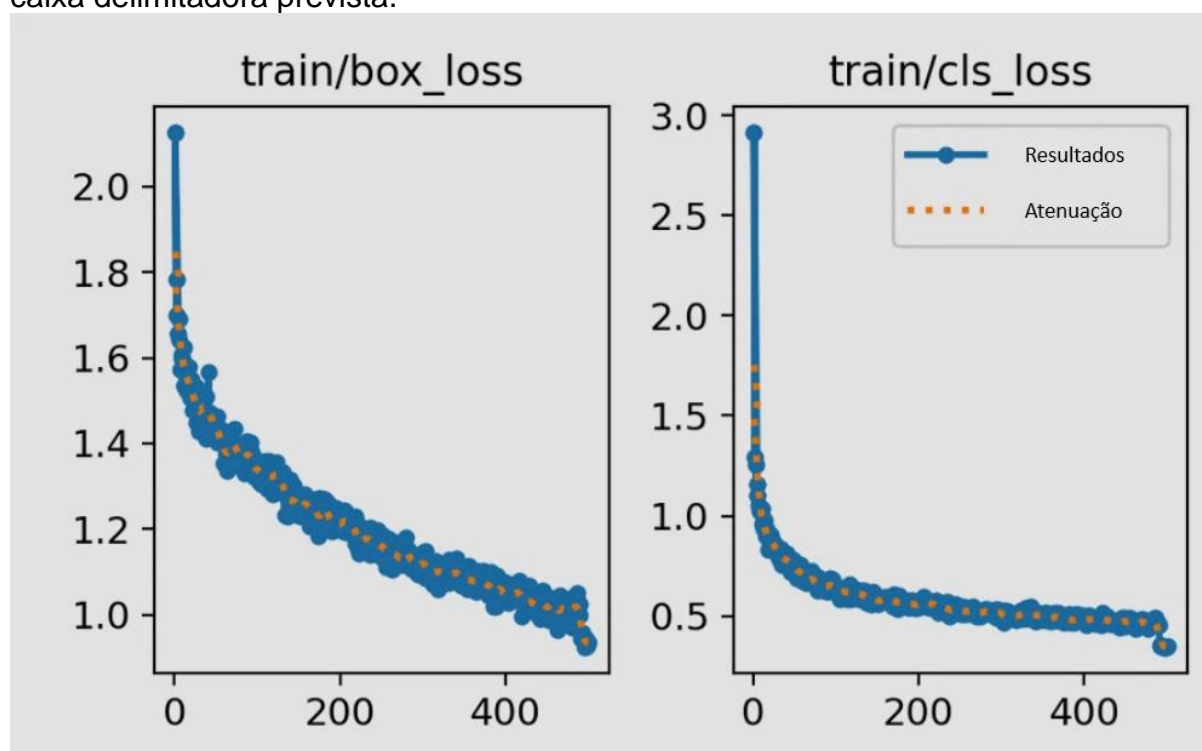
Esse resultado indicaria que o modelo conseguiria generalizar bem os dados de treinamento para dados não vistos no conjunto de validação. No entanto, é importante observar que o desempenho no conjunto de validação não garante necessariamente um desempenho semelhante em outros conjuntos de dados não vistos. A Figura 5 apresenta, separadamente, os gráficos de função de perda para caixas delimitadoras e para a classe ao longo das épocas de treinamento. O gráfico de função de perda para caixas delimitadoras reflete a precisão da localização das detecções, enquanto o gráfico de função de perda da classe mede a precisão na classificação dos objetos detectados. Esses gráficos de desempenho indicam uma convergência estável na função de perda para caixas delimitadoras após aproximadamente 400 épocas, enquanto o gráfico de função de perda da classe atinge estabilidade perto da época 200.

Figura 4 – Curva de precisão-revocação de detecção de cervos-do-pantanal em imagens aéreas para o conjunto de imagens de validação usando um modelo de detecção de objetos (YOLOv8).



Fonte: Autor

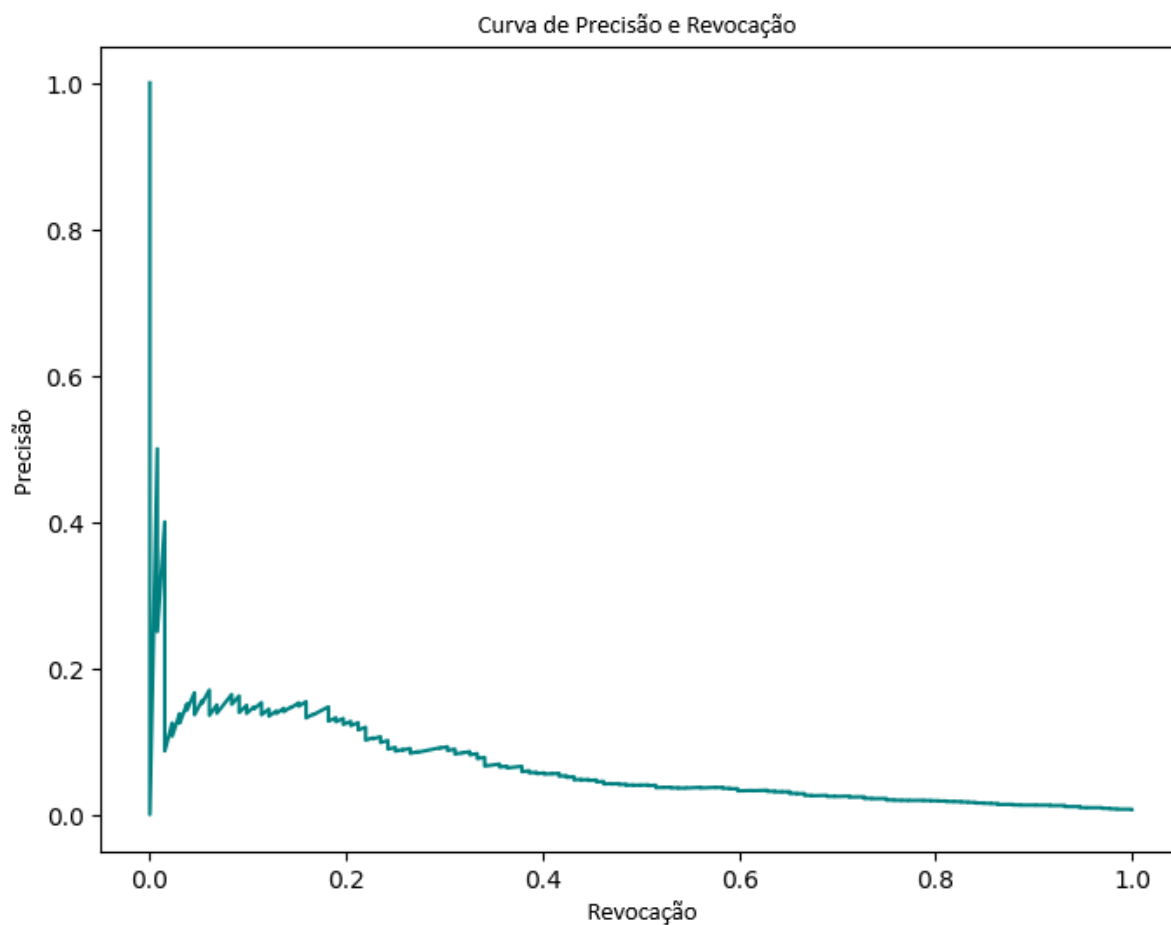
Figura 5 – Gráficos das métricas de desempenho da função de perda da validação do conjunto de treinamento do modelo. A primeira métrica de perda (*box_loss*) mede a "proximidade" das caixas delimitadoras previstas em relação ao objeto real anotado, a segunda métrica (*cls_loss*) mede a exatidão da classificação de cada caixa delimitadora prevista.



Fonte: Autor

No conjunto de teste final, nosso modelo conseguiu detectar 132 de 155 cervos-do-pantanal no conjunto de dados de teste (ver figura 7), alcançando uma taxa de verdadeiros positivos (TP) e de revocação de 0,8516. No entanto, observamos um número considerável de falsos positivos (FP), totalizando 18.652 (ver figura 8). Isso indica que objetos do complexo cenário do Pantanal foram frequentemente classificados erroneamente como cervos-do-pantanal pelo modelo, resultando em uma baixa precisão de 0,007. A figura 6 evidencia a diminuição substancial da precisão com o aumento da revocação.

Figura 6 – Curva de precisão-revocação de detecção de cervos-do-pantanal em imagens aéreas para o conjunto teste de imagens usando um modelo de detecção de objetos (YOLOv8).



Fonte: Autor

Figura 7 – Exemplo com recorte aproximado de detecção correta do cervo-do-pantanal em imagem aérea obtida através de algoritmo de detecção de objeto (YOLOv8) treinado.



Fonte: Autor

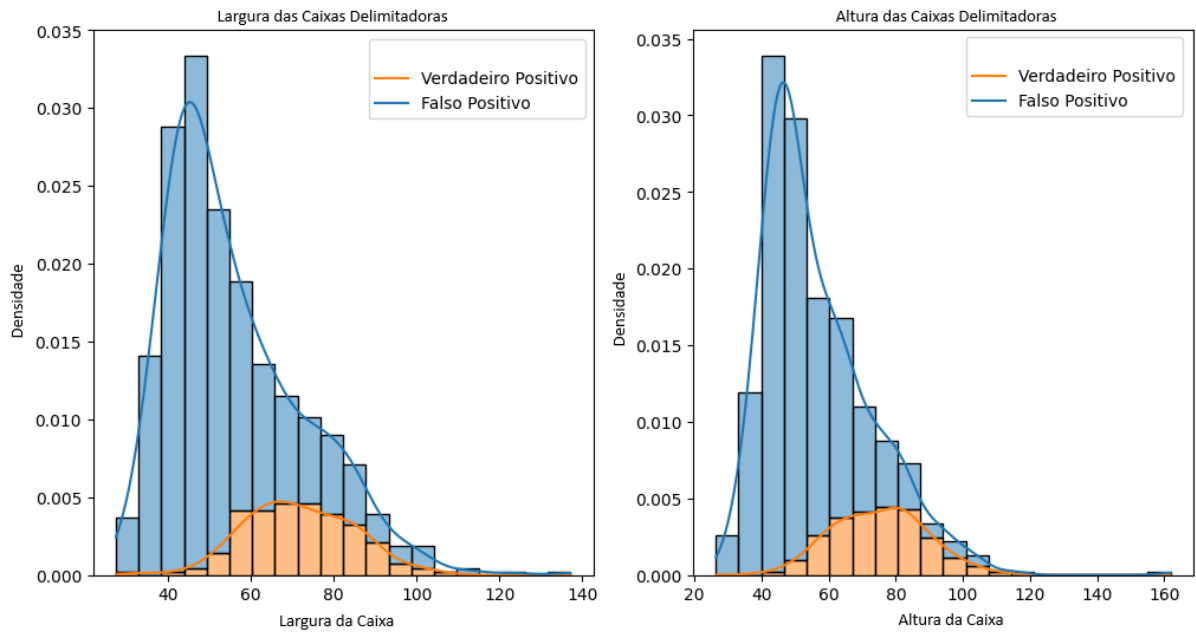
Figura 8 – Exemplo de detecções incorretas em imagem aérea de fundo complexo com atributos vegetais confundidos por cervos-do-pantanal obtidas através de algoritmo de detecção de objeto (YOLOv8) treinado



Fonte: Autor

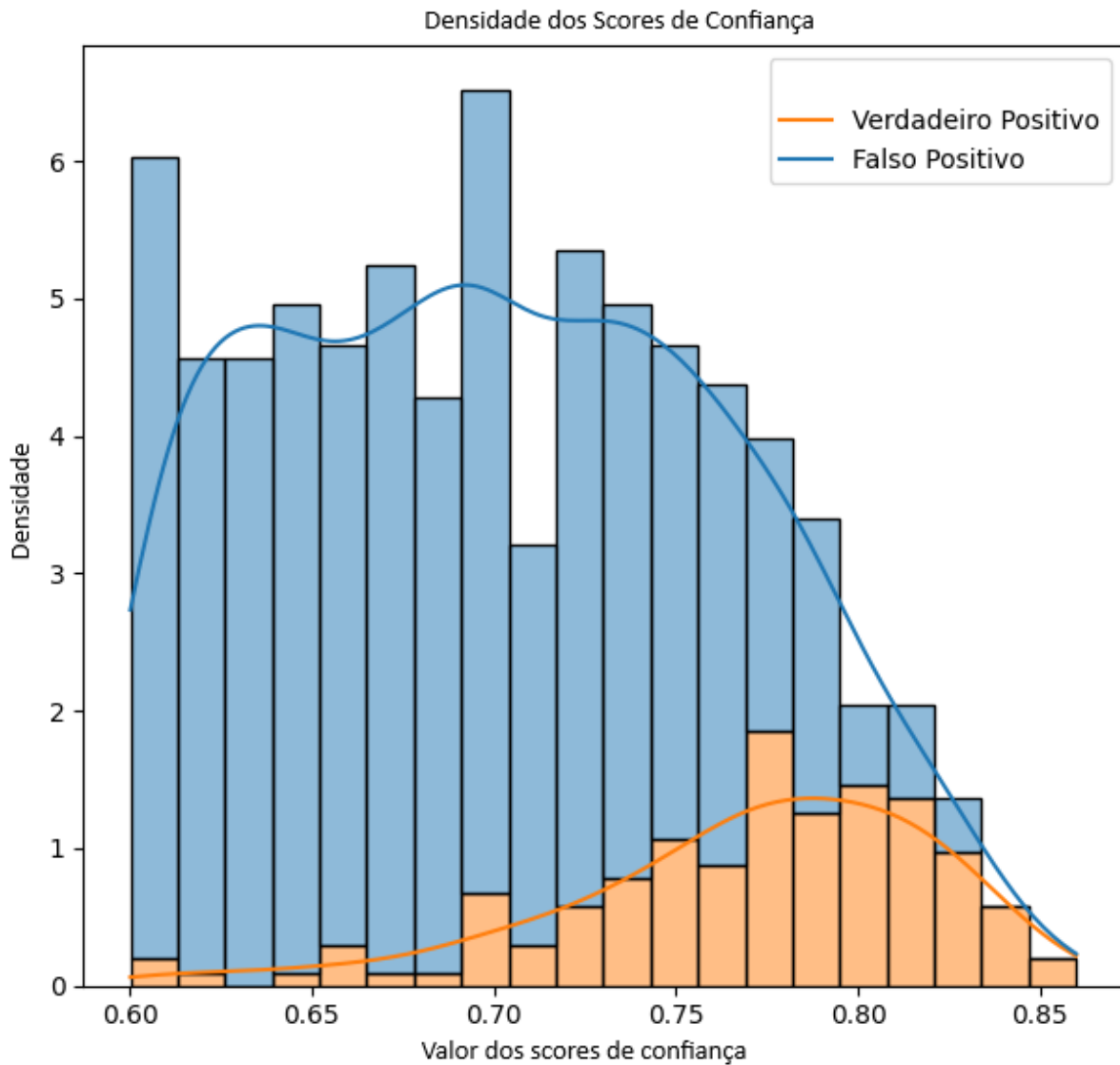
Nossas análises exploratórias nos dados de resultados de detecções do modelo revelaram algumas tendências relevantes. Observamos que muitos dos falsos positivos gerados pelo modelo corresponderam a caixas delimitadoras de tamanho menor em comparação com as caixas delimitadoras de detecções corretas. Isso sugere que o modelo pode estar identificando erroneamente características ou objetos menores no complexo plano de fundo do Pantanal como cervos-do-pantanal, contribuindo para a alta taxa de falsos positivos. Para melhor visualizar as relações entre as variáveis, uma amostragem dos resultados foi realizada de modo a manter a proporção de falsos positivos para verdadeiros positivos em 20:1, como ilustrado nas figuras 9 e 10. Na Figura 9, a densidade de detecções para diferentes valores de largura e altura das caixas delimitadoras é representada pelos histogramas. As médias das larguras e alturas foram 53,21 e 53,54 pixels, respectivamente, com desvios padrão de 14,57 e 15,75 pixels. A Figura 10 apresenta a densidade de detecções distribuídas por diferentes valores de score de confiança, agrupadas em 20 bins no intervalo de 0.6 a 0.85, com um tamanho de 0.0125 para cada bin. A curva de densidade suavizada indica que a maioria das detecções corretas têm scores de confiança entre 0.75 e 0.85, enquanto as detecções incorretas se concentram nas faixas menores, entre 0,625 e 0,75.

Figura 9 – Histogramas de amostragem do conjunto mostrando a densidade de detecções através de diferentes valores para largura e altura da caixa.



Fonte: Autor

Figura 10 – Histograma de amostragem do conjunto mostrando a densidade de detecções através de diferentes valores para score de confiança

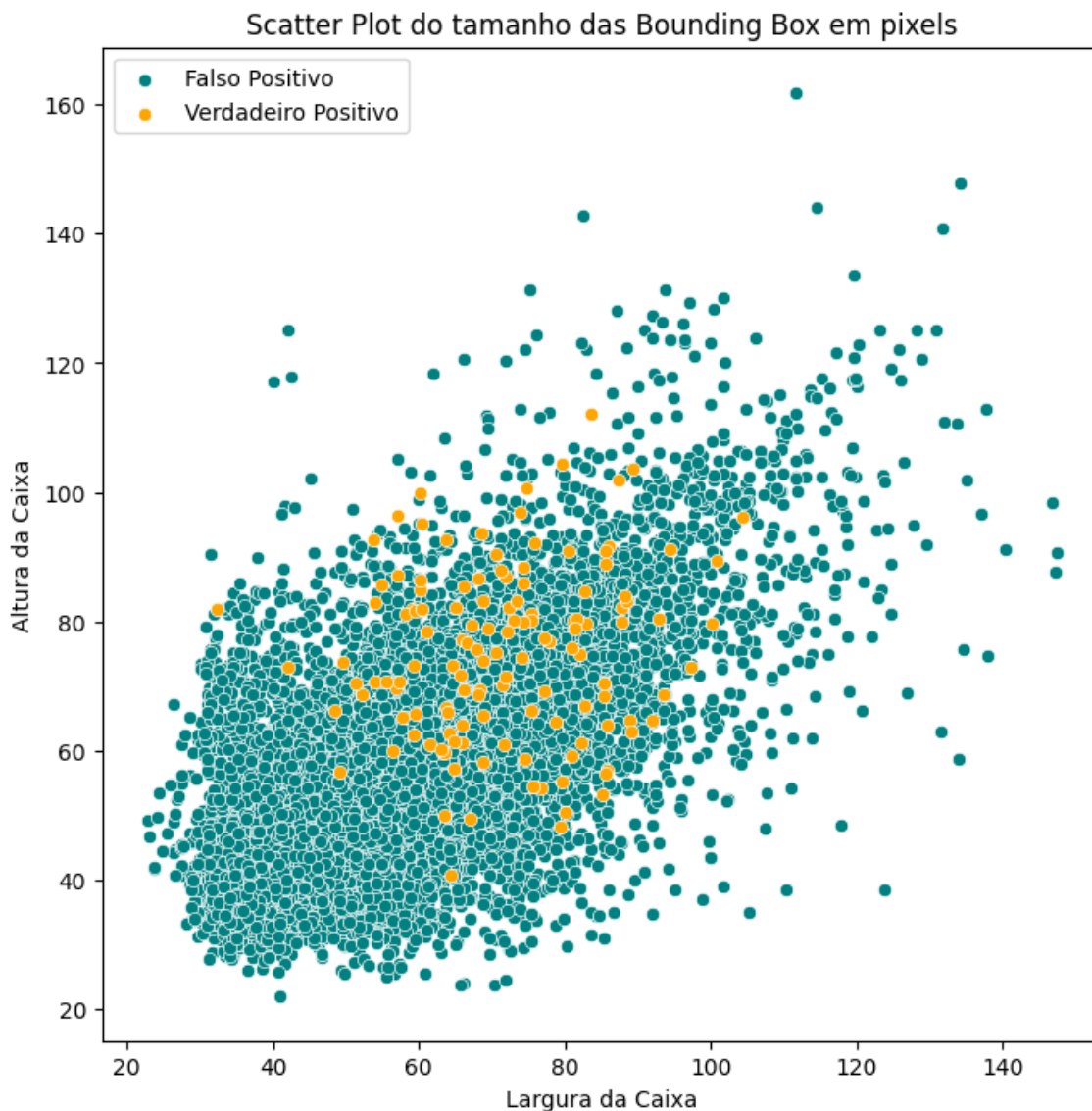


Fonte: Autor

A Figura 11 apresenta um *scatter plot* do tamanho das caixas delimitadoras em pixels, comparando a largura e altura das caixas para falsos positivos e verdadeiros positivos. Este gráfico mostra uma ampla dispersão de tamanhos para ambas as categorias, com uma concentração maior de detecções na faixa de 20 a 100 pixels tanto para largura quanto para altura. Nota-se que a distribuição dos falsos positivos (em azul) é mais ampla, incluindo muitas detecções de tamanho menor, enquanto os verdadeiros positivos (em laranja) parecem ter uma distribuição mais concentrada. A Figura 12 mostra um *scatter plot* do tamanho das caixas delimitadoras em comparação com o score de confiança, incluindo uma linha de tendência para os verdadeiros positivos. Este gráfico revela uma leve tendência positiva entre o tamanho da caixa e o score de confiança para as detecções corretas (verdadeiros

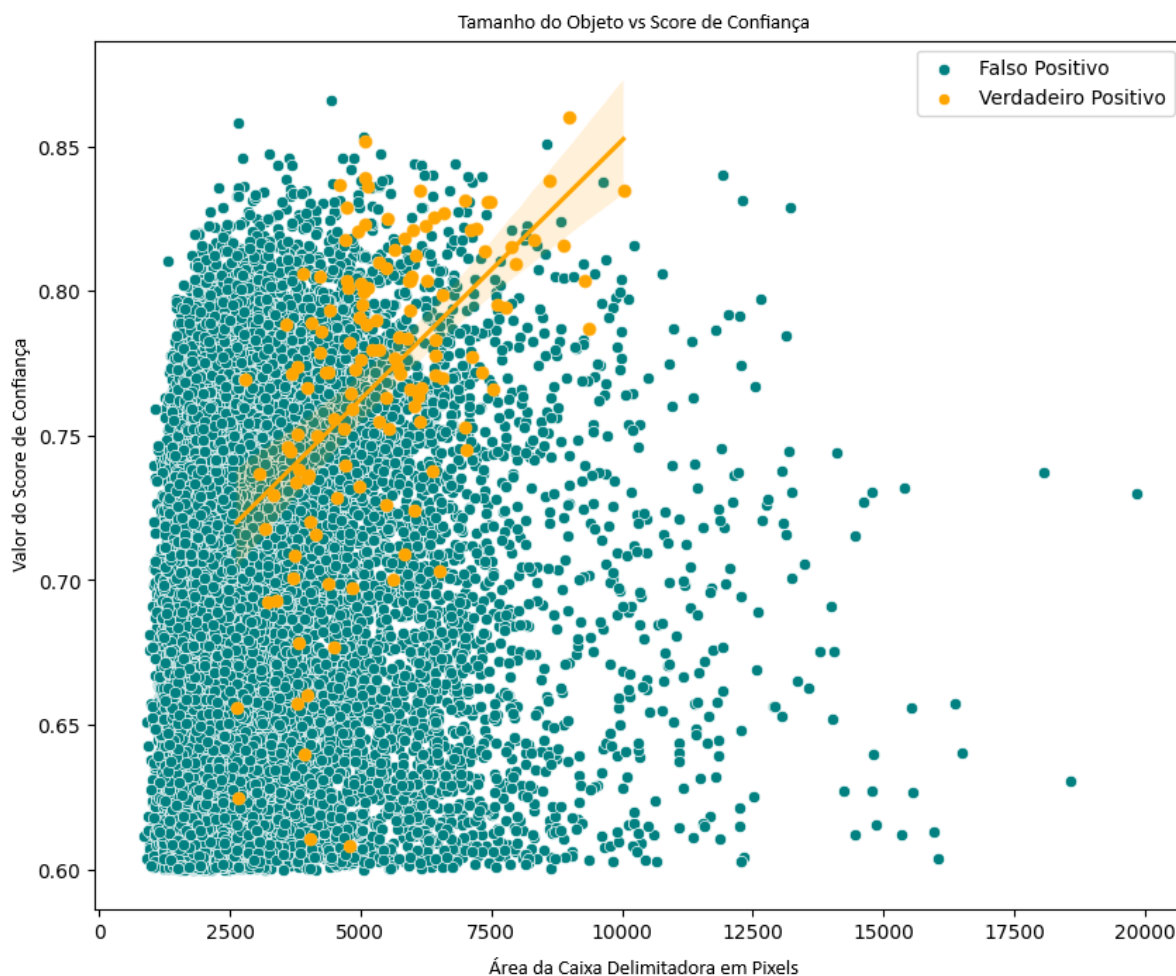
positivos). Observa-se que os falsos positivos se distribuem por uma ampla faixa de tamanhos e *scores* de confiança, enquanto os verdadeiros positivos tendem a se concentrar em *scores* de confiança ligeiramente mais altos, especialmente para objetos de tamanho médio.

Figura 11 – *Scatter plot* comparando altura e largura (em pixels) das caixas delimitadoras de detecções corretas e incorretas do modelo YOLOv8 treinado.



Fonte: Autor

Figura 12 – *Scatter plot* da área das caixas delimitadoras (em pixels) versus score de confiança com linha de tendência com campo de variação para os verdadeiros positivos



Fonte: Autor

Discussão

Nós treinamos um modelo algorítmico de aprendizado profundo de máquina de detecção de objeto para detectar cervos-do-pantanal em imagens aéreas obtidas com drones, utilizando o modelo YOLOv8. Avaliamos o desempenho do modelo analisando as taxas de precisão e revocação e realizamos análises exploratórias dos dados para entender os padrões de detecção.

A alta taxa de revocação demonstra a capacidade do modelo de reconhecer cervos-do-pantanal no conjunto de dados, sugerindo que ele aprendeu algumas características distintivas da espécie-alvo, apesar do tamanho limitado e da complexidade do conjunto de dados. No entanto, a precisão muito baixa revela um problema de previsões excessivas, com a tendência do modelo de identificar muitas

instâncias como cervos-do-pantanal, mesmo quando não eram, resultando em um elevado número de falsos positivos.

O tamanho relativamente pequeno do conjunto de dados de treinamento provavelmente limitou a capacidade de generalização do modelo para todas as variações presentes no ambiente complexo da paisagem do Pantanal. Conforme destacado por Krizhevsky *et al.* (2012) em seu trabalho seminal sobre aprendizado profundo para visão computacional, modelos de aprendizado profundo, como o YOLOv8 utilizado neste estudo, geralmente requerem conjuntos de dados de treinamento muito grandes, contendo milhões de exemplos, para atingir seu potencial máximo de generalização. Esses modelos possuem milhões de parâmetros ajustáveis e precisam ser expostos a uma ampla variedade de exemplos durante o treinamento para aprender efetivamente os padrões relevantes. No caso específico do YOLOv8, sua documentação recomenda pelo menos 1500 imagens por classe de interesse e pelo menos 10000 instâncias (objetos rotulados) por classe de interesse (Jocher *et al.*, 2023). A versão pré-treinada no conjunto COCO foi inicialmente exposta a mais de 300.000 imagens contendo mais de 2 milhões de instâncias rotuladas de 80 classes de objetos (Lin *et al.*, 2015). Esse grande volume de dados permitiu que o modelo adquirisse um conhecimento geral robusto sobre características visuais baixo-nível, como bordas, formas, texturas e padrões comumente encontrados em ambientes genéricos. No entanto, para a tarefa específica de detecção de cervos-do-pantanal no ambiente complexo do Pantanal, o conjunto de dados de treinamento utilizado neste estudo era consideravelmente menor, com apenas 573 amostras únicas de vídeos. Embora técnicas como aumento de dados tenham sido empregadas para gerar variações sintéticas, a diversidade e a quantidade de exemplos foi insuficiente para capturar toda a gama de variações presentes nesse ambiente natural desafiador (Zoph *et al.*, 2020).

Os ambientes naturais, especialmente aqueles com vegetação heterogênea e condições climáticas variáveis, como o Pantanal, apresentam uma complexidade muito maior em termos de iluminação, ângulos de visão, oclusões parciais e planos de fundo intrincados em comparação com os conjuntos de dados padrão. Estudos como o de Dai *et al.* (2021) sobre detecção de objetos em imagens aéreas reforçam os desafios de aplicar técnicas de aprendizado profundo de máquina em ambientes naturais complexos, onde a diversidade de condições ambientais exige que o modelo seja exposto a uma ampla variedade de exemplos durante o treinamento. O tamanho relativamente pequeno do conjunto de dados de treinamento pode ter limitado a capacidade do modelo de aprender a lidar com todas essas variações complexas, levando a dificuldades em distinguir com precisão os cervos-do-pantanal de outras características semelhantes presentes no ambiente. Isso pode ter contribuído para o alto número de falsos positivos observados durante a avaliação do modelo no conjunto de teste.

Os padrões observados nas análises exploratórias oferecem caminhos para possíveis melhorias. A incorporação de um filtro ou a criação de uma estratégia de treinamento que leve em conta o tamanho das caixas delimitadoras poderia aumentar a precisão

do modelo. Especificamente, um limite de tamanho poderia ser definido de modo que caixas delimitadoras abaixo de um determinado tamanho não fossem consideradas como possíveis detecções. Isso poderia reduzir significativamente o número de falsos positivos, sendo mais eficiente em um possível cenário de uso real. No entanto, deve-se tomar cuidado ao definir esse limite para evitar a exclusão de detecções válidas de cervos-do-pantanal que possam parecer menores devido a fatores como seu formato no momento da captura (por exemplo, se ele estiver curvado, diminuindo seu tamanho de caixa) (Liu *et al.*, 2021). Portanto, é necessário investigar mais para determinar um limite de tamanho ideal que equilibre a redução de falsos positivos com a preservação de verdadeiros positivos.

É possível que as técnicas de ampliação dos dados aplicadas durante o pré-processamento das imagens tenham exercido influência sobre a capacidade discriminativa do modelo YOLOv8. Em particular, as transformações de aumento de dados que envolvem alterações nos canais de cor, tais como ajustes de matiz (hue) e saturação, podem ter induzido o modelo a aprender representações pouco fidedignas quanto às características cromáticas intrínsecas aos cervos-do-pantanal. He *et al.* (2019), em seu trabalho sobre a lacuna de distribuição entre dados limpos e aumentados, destacam os desafios envolvidos no uso de técnicas de aumento de dados, como ajustes de matiz e saturação, que podem distorcer características importantes, como as cromáticas, prejudicando a capacidade discriminativa dos modelos. Essa hipótese ganha força ao considerarmos que a correta percepção das nuances de cor é um aspecto importante para a detecção acurada dessa espécie, haja vista a complexidade dos fundos presentes nas imagens capturadas pelos drones nas regiões de estudo. Ao ser exposto a variações excessivas nos valores de matiz e saturação durante o treinamento, o modelo pode ter perdido a capacidade de associar os padrões cromáticos típicos dos cervídeos a instâncias genuínas dessa classe, culminando em um aumento na taxa de falsos positivos para objetos de coloração distinta, porém com características estruturais semelhantes (Zoph *et al.*, 2020; He *et al.*, 2019).

Para aumentar a precisão do modelo, várias estratégias podem ser consideradas. Aumentar o tamanho do conjunto de dados de treinamento deve ajudar, pois forneceria uma gama mais ampla de instâncias para o modelo aprender. Isto destaca a importância da manutenção a longo prazo de um projeto de monitoramento via drones como o utilizado no estudo. Novas técnicas de aumento de dados podem ser exploradas também, como técnicas de geração de dados sintéticos (Premakumara *et al.*, 2023). Além disso, o ajuste fino do modelo por meio do ajuste de hiperparâmetros ou do emprego de técnicas de regularização (métodos para evitar o sobreajuste do modelo) pode ajudar a controlar a tendência do modelo de prever em excesso (Premakumara *et al.*, 2023). Uma abordagem promissora, conforme explorada por Knyaz (2019) em seu estudo sobre fusão de dados multimodais para detecção de objetos em imagens aéreas, é a combinação de imagens RGB com outras fontes de dados, como imagens térmicas ou dados LiDAR, fornecendo informações

complementares para auxiliar na distinção dos objetos de interesse. (Knyaz, 2019; Povlsen *et al.*, 2024; Lyu *et al.*, 2024)

A ocorrência significativa de falsos positivos observada aponta para a necessidade de mais desenvolvimento e refinamento antes que essa abordagem possa ser amplamente implementada. A implementação bem-sucedida dessa abordagem de detecção automática baseada em visão computacional em ambientes naturais enfrenta desafios práticos e econômicos significativos. Conforme discutido por Xu *et al.* (2024) em sua revisão sobre os desafios da aplicação de aprendizado profundo na conservação da vida selvagem, a aquisição de equipamentos como drones e a infraestrutura computacional necessária para o treinamento e implantação de modelos de visão computacional podem representar custos iniciais substanciais. Além disso, os autores destacam a necessidade de treinamento e capacitação adequados para os profissionais envolvidos no manejo e conservação das espécies-alvo, a fim de garantir a utilização eficaz dessas tecnologias. Embora os benefícios a longo prazo possam compensar os investimentos iniciais, é fundamental considerar esses fatores ao planejar a adoção de soluções baseadas em visão computacional em projetos de monitoramento e conservação (Fritsch *et al.*, 2020; Schad *et al.*, 2023).

É importante ressaltar que, mesmo com as melhorias necessárias, a detecção automática baseada em visão computacional não substitui completamente os esforços humanos de revisão de imagens. Em vez disso, ela deve ser vista como uma ferramenta complementar que, quando integrada de forma adequada a um plano abrangente de monitoramento e manejo, pode fornecer dados mais precisos e oportunos para embasar decisões críticas relacionadas à proteção de espécies ameaçadas e à preservação da biodiversidade.

Conclusão

Este estudo investigou o desempenho de um modelo YOLOv8 treinado para detectar cervos-do-pantanal em imagens capturadas no complexo ambiente do Pantanal. Embora o modelo tenha alcançado uma taxa de revocação considerável, de 85,1% (132 de 155 cervos), foi observado um número substancial de falsos positivos (18.542). Esses resultados revelam os desafios associados à detecção de objetos em ambientes naturais complexos, onde o modelo pode ser suscetível a classificar erroneamente elementos de fundo como o objeto-alvo.

Uma das principais vantagens da utilização de drones e algoritmos de detecção automática é a capacidade de cobrir extensas áreas de forma rápida e eficiente, reduzindo drasticamente o esforço humano necessário para revisar manualmente todas as imagens capturadas. Essa eficiência pode ser particularmente valiosa em regiões remotas e de difícil acesso, como o Pantanal, onde o monitoramento tradicional enfrenta desafios logísticos e financeiros significativos. No entanto, para que essa abordagem seja viável na prática, é essencial que o modelo atinja níveis

mais altos de precisão, minimizando os falsos positivos que podem comprometer a confiabilidade dos dados de monitoramento.

A acurácia e precisão relativamente baixa ressaltam o impacto significativo dos falsos positivos no desempenho geral do modelo. Embora a alta taxa de revocação indique a capacidade do modelo de identificar com êxito a maioria dos verdadeiros positivos, a compensação entre revocação e precisão precisa ser considerada com cuidado. Nesse caso, a obtenção de uma alta taxa de revocação tem o custo de uma alta taxa de falsos positivos, o que pode ser problemático em cenários do mundo real em que alarmes falsos podem acarretar custos ou consequências significativas.

Vários fatores provavelmente contribuíram para as limitações observadas no modelo. O tamanho limitado do conjunto dos dados de treinamento pode ter prejudicado a capacidade do modelo de aprender toda a gama de variações presentes no complexo fundo do Pantanal, levando a classificações errôneas de objetos de aparência semelhante. Além disso, a complexidade inerente do próprio ambiente natural, caracterizado por vegetação diversa e condições de iluminação variadas, pode representar desafios para os algoritmos de detecção de objetos. Por fim, as observações também ressaltam a importância de compreender as características e as nuances dos dados com os quais se está trabalhando. Ao examinar de perto os padrões dos falsos positivos gerados pelo modelo, podemos identificar possíveis limitações e elaborar estratégias para melhorar o desempenho do modelo, como o uso de uma seleção mais otimizada de métodos de transformação e aumento dos dados.

Referências:

BARASONA, J. Á. *et al.* Unmanned Aircraft Systems for Studying Spatial Abundance of Ungulates: Relevance to Spatial Epidemiology. *PLoS ONE*, v. 9, n. 12, p. e115608, 2014. Disponível em: <https://doi.org/10.1371/journal.pone.0115608>. Acesso em: 30 maio 2023.

BRACK, I. V. *et al.* Detection Errors in Wildlife Abundance Estimates from Unmanned Aerial Systems (UAS) Surveys: Synthesis, Solutions, and Challenges. *Methods in Ecology and Evolution*, v. 9, n. 8, p. 1864-1873, ago. 2018. Disponível em: <https://doi.org/10.1111/2041-210X.13026>. Acesso em: 30 maio 2023.

BRACK, I. V. *et al.* Spatial Variation on the Abundance of a Threatened South American Large Herbivore Using Spatiotemporally Replicated Drone Surveys. *Biodiversity and Conservation*, fev. 2023. Disponível em: <https://doi.org/10.1007/s10531-023-02553-7>. Acesso em: 30 maio 2023.

BRACK, I. V. Modelagem de abundância com drones: detectabilidade, desenho amostral e revisão automática de imagens em um estudo com cervos-do-pantanal.

2022. Disponível em: <https://lume.ufrgs.br/handle/10183/247228>. Acesso em: 30 maio 2023

BRODRICK, P. G. *et al.* Uncovering Ecological Patterns with Convolutional Neural Networks. *Trends in Ecology & Evolution*, v. 34, n. 8, p. 734-745, ago. 2019. Disponível em: <https://doi.org/10.1016/j.tree.2019.03.006>. Acesso em: 30 maio 2023.

CEBALLOS, G. *et al.* Biological Annihilation via the Ongoing Sixth Mass Extinction Signaled by Vertebrate Population Losses and Declines. *Proceedings of the National Academy of Sciences*, vol. 114, n. 30, julho de 2017. DOI.org (Crossref), <https://doi.org/10.1073/pnas.1704949114>. Acesso em: 30 maio 2023.

CEBALLOS, G. *et al.* Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences USA*, v. 117, p. 13596-13602, 2020.

CHABOT, D. *et al.* Computer-automated bird detection and counts in high-resolution aerial images: a review. *Journal of Field Ornithology*, v. 87, n. 4, p. 343-359, 2016. Disponível em: <https://doi.org/10.1111/jof.12171>. Acesso em: 30 maio 2023.

CHABOT, D. *et al.* An approach for using off-the-shelf object-based image analysis software to detect and count birds in large volumes of aerial imagery. *Avian Conservation and Ecology*, v. 13, n. 1, 2018. Disponível em: <https://doi.org/10.5751/ACE-01205-130115>. Acesso em: 30 maio 2023

CORCORAN, E. *et al.* Automated detection of koalas using low-level aerial surveillance and machine learning. *Scientific Reports*, v. 9, n. 1, p. 3208, 2019. Disponível em: <https://doi.org/10.1038/s41598-019-39917-5>. Acesso em: 30 maio 2023.

CHRÉTIEN, L.-P.; THÉAU, J.; MÉNARD, P. Visible and thermal infrared remote sensing for the detection of white-tailed deer using an unmanned aerial system. *Wildlife Society Bulletin*, v. 40, n. 1, p. 181-191, 2016. Disponível em: <https://doi.org/10.1002/wsb.629> <https://doi.org/10.1002/wsb.629> Acesso em: 30 maio 2023.

COSTA, S. S. *et al.* Plants Composing the Diet of Marsh and Pampas Deer in the Brazilian Pantanal Wetland and Their Ethnomedicinal Properties. *Journal of Biological Sciences*, v. 6, n. 5, p. 840–846, 2006. Disponível em: <https://doi.org/10.3923/jbs.2006.840.846>. Acesso em: 30 maio 2023.

DUARTE, J. M. B. *et al.* *Blastocerus dichotomus*. The IUCN Red List of Threatened Species, 2016. Disponível em: <https://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T2828A22160916.en>. Acesso em: 30 maio 2023.

Dai, J. *et al.* Overview of UAV Target Detection Algorithms Based on Deep Learning. *2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, Chongqing, China, 2021, pp. 736-745, Disponível em: doi: 10.1109/ICIBA52610.2021.9688305. Acesso em: 30 maio 2024.

- EIKELBOOM, J. A. J. *et al.* Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, v. 10, n. 11, p. 1875–1887, 29 ago. 2019. Disponível em: <https://doi.org/10.1111/2041-210X.13277>. Acesso em: 30 julho 2024
- FERREIRA, S. M.; AARDE, R. J. Van. Aerial Survey Intensity as a Determinant of Estimates of African Elephant Population Sizes and Trends. *South African Journal of Wildlife Research*, v. 39, n. 2, p. 181–191, 2009. Disponível em: <https://doi.org/10.3957/056.039.0205>. Acesso em: 30 maio 2023.
- FRANKE, U. *et al.* Aerial ungulate surveys with a combination of infrared and high-resolution natural colour images. *Animal Biodiversity and Conservation*, v. 35, n. 2, p. 285–293, dez. 2012. Disponível em: <https://doi.org/10.32800/abc.2012.35.0285>. Acesso em: 30 julho 2024
- FRITSCH, C. J.; DOWNS, C. T. Evaluation of low-cost consumer-grade UAVs for conducting comprehensive high-frequency population censuses of hippopotamus populations. *Conservation Science and Practice*, v. 2, n. 12, p. 1–11, 2020. Disponível em: <https://doi.org/10.1111/csp2.281>. Acesso em: 30 maio 2023.
- GENTLE, M. *et al.* A comparison of unmanned aerial vehicles (drones) and manned helicopters for monitoring macropod populations. *Wildlife Research*, v. 45, n. 7, p. 586, 2018. Disponível em: <https://doi.org/10.1071/WR18034>. Acesso em: 30 maio 2023.
- GERUM, R. C. *et al.* ClickPoints: an expandable toolbox for scientific image annotation and analysis. *Methods in Ecology and Evolution*, v. 8, n. 6, p. 750–756, 2017. Disponível em: <https://doi.org/10.1111/2041-210X.12702>. Acesso em: 30 maio 2023.
- HABIB, T. J. *et al.* Detection and stratification approaches for aerial surveys of deer in prairie-parklands. *Wildlife Research*, v. 39, n. 7, p. 593, 2012. Disponível em: <https://doi.org/10.1071/WR11175>. Acesso em: 30 julho 2024
- HODGSON, J. C. *et al.* Drones Count Wildlife More Accurately and Precisely than Humans. *Methods in Ecology and Evolution*, v. 9, n. 5, p. 1160–1167, May 2018. Disponível em: <https://doi.org/10.1111/2041-210X.12974>. Acesso em: 30 maio 2023.
- HOLLINGS, T. *et al.* How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *Methods in Ecology and Evolution*, v. 9, n. 4, p. 881–892, 2018. Disponível em: <https://doi.org/10.1111/2041-210X.12973>. Acesso em: 30 maio 2023.
- IVANOVA, S.; PROSEKOV, A.; KALEDIN, A. A Survey on Monitoring of Wild Animals during Fires Using Drones. *Fire*, v. 5, n. 3, p. 60, 29 abr. 2022. Disponível em: <https://doi.org/10.3390/fire5030060>. Acesso em: 30 julho 2024
- JOCHER, G. *et al.* Ultralytics YOLO. Código Fonte. Ultralytics, 2023. Disponível em: <https://github.com/ultralytics/ultralytics>. Acesso em: 30 maio 2023.

KOGER, B. *et al.* Quantifying the Movement, Behaviour and Environmental Context of Group-living Animals Using Drones and Computer Vision. *Journal of Animal Ecology*, p. 1365-2656.13904, Mar. 2023. Disponível em: <https://doi.org/10.1111/1365-2656.13904>. Acesso em: 30 maio 2023.

KOH, L. P.; WICH, S. A. Dawn of drone ecology: Low-cost autonomous aerial vehicles for conservation. *Tropical Conservation Science*, v. 5, p. 121–132, 2012. Disponível em: <https://doi.org/10.1177/194008291200500202>. Acesso em: 30 maio 2023.

LENZI, J. *et al.* Artificial intelligence for automated detection of large mammals creates path to upscale drone surveys. *Scientific Reports*, v. 13, n. 1, 18 jan. 2023. Disponível em: <https://www.nature.com/articles/s41598-023-28240-9>. Acesso em: 30 maio 2023.

LEUNG, B.; GONZALEZ, A. Global monitoring for biodiversity: Uncertainty, risk, and power analyses to support trend change detection. *Science Advances*, v. 10, n. 7, p. eadj1448, 16 fev. 2024. Disponível em: <https://www.science.org/doi/epdf/10.1126/sciadv.adj1448> Acesso em: 20 agosto 2024.

LIN, Tsung-Yi *et al.* Microsoft COCO: Common Objects in Context. 2015. Disponível em: <https://arxiv.org/abs/1405.0312>. Acesso em: 30 maio 2023.

LIU, Yang *et al.* A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, v. 172, 114602, 2021. DOI: 10.1016/j.eswa.2021.114602. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417421000439>. Acesso em: 30 abril 2024.

MARIN, V. C. *et al.* Diet of the marsh deer in the Paraná River Delta, Argentina—a vulnerable species in an intensive forestry landscape. *European Journal of Wildlife Research*, v. 66, n. 1, p. 16-18, 2020. Disponível em: <https://doi.org/10.1007/s10344-019-1358-3>. Acesso em: 30 maio 2023.

MILLIGAN, N. *et al.* AERIAL SURVEYS IN THE STUDY OF ANIMAL POPULATIONS AND RANGE CONDITIONS. *Nigerian Journal of Animal Production*, v. 4, n. 1, p. 1–18, 16 jan. 2021. Disponível em: <https://doi.org/10.51791/NJAP.V4I1.2561>. Acesso em: 30 julho 2024.

MULERO-PÁZMÁNY, M. *et al.* Remotely piloted aircraft systems as a rhinoceros anti-poaching tool in Africa. *PLoS ONE*, v. 9, p. e83873, 2014. Disponível em: <https://doi.org/10.1371/journal.pone.0083873>. Acesso em: 30 maio 2023.

NAVARRO, L. M. *et al.* Monitoring Biodiversity Change through Effective Global Coordination. *Current Opinion in Environmental Sustainability*, vol. 29, dezembro de 2017, p. 158–169. DOI.org (Crossref), <https://doi.org/10.1016/j.cosust.2018.02.005>. Acesso em: 30 maio 2023.

OLLIER, C. *et al.* Matching visual and acoustic events to estimate detection probability for small cetaceans in the ACCOBAMS Survey Initiative. *Frontiers in marine science*, v. 10, 16 out. 2023. Disponível em: <https://doi.org/10.3389/fmars.2023.1244474>. Acesso em: 30 de julho de 2024.

OLIVEIRA, L. F. B. *et al.* Padrões e tendências espaço-temporais na estrutura de uma paisagem antropizada no norte do Pantanal. In: PERES, C. A. *et al.* (Org.). *Conservação da Biodiversidade em Paisagens Antropizadas do Brasil*. Editora UFPR, 2013. p. 231-26. Acesso em: 30 abril 2024.

OLEKSYN, S. *et al.* Going Batty: The Challenges and Opportunities of Using Drones to Monitor the Behaviour and Habitat Use of Rays. *Drones*, v. 5, n. 1, p. 12, 2 fev. 2021. Disponível em: <https://doi.org/10.3390/DRONES5010012>. Acesso em: 30 julho 2024

PADILLA, R. *et al.* A survey on performance metrics for object-detection algorithms. In: INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING (IWSSIP), 2020. p. 237–242. Disponível em: <https://doi.org/10.1109/IWSSIP48289.2020.9145130>. Acesso em: 30 abril 2024.

PATTERSON, C. *et al.* Evaluation of an unmanned aircraft system for detecting surrogate caribou targets in Labrador. *Journal of Unmanned Vehicle Systems*, vol. 4, n. 1, março de 2016, p. 53–69. Atypon, <https://doi.org/10.1139/juvs-2015-0014>. Acesso em: 30 maio 2023.

POVLSEN, P. *et al.* A Novel Scouring Method to Monitor Nocturnal Mammals Using Uncrewed Aerial Vehicles and Thermal Cameras—A Comparison to Line Transect Spotlight Counts. *Drones*, v. 7, n. 11, p. 661, 1 nov. 2023. Disponível em: <https://doi.org/10.3390/drones7110661>. Acesso em: 30 julho 2024

PREMAKUMARA, N. *et al.* Improving Object Detection Robustness against Natural Perturbations through Synthetic Data Augmentation. 28 abr. 2023. Disponível em: <https://doi.org/10.1145/3596286.3596293>. Acesso em: 30 julho 2024

RANČIĆ, K. *et al.* Animal detection and counting from UAV images using convolutional neural networks. *Drones*, v. 7, n. 3, p. 179, 2023. Disponível em: <https://doi.org/10.3390/drones703017>. Acesso em: 30 maio 2023.

SCHAD, L.; FISCHER, J. Opportunities and Risks in the Use of Drones for Studying Animal Behaviour. *Methods in Ecology and Evolution*, p. 2041-210X.13922, June 2022. Disponível em: <https://doi.org/10.1111/2041-210X.13922>. Acesso em: 30 maio 2023.

TUIA, D. *et al.* Perspectives in Machine Learning for Wildlife Conservation. *Nature Communications*, v. 13, n. 1, p. 792, Feb. 2022. Disponível em: <https://doi.org/10.1038/s41467-022-27980-y>. Acesso em: 30 maio 2023.

- WITMER, G. W. Wildlife Population Monitoring: Some Practical Considerations. *Wildlife Research*, v. 32, n. 3, p. 259–263, June 2005. Disponível em: <https://doi.org/10.1071/WR04003>. Acesso em: 30 maio 2023.
- YING, X. An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, v. 1168, n. 2, p. 022022, 2019. DOI: 10.1088/1742-6596/1168/2/022022. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1168/2/022022/meta>. Acesso em: 30 abril 2024.
- KNYAZ, V. Multimodal data fusion for object recognition. In: *Multimodal sensing: technologies and applications*. Proceedings of SPIE, v. 11059, p. 110590P, 21 jun. 2019. DOI: 10.1117/12.2526067. Disponível em: <https://doi.org/10.1117/12.2526067>. Acesso em: 30 abril 2024.
- GROOM, G. *et al.* Using Object-Based Analysis of Image Data to Count Birds: Mapping of Lesser Flamingos at Kamfers Dam, Northern Cape, South Africa. *International Journal of Remote Sensing*, v. 32, n. 16, p. 4611–4639, Aug. 2011. Disponível em: <https://doi.org/10.1080/01431161.2010.489068>. Acesso em: 30 maio 2023.
- MAWDSLEY, J. R.; O'MALLEY, R.; OJIMA, D. S. A Review of Climate-Change Adaptation Strategies for Wildlife Management and Biodiversity Conservation. *Conservation Biology*, v. 23, n. 5, p. 1080–1089, out. 2009. Disponível em: <https://doi.org/10.1111/j.1523-1739.2009.01264.x> Acesso em: 30 janeiro 2024.
- MOURÃO, G. *et al.* Aerial surveys of caiman, marsh deer and pampas deer in the Pantanal Wetland of Brazil. *Biological Conservation*, v. 92, n. 2, p. 175–183, fev. 2000. Disponível em: [https://doi.org/10.1016/S0006-3207\(99\)00051-8](https://doi.org/10.1016/S0006-3207(99)00051-8) Acesso em: 30 janeiro 2024.
- AKYON, F. C. *et al.* Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. 2022 IEEE International Conference on Image Processing (ICIP). IEEE, Out. 2022. DOI: 10.1109/icip46576.2022.9897990. Disponível em: <http://dx.doi.org/10.1109/ICIP46576.2022.9897990>. Acesso em: 28 fevereiro 2024.
- POVLSEN, P. *et al.* Using YOLO object detection to identify hare and roe deer in thermal aerial video footage—possible future applications in real-time automatic drone surveillance and wildlife monitoring. *Drones*, v. 8, n. 1, p. 2, 2024. DOI: 10.3390/drones8010002. Disponível em: <https://doi.org/10.3390/drones8010002>. Acesso em: 30 abril 2024.
- HE, Z. *et al.* Data augmentation revisited: rethinking the distribution gap between clean and augmented data. 2019. Disponível em: <https://arxiv.org/abs/1909.09148>. Acesso em: 30 abril 2024.
- ZOPH, B. *et al.* Learning data augmentation strategies for object detection. In: VEDALDI, A.; BISCHOF, H.; BROX, T.; FRAHM, J. M. (eds.). *Computer vision – ECCV 2020*. ECCV 2020. Lecture Notes in Computer Science, v. 12372. Springer,

Cham, 2020. DOI: 10.1007/978-3-030-58583-9_34. Disponível em: https://doi.org/10.1007/978-3-030-58583-9_34. Acesso em: 30 abril 2024.

XU, Z. *et al.* A review of deep learning techniques for detecting animals in aerial and satellite images. *International Journal of Applied Earth Observation and Geoinformation*, v. 128, 103732, 2024. DOI: 10.1016/j.jag.2024.103732. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1569843224000864>. Acesso em: 30 abril 2024.

KRIZHEVSKY, A. *et al.* ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, v. 60, n. 6, p. 84–90, jun. 2017. DOI: 10.1145/3065386. Disponível em: <https://doi.org/10.1145/3065386>. Acesso em: 30 abril 2024.

LYU, H. *et al.* Deer survey from drone thermal imagery using enhanced faster R-CNN based on ResNets and FPN. *Ecological Informatics*, v. 79, 102383, 2024. DOI: 10.1016/j.ecoinf.2023.102383. Disponível em: <https://doi.org/10.1016/j.ecoinf.2023.102383>. Acesso em: 30 abril 2024.

