

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ARQUITETURA E URBANISMO**

Jaceguay Zukoski

Visão Computacional - Relatório:
Análise de Áreas Costeiras com Ortoimagens de Drones: Segmentação e Classificação

Florianópolis

2024

1. Introdução

Este estudo investiga o uso de técnicas de visão computacional na análise de ambientes litorâneos a partir de ortoimagens obtidas por drones. A fotogrametria possibilita gerar imagens referenciadas em sistemas métricos, onde cada pixel representa uma medida espacial conhecida. Esta condição facilita o monitoramento de áreas costeiras, permitindo quantificar a extensão de vegetação ou areia e avaliar a presença de pessoas em escala útil à gestão ambiental. Aqui, compara-se uma abordagem clássica de segmentação por cores com métodos baseados em aprendizado profundo, examinando suas vantagens, limitações e aplicabilidade.

2. Contextualização do Problema

A ocupação humana em áreas costeiras impacta a distribuição da vegetação e a integridade das faixas de areia. A necessidade de acompanhar a evolução dessas alterações justifica o uso de ferramentas automáticas de análise. Imagens obtidas por drones cobrem grandes áreas de modo ágil, enquanto a fotogrametria gera ortoimagens métricas precisas, como no sistema SIRGAS 2000 UTM 22S. Essas condições permitem quantificar e comparar dados ao longo do tempo, contribuindo para decisões de manejo, conservação e planejamento territorial.

3. Metodologia e Resultados

3.1 Estratégias Clássicas de Segmentação por Cores

A metodologia inicial envolveu processamento de imagens com OpenCV. A ortoimagem, carregada a partir de um arquivo GeoTIFF, foi convertida do espaço de cor BGR (padrão do OpenCV) para RGB e, em seguida, para HSV. Ajustaram-se intervalos de matiz, saturação e valor para segmentar a vegetação e a areia. Ao aplicar máscaras binárias, isolaram-se áreas vegetadas e obteve-se uma máscara homogênea, permitindo o cálculo da área em metros quadrados, multiplicando o número de pixels pela resolução espacial do levantamento (0,016 m/pixel).

A área de restinga foi destacada de modo homogêneo com sucesso obtendo-se a área de 2241.28 m². Já no caso da areia, mesmo após o teste com várias faixas de cores para a presença de objetos heterogêneos (guarda-sóis, toalhas e pessoas) introduziu cores atípicas, dificultando a separação precisa. Mesmo com filtros adicionais, não foi possível obter uma máscara pura. Essas variações cromáticas destacam a limitação dos métodos clássicos: são rápidos e diretos, porém altamente sensíveis à variabilidade de cor, iluminação e objetos inesperados.

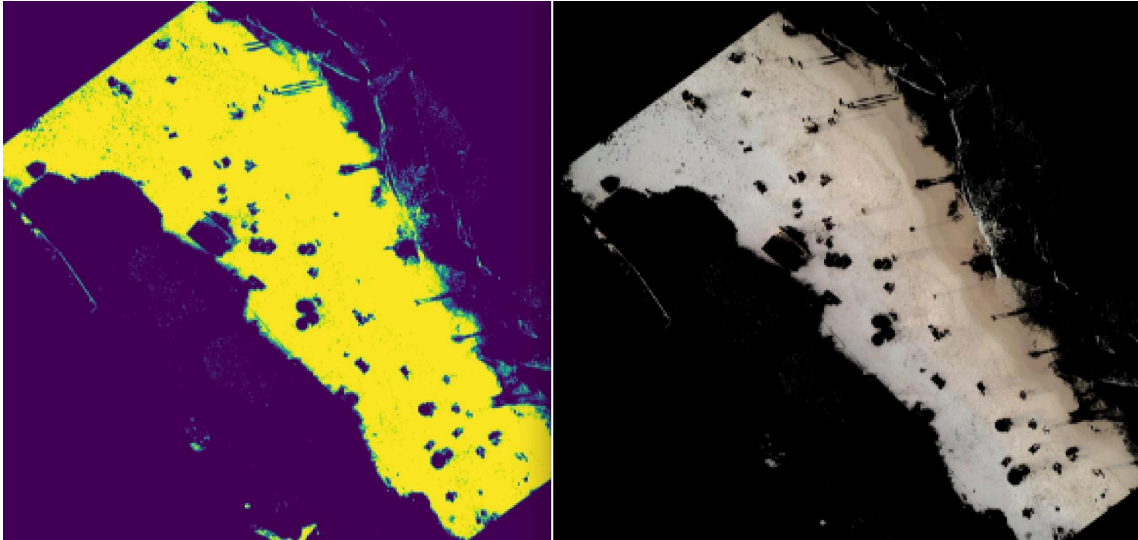


Figura 1. Máscara e recorte a partir do intervalo de cores atribuído a areia. (fonte autor)

Apesar das limitações, o método por cores forneceu uma linha de base, demonstrando a viabilidade de extrair áreas aproximadas de vegetação e areia. Porém, a necessidade de técnicas mais robustas tornou-se evidente.

3.2 Treinamento de um Modelo Pré-treinado em Aprendizado Profundo

Para superar as limitações da abordagem clássica, adotou-se uma estratégia baseada em aprendizado profundo, empregando PyTorch e Torchvision. Inicialmente, organizou-se um conjunto de dados a partir de multiplas ortoimagens de diferentes regiões: as imagens foram divididas em quadrículas de 1x1 metro, refletindo a resolução espacial da cena. Cada quadrícula foi classificada manualmente em uma das classes alvo (restinga e areia). Estas amostras foram responsáveis pelo conjunto de treinamento e teste.

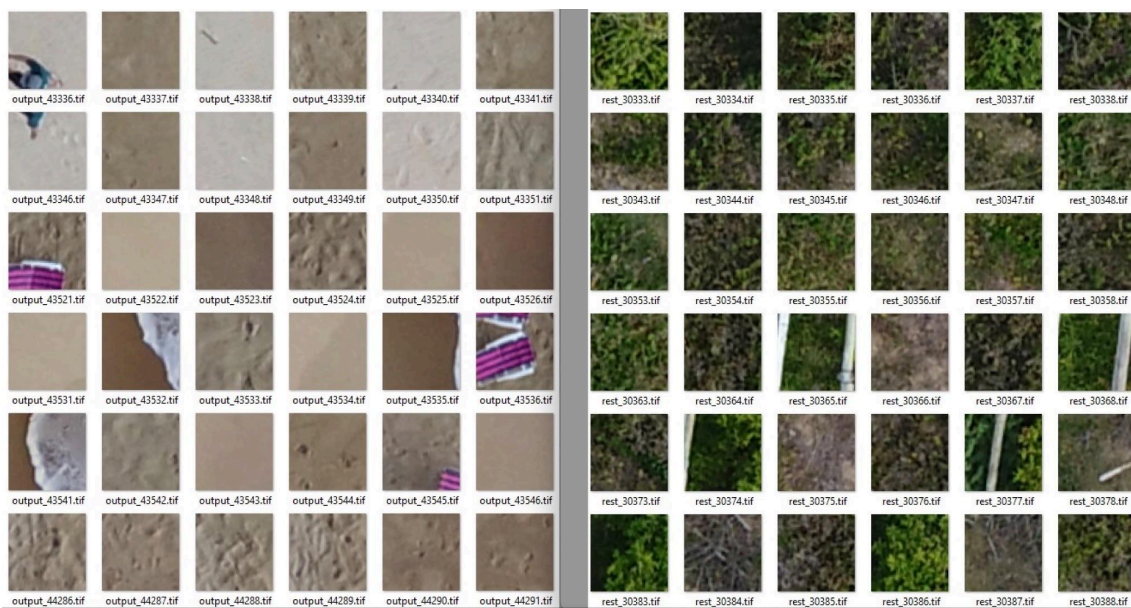


Figura 2. Amostras separadas em pastas distintas, restinga e areia. (fonte autor)

Utilizou-se um modelo ResNet18 pré-treinado no ImageNet. A última camada totalmente conectada (fc) foi substituída para refletir o número de classes do problema. O treinamento ocorreu em 10 épocas, usando o otimizador Adam com taxa de aprendizado de 0,001 e a função de perda CrossEntropyLoss. Durante o treinamento, foram aplicadas transformações de dados (data augmentation), como redimensionamento para 256x256 pixels, normalização dos valores (mean=[0.5,0.5,0.5], std=[0.5,0.5,0.5]), rotação aleatória em ângulos de 90°, 180°, 270° e inversões horizontais. Essas transformações tem por objetivo aumentar a variedade das amostras.

Os resultados mostraram rápida queda do loss de treinamento nas primeiras épocas e alta acurácia de validação já nos estágios iniciais, atingindo cerca de 0,99. Isso indica que o modelo pré-treinado captou facilmente as características visuais relevantes.

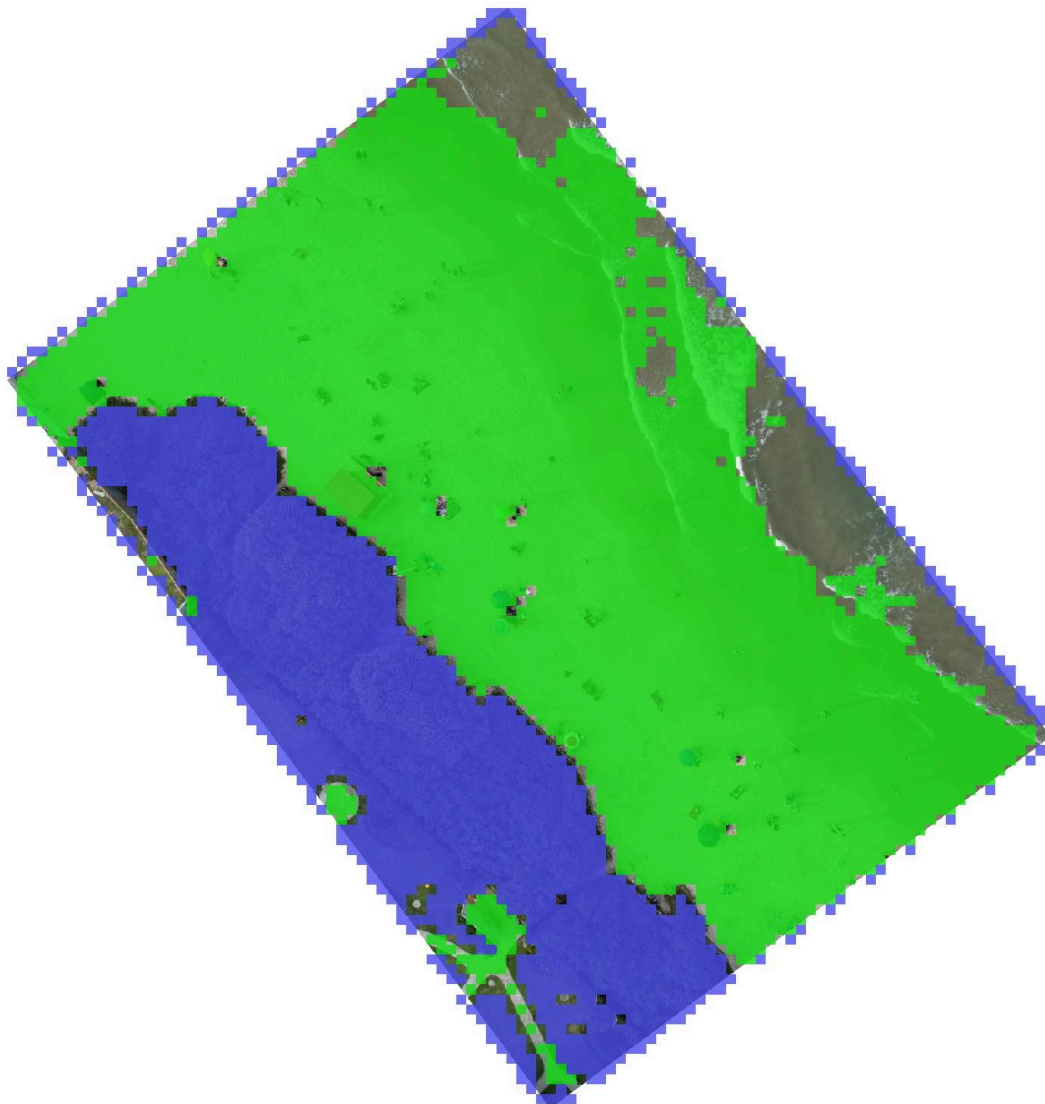


Figura 3. Utilização do modelo em um nova imagem, grade de 1x1 metros, classes areia(verde) e restinga(azul). (fonte autor)

Diferentemente da segmentação por cores, a rede conseguiu abstrair interferências cromáticas (objetos na areia) e classificar corretamente as quadrículas mesmo em contextos visuais complexos. A limitação principal reside na necessidade de um

conjunto de treinamento abrangente. Sem amostras que reflitam todas as condições visuais possíveis, o modelo pode falhar em cenários desconhecidos, foram identificados trechos junto a área de areia que ainda não foram corretamente classificados, foi atribuída a limitação da variedade de amostras, a extensa diversidade de interferências possíveis na faixa de areia não foi bem representada nas poucas amostras que utilizamos. Ainda assim, o método de aprendizado profundo mostrou-se robusto, reduzindo a dependência de faixas de cor fixas e aproximando-se de um entendimento semântico da imagem.

3.3 Utilização de um Modelo YOLO para Detecção de Pessoas

A detecção de pessoas seguiu um caminho distinto. Em vez de segmentar classes de cobertura do solo, buscou-se localizar objetos específicos. Para isso, utilizou-se um modelo do tipo YOLO (You Only Look Once), conhecido por detectar objetos em tempo real. O código dividiu a ortoimagem em blocos (tiles) menores utilizando a biblioteca rasterio, de modo que cada tile pudesse ser processado separadamente. Em seguida, o YOLO foi aplicado sobre cada tile, fornecendo coordenadas dos objetos detectados.



Figura 4. Caixas delimitadoras das pessoas detectadas a partir do modelo YOLO11. (fonte autor)

O modelo YOLO11, pré-treinado para detecção de pessoas, encontrou dificuldades ao lidar com as especificidades das imagens obtidas por drones, notadamente o ângulo superior de visão, possivelmente com uma representatividade menor no conjunto de treinamento original. Além disso, a fotogrametria e o georreferenciamento das ortoimagens podem introduzir distorções (warp) que alteram o formato aparente dos indivíduos, dificultando a identificação. Como consequência, algumas pessoas não foram detectadas. Uma solução consiste em incorporar amostras que reflitam essas

características, seja pela obtenção de novas imagens ou pela simulação de deformações, tornando o YOLO mais adequado às condições reais.

4. Conclusões

O estudo mostrou que as ortoimagens métricas obtidas por drones, após processamento fotogramétrico, servem como uma base sólida para analisar ambientes litorâneos. A segmentação por cores, embora simples e ágil, revelou-se pouco flexível diante de variações cromáticas e da presença de objetos incomuns.

O uso de um modelo pré-treinado em aprendizado profundo (ResNet18) mostrou-se mais consistente. Ao classificar quadrículas de 1x1 metro, o modelo reduziu a dependência de cores fixas e capturou padrões visuais de maior nível. A alta acurácia deve-se à capacidade do modelo em generalizar, embora sua eficácia dependa de conjuntos de treinamento abrangentes. Esse método permitiu abstrair as interferências presentes na faixa de areia, considerando-as parte da classe, resultando em uma classificação mais coerente com o contexto.

Na detecção de pessoas, o YOLO apresentou um bom desempenho, embora as distorções fotogramétricas tenham reduzido sua eficácia. Ainda assim, o resultado permaneceu aceitável para os propósitos deste estudo. Ajustar o modelo às condições reais, incorporando dados mais diversos ao treinamento, pode aumentar sua precisão mesmo em situações visuais adversas.

Como conclusão, a combinação de métodos clássicos e aprendizado profundo oferece um conjunto mais abrangente de ferramentas para monitorar ambientes litorâneos. A existência de ortoimagens métricas e a flexibilidade da visão computacional permitem, de forma integrada, quantificar a extensão de vegetação e areia, bem como detectar pessoas, contribuindo para uma compreensão mais clara dos impactos antrópicos e fornecendo subsídios para decisões mais embasadas na gestão e conservação costeira.

Bibliografia

PYTORCH. Transfer Learning for Computer Vision Tutorial. Disponível em: https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html. Acesso em: 7 dez. 2024.

ULTRALYTICS. Tutoriais e guias do YOLO. Disponível em: <https://docs.ultralytics.com/pt/guides/>. Acesso em: 7 dez. 2024.

ULTRALYTICS. Treinamento de conjuntos de dados personalizados com Ultralytics YOLOv8 no Google Colab. Disponível em: <https://www.ultralytics.com/pt/blog/training-custom-datasets-with-ultralytics-yolov8-in-google-colab>. Acesso em: 7 dez. 2024.