Representação Postural de Baleias-Jubarte em Imagens Aéreas usando Aprendizado Profundo

Reinan L. Argolo[1], Alexandre E. C. Oliveira[1], Lucas S. C. Santos[1], Susana M. Iglesias[2], Paulo E. Ambrósio[2], Bianca M. Righi[3,4], Yvonnick Le Pendu[3]

[1]Departamento de Engenharias e Computação
Universidade Estadual De Santa Cruz (UESC) – Ilhéus, BA – Brazil

[2]Programa de Pós-Graduação em Modelagem Computacional
Universidade Estadual De Santa Cruz (UESC) – Ilhéus, BA – Brazil

[3]Programa de Pós-Graduação em Zoologia
Universidade Estadual De Santa Cruz (UESC) – Ilhéus, BA – Brazil

[4]Projeto Baleia Jubarte
Instituto Baleia Jubarte – Caravelas, BA – Brazil

{rlargolo, aecoliveira, lscsantos}.cic@uesc.br

{smiglesias, peambrosio}@uesc.br

bmrighi.ppgzoo@uesc.br

yvonnick@uesc.br

Abstract. *This paper investigates the postural representation of humpback whales (Megaptera novaeangliae) in aerial images acquired by Unmanned Aerial Vehicles, focusing on deep learning-based anatomical keypoint estimation. Although UAVs have expanded the capacity to monitor cetaceans in natural environments, the increasing volume of visual data and the presence of reflections, foam, partial submersion, and posture variation make manual analysis time-consuming and difficult to scale. To address this problem, we propose a pilot study based on the annotation of a minimal set of 5 anatomical keypoints and the training of a YOLO-based pose estimation model for humpback whale imagery. The model was evaluated on an independent test set using detection metrics, pose estimation metrics, normalized per-keypoint error, and Percentage of Correct Keypoints. The results indicate promising performance, with a pose mAP50 of 0.6539 and a matched-instance PCK of 89.26% at a normalized error threshold of 0.10, although the end-to-end results remains limited by detection and association failures. These findings support the feasibility of keypoint-based postural representation as an initial step toward future analyses of body orientation, movement, and social interactions of humpback whales in real aerial monitoring scenarios.*

Resumo. *Este artigo investiga a representação postural de baleias-jubarte (Megaptera novaeangliae) em imagens aéreas obtidas por Veículos Aéreos Não Tripulados, com foco no uso de aprendizado profundo para estimativa de pontos-chave anatômicos. Embora VANTs ampliem a capacidade de monitoramento de cetáceos em ambiente natural, o grande volume de registros visuais*

e a presença de reflexos, espuma, submersão parcial e variações de postura dificultam a análise manual e a extração sistemática de informações comportamentais. Para enfrentar esse problema, propõe-se um estudo piloto baseado na anotação de 5 pontos anatômicos mínimos e no treinamento de um modelo YOLO para estimativa de pose em imagens de baleias-jubarte. O modelo foi avaliado em um conjunto de teste independente, considerando métricas de detecção, métricas de pose, erro normalizado por ponto e PCK. Os resultados indicam desempenho promissor, com mAP50 de 0,6539 para pose e PCK pareado de 89,26% no limiar de erro normalizado 0,10, embora o resultado end-to-end ainda seja limitado por falhas de detecção e associação. Os achados sustentam a viabilidade da representação postural por pontos-chave como etapa inicial para análises futuras de orientação corporal, deslocamento e interações sociais de jubartes em cenários reais de monitoramento aéreo.

1. Introdução

A conservação de ecossistemas marinhos depende de métodos capazes de produzir informações confiáveis sobre a presença, o deslocamento e o comportamento de espécies em ambiente natural. Entre os mamíferos marinhos, a baleia-jubarte (*Megaptera novaeangliae*) possui especial relevância ecológica e científica, tanto por sua ampla distribuição migratória quanto pelo interesse associado às suas interações sociais, padrões de deslocamento e uso de áreas reprodutivas. A observação sistemática desses animais contribui para compreender aspectos da dinâmica populacional e subsidiar ações de manejo e conservação [Castrillon and Bengtson Nash 2020, Guirado et al. 2019].

Apesar dessa importância, a análise visual de cetáceos em ambiente natural ainda apresenta limitações metodológicas relevantes. As baleias permanecem parte significativa do tempo parcial ou totalmente submersas, e sua observação ocorre em cenas marcadas por variações de luminosidade, turbidez, ondulação, espuma e reflexos na superfície da água. Além disso, mudanças de postura, mergulho e exposição parcial do corpo dificultam a identificação precisa de estruturas anatômicas em imagens aéreas [Castrillon and Bengtson Nash 2020, Boulent et al. 2023]. Como consequência, a interpretação manual de vídeos pode se tornar lenta, subjetiva e pouco escalável, sobretudo diante do aumento no volume de registros audiovisuais.

O uso de Veículos Aéreos Não Tripulados (VANTs) tem ampliado as possibilidades de monitoramento de mamíferos marinhos, permitindo a coleta de imagens em perspectiva aérea com menor interferência direta sobre os animais [Hodgson et al. 2017]. No entanto, a ampliação da capacidade de coleta também intensifica o desafio de processamento dos dados: quanto maior o volume de imagens e vídeos, maior a necessidade de ferramentas capazes de organizar, filtrar e extrair informações de forma consistente [Weinstein 2018, Tuia et al. 2022]. Técnicas de visão computacional e aprendizado profundo têm sido empregadas nesse contexto, principalmente em tarefas de detecção, contagem e rastreamento de indivíduos. Essas abordagens, embora úteis, frequentemente representam o animal por caixas delimitadoras, o que limita a descrição da configuração corporal interna.

Para análises comportamentais mais refinadas, não basta identificar apenas onde o animal está na imagem. Também é relevante compreender como o corpo está orien-

tado, quais estruturas anatômicas estão visíveis e de que modo regiões como cabeça, eixo corporal, pedúnculo caudal e cauda se organizam espacialmente. A estimativa de postura por pontos-chave oferece um caminho metodológico para essa representação, pois associa partes corporais específicas a coordenadas no plano da imagem [Mathis et al. 2018, Pereira et al. 2022]. Com isso, a imagem deixa de ser tratada apenas como uma área de detecção e passa a ser convertida em uma descrição espacial mais informativa do corpo do animal.

Diante desse cenário, este trabalho propõe um estudo piloto para estimativa de postura corporal de baleias-jubarte em imagens aéreas obtidas por VANTs. A proposta consiste na definição de um conjunto inicial de pontos-chave anatômicos, na anotação manual de imagens selecionadas e na avaliação preliminar de um modelo de *pose estimation* para localizar esses pontos em diferentes condições visuais. O objetivo não é apresentar uma solução definitiva para análise comportamental automática, mas investigar a viabilidade de representar a postura corporal de jubartes a partir de imagens aéreas reais.

As principais contribuições deste trabalho são: (i) a definição de um protocolo inicial de anotação de pontos-chave para baleias-jubarte; (ii) a avaliação de um modelo de estimativa de pose em imagens aéreas de um cenário marinho real; e (iii) a análise quantitativa e qualitativa dos fatores que afetam a localização anatômica, como submersão parcial, reflexos, espuma e variações de postura. Os resultados obtidos buscam apoiar trabalhos futuros voltados à análise comportamental, à integração com rastreamento multiobjeto e à extração de métricas espaciais mais refinadas sobre cetáceos em ambiente natural.

2. Trabalhos Relacionados

O uso de imagens aéreas tem se consolidado como uma estratégia relevante para o estudo de mamíferos marinhos, por permitir a observação de indivíduos em ambiente natural com menor interferência direta do que métodos tradicionais. Em baleias-jubarte (*Megaptera novaeangliae*), VANTs vêm sendo empregados principalmente em fotogrametria, estimativa de comprimento corporal e avaliação de condição física. Estudos como os de Napoli et al. [Napoli et al. 2024] e Blázquez et al. [Blázquez et al. 2025] demonstram que imagens aéreas fornecem informações importantes sobre tamanho, forma e estado corporal, embora permaneçam centrados sobretudo em medidas morfométricas, e não na representação postural por pontos anatômicos.

Técnicas de visão computacional e aprendizado profundo também têm sido aplicadas ao monitoramento de cetáceos, especialmente em tarefas de detecção, contagem e rastreamento. Alsaidi et al. [Alsaidi et al. 2024], por exemplo, utilizaram modelos de aprendizado profundo para localizar e rastrear belugas em vídeos aéreos, distinguindo adultos e filhotes. Apesar de aproximarem a análise automática de questões comportamentais, essas abordagens geralmente representam os animais por caixas delimitadoras ou trajetórias, o que limita a descrição da configuração corporal interna.

Entre a simples detecção e a análise postural, alguns trabalhos exploram descritores geométricos do corpo. Bierlich et al. [Bierlich et al. 2024] desenvolveram modelos para extrair medidas de comprimento e condição corporal em vídeos de drone, enquanto Bagchi et al. [Bagchi et al. 2025] utilizaram Mask R-CNN para gerar máscaras corporais

e estimar pontos ao longo do eixo de baleias-francas-austrais. Embora não tratem diretamente de *pose estimation*, esses estudos indicam que a representação geométrica do corpo é uma etapa importante para análises mais detalhadas.

A estimativa de pose por pontos-chave tem avançado em estudos comportamentais com animais, com ferramentas como DeepLabCut [Mathis et al. 2018], DeepPoseKit [Graving et al. 2019] e SLEAP [Pereira et al. 2022], que permitem rastrear partes corporais definidas pelo usuário. No entanto, sua aplicação a baleias-jubarte em imagens aéreas reais ainda é desafiadora devido a corpos alongados e não rígidos, submersão parcial, baixo contraste com a água, reflexos, espuma e variações de visibilidade. Assim, embora a literatura apresente avanços em detecção, rastreamento, fotogrametria e morfometria de cetáceos, ainda há uma lacuna na representação postural de jubartes por pontos-chave anatômicos em cenários reais de monitoramento aéreo.

3. Materiais e Métodos

Este trabalho avalia uma abordagem para a estimativa de pose de baleias-jubarte em imagens aéreas obtidas por Veículos Aéreos Não Tripulados (VANTs). A proposta busca verificar a viabilidade de representar a postura corporal do animal por meio de pontos-chave anatômicos, permitindo uma descrição mais informativa do que a obtida apenas por caixas delimitadoras. A metodologia compreende quatro etapas: construção do conjunto de dados, definição e anotação dos pontos anatômicos, treinamento do modelo e avaliação quantitativa dos resultados.

3.1. Conjunto de dados e anotação

O conjunto de dados é composto por 504 imagens de baleias-jubarte, extraídas de vídeos aéreos capturados na região do Banco dos Abrolhos, uma das principais áreas de ocorrência reprodutiva da espécie no Atlântico Sul Ocidental [Martins et al. 2001]. As imagens contemplam diferentes condições visuais, incluindo variações de iluminação, reflexos na superfície da água, presença de espuma, submersão parcial do corpo e mudanças na orientação dos indivíduos.

Como os quadros foram extraídos de vídeos, a divisão dos dados considerou a origem temporal das imagens. Essa decisão buscou reduzir o risco de vazamento de informação entre treinamento, validação e teste, evitando que frames muito semelhantes de uma mesma sequência fossem distribuídos em subconjuntos distintos. Assim, a avaliação final reflete de modo mais realista a capacidade de generalização do modelo para novas cenas.

Para representar a postura corporal das baleias, foram definidos cinco pontos-chave anatômicos ilustrados na Figura 1, sendo eles: (A) rostro, (B) nadadeira dorsal, (C) pedúnculo caudal, (D) extremidade caudal direita e (E) extremidade caudal esquerda. A seleção desses pontos foi orientada por critérios biológicos e fotogramétricos associados à visibilidade de estruturas corporais em imagens aéreas. Em estudos com VANTs, fotografias adequadas de jubarte são frequentemente caracterizadas pela visualização da face dorsal do animal na superfície, com eixo corporal retilíneo e pedúnculo caudal não arqueado [Christiansen et al. 2016]. Além disso, medidas morfométricas em imagens aéreas utilizam o rostro, a nadadeira dorsal e o pedúnculo caudal como referências para estimar o comprimento, larguras e condições corporais

[Christiansen et al. 2016, Napoli et al. 2024]. As nadadeiras peitorais não foram incluídas, pois, nas posturas dorsais adotadas para análise [Christiansen et al. 2016], tenderam a permanecer junto ao corpo, dificultando a identificação de suas extremidades. Assim, optou-se por um conjunto mínimo de pontos que representasse o eixo corporal aparente mantendo viabilidade de anotação em imagens com visibilidade parcial.

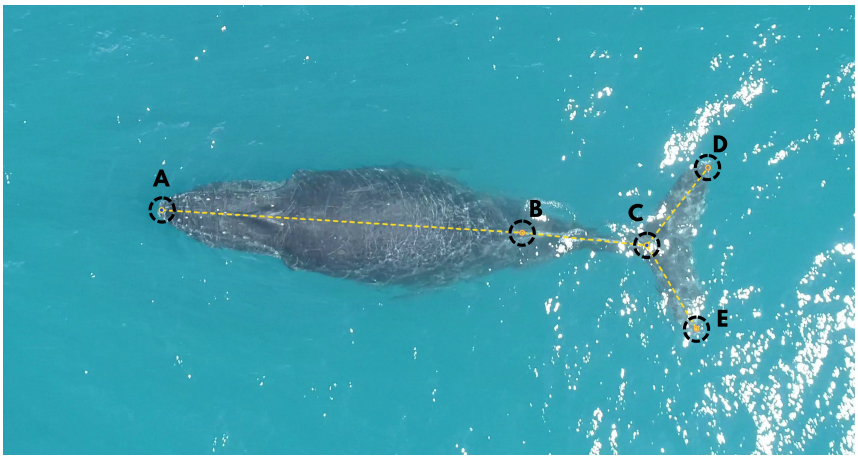


Figura 1. Exemplo de anotação de pose em imagem aérea de baleia-jubarte. As linhas tracejadas representam as conexões utilizadas para visualização da estrutura de pose.

As imagens foram anotadas manualmente, marcando-se os cinco pontos para cada indivíduo visível. Em casos de baixa visibilidade, submersão parcial ou ambiguidade anatômica, a marcação foi realizada com base na melhor estimativa visual da estrutura correspondente. Como se trata de um estudo piloto, a ausência de validação por especialistas e de análise de concordância entre anotadores constitui uma limitação do protocolo inicial, a ser tratada em etapas futuras.

3.2. Treinamento do modelo

Para a estimativa dos pontos-chave, foi utilizado um modelo baseado na arquitetura *YOLO26* para *pose estimation*. O treinamento considerou simultaneamente a localização da baleia por caixa delimitadora e a predição dos pontos anatômicos associados a cada indivíduo. Essa estratégia permite que o modelo aprenda uma representação conjunta da presença do animal e de sua configuração corporal simplificada.

O treinamento foi conduzido com o objetivo de obter um modelo de referência para a tarefa, e não uma solução operacional definitiva. O desempenho foi acompanhado no conjunto de validação por métricas de detecção e de pose. O checkpoint selecionado para a avaliação final corresponde ao arquivo `best.pt`, escolhido antes da aplicação no conjunto de teste.

3.3. Protocolo de avaliação

A avaliação foi realizada exclusivamente no conjunto de teste, composto por 163 imagens e 154 instâncias anotadas de baleias-jubarte. Esse conjunto não foi utilizado no treinamento, na validação, na seleção do checkpoint ou no ajuste de hiperparâmetros. A

inferência foi conduzida com a implementação Ultralytics, utilizando imagens redimensionadas para 640 pixels, limiar mínimo de confiança de 0,001 para a construção das curvas de precisão-revocação e limiar de IoU de 0,7 na etapa de *Non-Maximum Suppression* (NMS).

Foram reportadas métricas separadas para a detecção de caixas delimitadoras, indicadas por (B), e para a estimativa dos pontos-chave, indicadas por (P). Para a detecção, utilizaram-se *Precision*, *Recall*, mAP50 e mAP50-95 [Everingham et al. 2010, Lin et al. 2014]. Para a pose, foram consideradas métricas equivalentes baseadas em *Object Keypoint Similarity* (OKS), conforme protocolos derivados do COCO [Lin et al. 2014, Maji et al. 2022].

Além das métricas globais, foi realizada uma análise por ponto anatômico. As caixas preditas e anotadas foram pareadas pelo algoritmo Húngaro [Kuhn 1955], considerando válidas apenas associações com IoU maior ou igual a 0,5. Para cada ponto-chave, o erro foi calculado como a distância euclidiana entre a coordenada predita e a coordenada anotada, normalizada pela diagonal da caixa delimitadora anotada. Essa normalização permite comparar erros entre instâncias de diferentes tamanhos na imagem.

Como medida complementar, calculou-se o *Percentage of Correct Keypoints* (PCK), que mede a proporção de pontos cujo erro normalizado não ultrapassa um limiar τ [Yang and Ramanan 2011, Andriluka et al. 2014]. Para um conjunto $\mathcal{V}$ de pontos avaliados, o PCK foi definido como:

$$PCK(\tau) = \frac{1}{|\mathcal{V}|} \sum_{(i,k)\in\mathcal{V}} \mathbf{1}(e_{i,k} \leq \tau), \tag{1}$$

em que $e_{i,k}$ representa o erro normalizado do ponto k na instância i, e $\mathbf{1}(\cdot)$ é a função indicadora. O PCK foi calculado em duas condições: uma avaliação *end-to-end*, na qual caixas não associadas e pontos com confiança inferior a 0,25 foram tratados como falhas; e uma avaliação restrita às instâncias corretamente pareadas, voltada a medir a qualidade da localização dos pontos quando a baleia foi detectada com sucesso.

4. Resultados e Discussão

O modelo foi avaliado no conjunto de teste composto por 163 imagens e 154 instâncias anotadas de baleias-jubarte. Esse conjunto não foi utilizado nas etapas de treinamento, validação, seleção do checkpoint ou ajuste de hiperparâmetros. A Tabela 1 apresenta os resultados obtidos para detecção de caixas delimitadoras, indicadas por (B), e para estimativa de pose, indicada por (P).

Tabela 1. Métricas globais do modelo no conjunto de teste.

Tarefa	Precision	Recall	mAP50	mAP50-95
Detecção (B)	0,8263	0,5584	0,6365	0,3716
Pose (P)	0,8557	0,5777	0,6539	0,5678

A precisão foi superior à revocação tanto na detecção quanto na estimativa de pose, indicando que o modelo tende a produzir predições confiáveis, embora ainda deixe

de recuperar parte das instâncias presentes no conjunto de teste. Na tarefa de pose, o modelo alcançou mAP50(P) de 0,6539 e mAP50-95(P) de 0,5678, resultado que indica a viabilidade inicial da representação corporal por pontos-chave em imagens aéreas de jubartes, mesmo diante das variações visuais próprias do ambiente marinho.

A análise por ponto anatômico, apresentada na Tabela 2, evidencia diferenças relevantes entre os marcos avaliados. O menor erro médio normalizado foi observado no pedúnculo caudal, seguido pela cauda esquerda, enquanto os maiores erros ocorreram na nadadeira dorsal e na cauda direita.

Tabela 2. Erro médio por ponto anatômico no conjunto de teste.

Ponto	Pareados	Erro px	Erro norm.
Rostro	90	84,37	0,0621
Nadadeira dorsal	88	94,51	0,0678
Pedúnculo caudal	90	61,20	0,0449
Cauda direita	89	88,99	0,0670
Cauda esquerda	90	66,29	0,0486

Essa variação sugere que pontos associados ao eixo longitudinal do corpo, como o pedúnculo caudal, tendem a ser mais estáveis quando a silhueta da baleia está visível. Em contraste, estruturas laterais ou parcialmente expostas, como a nadadeira dorsal e as extremidades caudais, são mais sensíveis a oclusões, ondulação e alterações de contraste. Para a proposta deste trabalho, esse resultado indica que a estimativa da orientação corporal pode ser mais robusta quando apoiada nos pontos centrais do eixo corporal.

A Figura 2 apresenta a curva de PCK em diferentes limiares de erro normalizado. No limiar 0,10, o PCK *end-to-end* foi de 53,20%, enquanto o PCK calculado apenas sobre instâncias corretamente pareadas alcançou 89,26%.

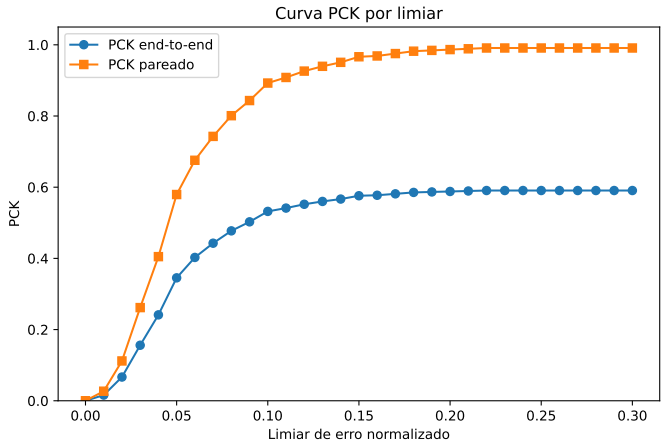


Figura 2. Curva PCK para avaliação *end-to-end* e avaliação restrita às instâncias pareadas.

A redução de 36,06 pontos percentuais entre o PCK pareado e o *end-to-end* mostra que a localização dos pontos é consistente quando a baleia é detectada e associada, mas se degrada no fluxo completo. Como caixas não associadas e pontos com confiança

inferior a 0,25 são contabilizados como falhas no segundo cenário, a diferença decorre principalmente da recuperação incompleta das instâncias, interpretação corroborada pelos valores de *Recall* de 0,5584 na detecção e 0,5777 na pose. O principal gargalo está, portanto, nas etapas anteriores à localização anatômica: detecção, associação e filtragem dos pontos.

Do ponto de vista computacional, o modelo apresenta características compatíveis com o processamento de grandes volumes de imagens aéreas. Com 2.797.824 parâmetros, tamanho de 5,76 MB e tempo médio de inferência de 21,24 ms por imagem, trata-se de uma solução leve, com potencial de uso em etapas futuras de triagem, pré-anotação ou extração inicial de descritores corporais.

A análise qualitativa, ilustrada na Figura 3, reforça os resultados quantitativos. Nos melhores casos, os pontos previstos acompanham adequadamente o eixo corporal da baleia e permitem inferir sua orientação aparente no plano da imagem. Nos casos intermediários, a estrutura global tende a ser preservada, ainda que ocorram deslocamentos locais em pontos específicos. As maiores falhas aparecem em situações de baixa visibilidade ou separação pouco nítida entre o corpo do animal e a superfície da água.

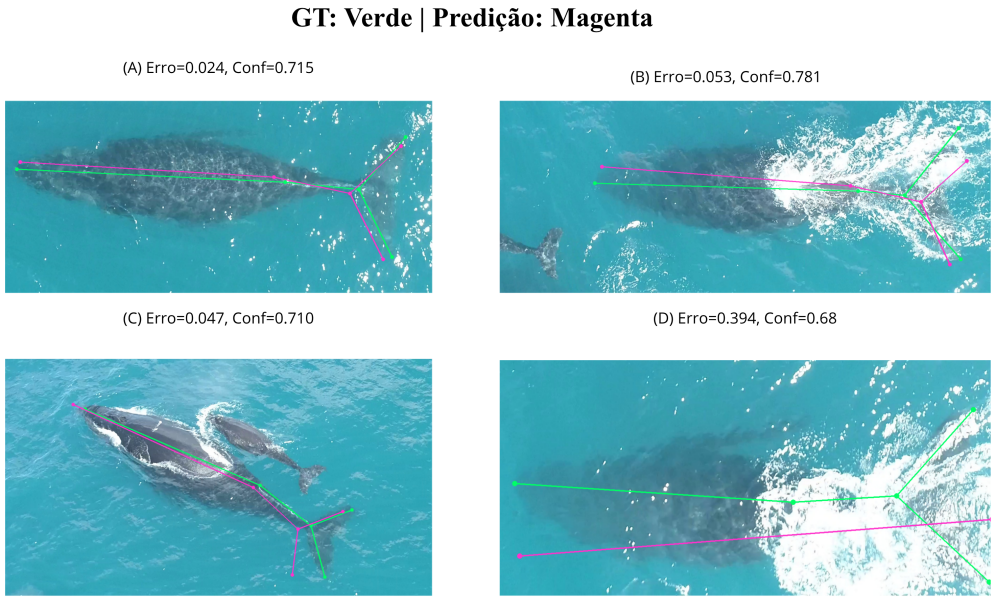


Figura 3. Exemplos qualitativos de predições do modelo. As anotações manuais são indicadas em verde e as predições em magenta.

5. Conclusão e Trabalhos Futuros

Este trabalho avaliou a estimativa de pose de baleias-jubarte em imagens aéreas por um conjunto mínimo de pontos-chave anatômicos. O modelo alcançou mAP50(P) de 0,6539 e PCK pareado de 89,26% no limiar 0,10. O PCK *end-to-end* de 53,20% confirma que o fluxo completo ainda é limitado por falhas de detecção e associação.

A principal contribuição do estudo está em demonstrar a viabilidade de representar a postura corporal aparente de jubartes por meio de pontos anatômicos em imagens aéreas reais. Essa representação oferece uma descrição mais informativa do que caixas

delimitadoras, pois permite observar a organização espacial do corpo e fornece uma base inicial para análises de orientação, deslocamento e interação entre indivíduos.

Como trabalhos futuros, pretende-se ampliar o conjunto de dados para abranger diferentes níveis de turbidez e profundidade de submersão, além de refinar o protocolo com múltiplos anotadores e validação por especialistas. Para reduzir as falhas do fluxo completo, serão avaliados limiares de confiança combinados com rastreamento temporal, por métodos como ByteTrack ou BoT-SORT. Essa integração permitirá explorar métricas temporais e relacionais associadas ao comportamento de baleias-jubarte.

Referências

Alsaidi, M., Al-Jassani, M. G., Bang, C., O'Corry-Crowe, G., Watt, C., Ghazal, M., and Zhuang, H. (2024). Localization and tracking of beluga whales in aerial video using deep learning. *Frontiers in Marine Science*, 11:1445698.

Andriluka, M., Pishchulin, L., Gehler, P., and Schiele, B. (2014). 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3686–3693, Columbus, OH, USA. IEEE.

Bagchi, C., Medina, J., Irschick, D. J., Maji, S., and Christiansen, F. (2025). Automated extraction of right whale morphometric data from drone aerial photographs. *Remote Sensing in Ecology and Conservation*, 11(4):438–453.

Bierlich, K. C., Karki, S., Bird, C. N., Fern, A., and Torres, L. G. (2024). Automated body length and body condition measurements of whales from drone videos for rapid assessment of population health. *Marine Mammal Science*, 40(4):e13137.

Blázquez, M., Massett, N., O'Callaghan, S., Hogan, M., López-Suárez, P., Lopes, K., Whooley, P., O'Brien, J. M., Wenzel, F. W., O'Connor, I., and Berrow, S. D. (2025). Humpback whale (Megaptera novaeangliae) maturity and body condition in Ireland, Norway and Cabo Verde based on UAS photogrammetry data. *Frontiers in Marine Science*, 12:1604106.

Boulent, J., Charry, B., Kennedy, M. M., Tissier, E., Fan, R., Marcoux, M., Watt, C. A., and Gagné-Turcotte, A. (2023). Scaling whale monitoring using deep learning: A human-in-the-loop solution for analyzing aerial datasets. *Frontiers in Marine Science*, 10:1099479.

Castrillon, J. and Bengtson Nash, S. (2020). Evaluating cetacean body condition; a review of traditional approaches and new developments. *Ecology and Evolution*, 10(12):6144–6162.

Christiansen, F., Dujon, A. M., Sprogis, K. R., Arnould, J. P. Y., and Bejder, L. (2016). Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales. *Ecosphere*, 7(10):e01468.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338.

Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R., and Couzin, I. D. (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife*, 8:e47994.

Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., and Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, 9(1):14259.

Hodgson, A., Peel, D., and Kelly, N. (2017). Unmanned aerial vehicles for surveying marine fauna: assessing detection probability. *Ecological Applications*, 27(4):1253–1267.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2014). Microsoft COCO: Common Objects in Context.

Maji, D., Nagori, S., Mathew, M., and Poddar, D. (2022). YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss.

Martins, C., Morete, M., H.Engel, M., Freitas, A., Secchi, E., and Kinas, P. (2001). Aspects of habitat use patterns of humpback whales in the Abrolhos Bank, Brazil, breeding ground. *Memoirs of the Queensland Museum*, 47:563–570.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289.

Napoli, C., Hirtle, N., Stepanuk, J., Christiansen, F., Heywood, E. I., Grove, T. J., Stoller, A., Dodds, F., Glarou, M., Rasmussen, M. H., Lonati, G. L., Davies, K. T., Videsen, S., Simon, M. J., Boye, T. K., Zoidis, A., Todd, S. K., and Thorne, L. H. (2024). Drone-based photogrammetry reveals differences in humpback whale body condition and mass across North Atlantic foraging grounds. *Frontiers in Marine Science*, 11:1336455.

Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., Papadoyannis, E. S., Normand, E., Deutsch, D. S., Wang, Z. Y., McKenzie-Smith, G. C., Mitelut, C. C., Castro, M. D., D'Uva, J., Kislin, M., Sanes, D. H., Kocher, S. D., Wang, S. S.-H., Falkner, A. L., Shaevitz, J. W., and Murthy, M. (2022). SLEAP: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19(4):486–495.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., Van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., Van Horn, G., Crofoot, M. C., Stewart, C. V., and Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1):792.

Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3):533–545.

Yang, Y. and Ramanan, D. (2011). Articulated pose estimation with flexible mixtures-of-parts. In *CVPR 2011*, pages 1385–1392, Colorado Springs, CO, USA. IEEE.