

Detecção de Lixo em Áreas Costeiras: Uma Aplicação de Segmentação com R-CNNs da Família YOLO

Melinne Diniz de Oliveira, Elloá B. Guedes

¹Laboratório de Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas
{mddo.snf21, ebgcosta}@uea.edu.br

Abstract. *This work aims to train and evaluate Deep Learning models from the YOLO Family to segment instances of artificial waste in images of coastal regions, aiming to contribute to the monitoring and depollution of these environments. A data preparation pipeline was developed and then YOLOv7 and YOLOv8 models were tested, with YOLOv7 standing out for its best performance, possibly due to its lesser specialization for benchmarks. The results aim to assist in the creation of automatic systems to detect and remove plastic waste in coastal areas, in order to reduce the environmental impacts of this form of pollution.*

Resumo. *Este trabalho visa treinar e avaliar modelos de Deep Learning da Família YOLO para segmentar instâncias de lixo artificial em imagens de regiões costeiras, visando contribuir para o monitoramento e despoluição desses ambientes. Foi desenvolvido um pipeline de preparação de dados e, em seguida, foram testados os modelos YOLOv7 e YOLOv8, destacando-se a YOLOv7 pelo melhor desempenho, possivelmente devido à sua menor especialização para benchmarks. Os resultados obtidos visam auxiliar na criação de sistemas automáticos para detectar e remover resíduos plásticos nas áreas costeiras, com o intuito de reduzir os impactos ambientais dessa forma de poluição.*

1. Introdução

Os Objetivos de Desenvolvimento Sustentável das Organizações das Nações Unidas (ONU), adotados em 2015, são um chamado à ação para pôr fim à pobreza, proteger o planeta e garantir que até 2030 todas as pessoas desfrutem de paz e prosperidade [ONU 2015]. O Décimo Quarto Objetivo – Vida na Água – trata da conservação e uso sustentável dos oceanos, mares e recursos marinhos, pois estes (i) contém 97 % da água da Terra; (ii) contemplam o maior ecossistema do mundo, abrigando quase um milhão de espécies conhecidas e contendo um vasto potencial inexplorado para descobertas científicas; (iii) fornecem recursos naturais essenciais, incluindo alimentos, medicamentos, biocombustíveis, etc.; (iv) absorvem cerca de 23 % das emissões anuais de CO₂ geradas pela atividade humana, contribuindo para atenuar os impactos das mudanças climáticas; e (v) é fonte de emprego e renda para diversas pessoas; dentre outros [ONU 2023].

Nos últimos 70 anos, a produção de plástico cresceu exponencialmente, passando de dois milhões de toneladas em 1950 para 450 milhões de toneladas em 2019, um aumento de 230 vezes. Embora seja um material versátil e amplamente utilizado devido à sua acessibilidade e diversas aplicações, sua gestão inadequada resulta em sérios danos ambientais. Anualmente, entre 5 e 12 Mi t de resíduos plásticos são depositados nos oceanos, com 88 %

permanecendo próximos à costa, causando impactos financeiros significativos, estimados em cerca de US \$13 Bi em custos de limpeza e perdas nas indústrias pesqueiras e em outros setores. Além disso, uma investigação recente revelou que no Brasil, assim como em outras regiões do mundo, o plástico se destacou como o resíduo mais abundantemente encontrado em um conjunto de 44 praias analisadas. [Andrades et al. 2020]

Por serem combustíveis fósseis, desenvolver soluções para mitigar o problema dos resíduos plásticos nas áreas costeiras é importante, pois, dentre outros, colabora para mitigar as consequências ecológicas, uma vez que o plástico representa uma enorme ameaça à vida marinha; para preservar a biodiversidade, visto que os oceanos são lar de diversas espécies; para diminuir os impactos na saúde humana, especialmente em decorrência da absorção dos microplásticos; para favorecer atividades econômicas, como pesca e turismo, permitindo a subsistência de mais de três bilhões de pessoas; e também diminuindo a acidificação dos oceanos [Simul Bhuyan et al. 2021, Romera-Castillo et al. 2023]. Em síntese, ao abordar a poluição por plásticos nos mares e áreas costeiras, garante-se não apenas a vitalidade dos oceanos, mas também a promoção de um futuro sustentável tanto para o meio ambiente quanto para a humanidade [ONU 2024].

Em face do contexto apresentado, o objetivo deste trabalho é apresentar uma solução inteligente de Visão Computacional (VC) para segmentação de instâncias de lixo artificial em regiões costeiras, colaborando também para o posterior desenvolvimento de soluções integradas e automatizadas de coleta e destinação de tais resíduos, diminuindo o impacto ambiental. Essa solução foi desenvolvida a partir de uma base de dados realística contendo 3500 imagens, contempla etapas de pré-processamento e parametrização e treinamento de Redes Neurais Convolucionais Baseadas em Regiões (R-CNNs, do inglês *Region-Based Convolutional Neural Networks* da Família YOLO (do inglês, *You Only Look Once*).

Para apresentar o que se propõe, este artigo está organizado como segue. Os trabalhos relacionados são discutidos na Seção 2. Os materiais e métodos são apresentados na Seção 3. Os resultados e discussão encontram-se dispostos na Seção 4. Por fim, as considerações finais e trabalhos futuros são mostrados na Seção 5.

2. Trabalhos Relacionados

A preocupação global com a poluição plástica em áreas costeiras é inegável. Embora as tecnologias de monitoramento remoto apresentem um potencial significativo para proporcionar vigilância de longo prazo em escala global, os estudos referentes aos resíduos plásticos em ambientes marinhos ainda se encontram em fase de desenvolvimento. Essa complexidade é atribuída, em parte, às restrições nas resoluções espaciais e espectrais, bem como às divergências nas propriedades de reflectância de superfície entre diferentes polímeros [Veettil et al. 2022].

O monitoramento de resíduos plásticos em diversas localidades costeiras tem sido conduzido com Veículos Aéreos Não Tripulados (VANTs) devido à alta frequência de aquisição de imagens, alta resolução espacial, capacidade de voar em baixas altitudes abaixo das nuvens e elevada mobilidade. Contudo, as técnicas de processamento de imagens ainda carecem de uma padronização uniforme nesse contexto, impactando o desempenho da detecção e resultando em classificações incorretas inevitáveis devido à sobreposição parcial dos espectros de objetos com cores distintas [Bao et al. 2018]. Em um dos cenários no qual as imagens coletadas por VANTs foram classificadas utilizando *Machine Learning*, houve detecção de 44 % de garrafas plásticas, de 5 % de tampas de garrafa e de 3,7 % de sacolas, ressaltando as dificulda-

des de detecção inversamente proporcionais ao tamanho do objeto considerado [Martin et al. 2018]

No tocante à captura de imagens em áreas costeiras para detecção de resíduos plásticos, o trabalho de Tamin *et al.* [2023] propõe um sistema que considera o posicionamento de múltiplas câmeras capazes de capturar tanto imagens convencionais (luz visível) quanto no espectro infravermelho próximo, sob o argumento das diferentes propriedades dos resíduos plásticos em diferentes contextos ambientais. Após a captura das imagens, os autores propuseram duas bases de dados, cada uma contendo um dos espaços de cores considerados. Em seguida, utilizaram a R-CNN YOLOv5m em uma validação cruzada 10-folds. No contexto das imagens convencionais, o mAP@0.5 médio foi de 93,31 %, enquanto nas imagens infravermelho próximo foi de 92,24 %. Embora as métricas de desempenho de detecção tenham se mostrado altas, é interessante ressaltar que ambos os *datasets* possuíam poucos exemplos (284 imagens cada), as anotações de lixo possuíam formato retangular e houve uma melhoria do treino com o acréscimo de imagens de *background*, utilizadas com o propósito de fornecer experiência ao modelo sobre as características típicas da região costeira sem resíduos plásticos. Além disso, há que se mencionar que o uso de imagens infravermelho próximas em tarefas de detecção de objetos requer uma cuidadosa consideração de suas possíveis implicações, incluindo condições atmosféricas, reflectância de superfície e a presença de outros materiais que podem afetar a qualidade e confiabilidade da imagem. Tais informações nem sempre são relevantes ou necessárias para certas tarefas de detecção de objetos, e sua inclusão no conjunto de dados de treinamento pode introduzir complexidade e vieses desnecessários [Tamin et al. 2023].

O trabalho de Hidaka *et al.* [2022] utilizou a rede HRNet (do inglês, *High-Resolution Network*), um modelo do estado da arte de *Deep Learning* para segmentação semântica cuja principal característica é a capacidade de obter estimativas altamente precisas, mesmo para casos em que os objetos em uma imagem sejam extremamente grandes ou pequenos [Hidaka et al. 2022]. Os autores avaliaram o desempenho do modelo na partição de testes da base de dados BeachLitter [Sugiyama et al. 2022] conforme as métricas de precisão, revocação e IoU (do inglês, *Intersection over Union*) para as oito classes disponíveis. Ao analisar os resultados obtidos, teve-se que a mediana do IoU para a classe indicativa de lixo artificial foi de 38 %, a revocação de 52,2 % e a precisão de 71,6 %. Apesar dos resultados se mostrarem inovadores para o *dataset* em questão, os autores ressaltam que o modelo teve melhor desempenho em outras classes que não se relacionam com o problema da poluição, tais como céu (mediana do IoU igual a 92,4 %) ou mar (97,8 %). Ademais, mencionam também que uma das questões práticas que influenciou na segmentação foi o tamanho dos objetos, os quais não foram detectados se os rótulos fossem menores que 100 px.

Ao examinar a literatura sobre o tema, é possível perceber que a detecção de poluição plástica em áreas costeiras ainda é um desafio. Os métodos de monitoramento remoto apresentam baixa viabilidade, e os VANTs têm dificuldade em detectar objetos pequenos. As soluções baseadas em *Deep Learning* também apresentam limitações. Uma delas utiliza a R-CNN YOLO, mas necessita da prévia aquisição de instrumentos específicos de captura de imagens em outras frequências, o que pode ser um impeditivo financeiro para uma implementação mais abrangente. Um dos trabalhos mais recentes propõe uma solução viável baseada na segmentação semântica. Essa solução parece ser mais factível, mas ainda há potencial para melhorias nas métricas relacionadas à classe de lixo de origem artificial e na identificação de objetos pequenos.

3. Material e Métodos

A tarefa de detecção de lixo artificial em áreas costeiras a partir de imagens foi abordada neste trabalho conforme uma tarefa de segmentação de instâncias em VC, ou seja, cujo objetivo era prover anotações sob a forma de polígonos convexos que circunscrevem o lixo nas imagens, colaborando assim para soluções que quantifiquem a sua ocorrência e que fomentem a sua coleta, minimizando os impactos ambientais.

Para elaborar a solução proposta considerou-se o *BeachLitter dataset*, o qual contém 3500 imagens, coletadas pela Prefeitura de Yamagata, no Japão, entre 2011 e 2019, já previamente particionadas em exemplos de treino e teste (80 % e 20 %, respectivamente) [Sugiyama et al. 2022]. A rotulação dos exemplos, para fins de Aprendizado Supervisionado, foi feita para a tarefa de segmentação semântica, em que cada píxel da imagem recebeu um rótulo dentre oito classes consideradas. Segundo os autores, a anotação desse conjunto de dados foi um processo custoso, envolvendo a colaboração de 15 profissionais ao longo de dois meses, além da atuação de três controladores de qualidade e de um pesquisador encarregados de avaliar as anotações produzidas.

Para contornar as dificuldades de rotulação píxel a píxel, bem como os ônus financeiro e de tempo, a segmentação de instâncias se apresenta como uma alternativa viável para esse contexto. Porém, para tanto, foi necessário adaptar os rótulos inicialmente disponibilizados, como descrito na Seção 3.1. Os modelos considerados e sua respectiva parametrização são apresentados na Seção 3.2. As métricas de desempenho consideradas e o ambiente de desenvolvimento encontram-se dispostos nas Seções 3.3 e 3.4, respectivamente.

3.1. Dados Experimentais: Pré-processamento e Análise Exploratória

A preparação dos exemplos para a tarefa considerada seguiu o *pipeline* apresentado na Figura 1. A descrição detalhada dos passos é mostrada a seguir.



Figura 1: Sequência de passos para preparação dos exemplos para a tarefa de segmentação de instâncias.

No primeiro passo, *Obtenção das Anotações da Classe de Interesse*, dentre as oito classes disponibilizadas pelos autores, apenas a classe rotulada como “*artificial litter*” (lixo artificial) foi considerada relevante para o escopo do presente trabalho, tendo as demais sido descartadas (tais como céu, areia, etc.). Para cada exemplo, vide Figura 2a, identificou-se a máscara de segmentação da classe, denotada na cor vermelha na Figura 2b, e foram selecionados os píxels do contorno externo de cada instância de objeto;

Ao observar os resultados obtidos no passo anterior, verificou-se que a classe de interesse possui um grande número de objetos pequenos e delimitados por polígonos de formato complexo, conforme Figura 3a, características que constituem um desafio para algoritmos de segmentação em geral. Com o intuito de simplificar os polígonos delimitadores de lixo, adotou-se o algoritmo *Convex Hull*, como ilustrado na Figura 3b, oriundo da Geometria Computacional, que recebe como entrada um conjunto de pontos e retorna o menor envoltório (fecho) convexo [Cormen et al. 2009].

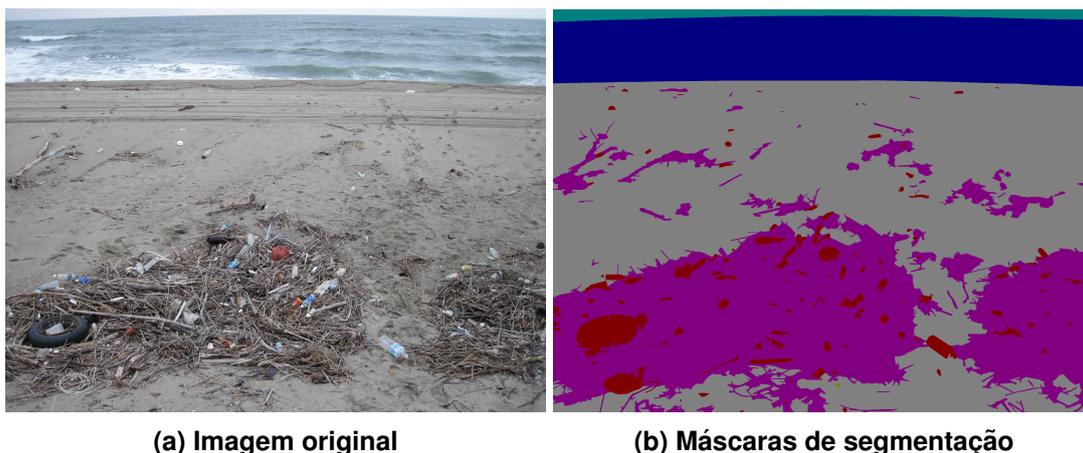


Figura 2: Exemplo da máscara de segmentação disponível na base de dados.

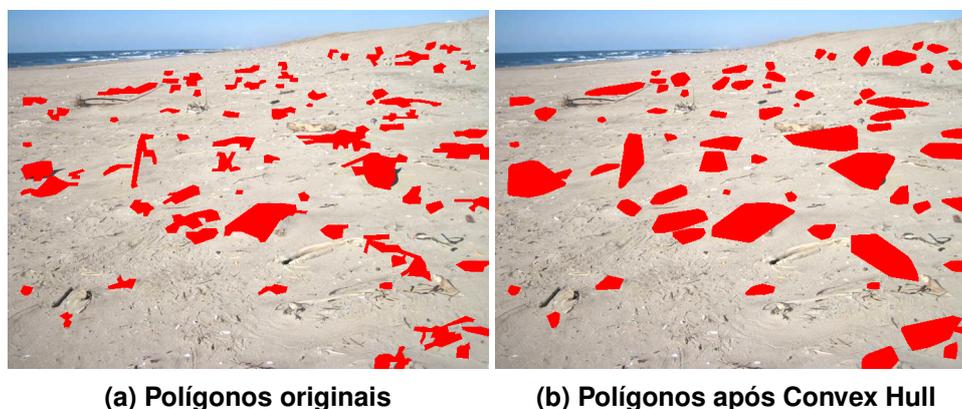


Figura 3: Comparação dos polígonos antes e depois da aplicação do *Convex Hull* em um exemplo da base de dados.

Na análise dos polígonos delimitadores obtidos na etapa anterior, foram identificadas sobreposições que violam o princípio de disjunção necessário para a segmentação, conforme a Figura 4a. Para solucionar esse problema, os polígonos sobrepostos foram mesclados em um único polígono delimitador, conforme demonstrado na Figura 4b. Esse processo foi repetido até que todos os exemplos atendessem aos critérios estabelecidos. Essa técnica resultou na simplificação dos polígonos e na redução do número de amostras necessárias para o treinamento do modelo de segmentação.

Por fim, para facilitar o aprendizado de um modelo de segmentação de instâncias, é recomendado que as coordenadas dos polígonos estejam normalizadas. Para tanto, dividiu-se cada coordenada pelos valores correspondentes à altura e largura da imagem original, resultando em valores no intervalo $[0, 1]$. Dessa forma, as coordenadas normalizadas representam a posição relativa do objeto na imagem, e não os valores absolutos em pixels.

Uma vez concluído o *pipeline* proposto, tende-se a Tabela 1, onde é possível quantificar a área dos objetos contidos nos conjuntos de treino e de teste.

O *pipeline* proposto restringe o problema de segmentação de instâncias para uma classe específica de elevado impacto ambiental. Além disso, simplifica o processo de rotulação de dados para novos contextos, diminuindo os esforços de anotação na ocasião de elaboração de novas bases de dados para este problema de abrangência global.

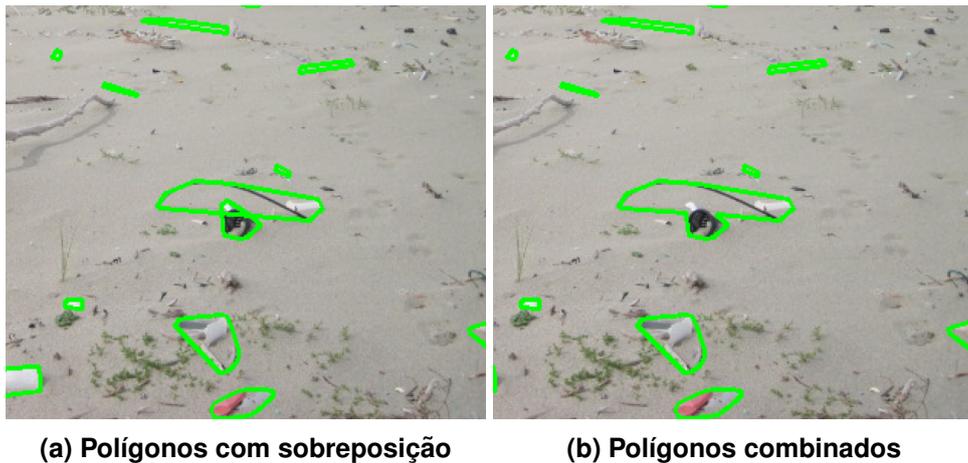


Figura 4: Comparação das máscaras (polígonos) de segmentação com e sem sobreposição

Tabela 1: Área (em px^2) dos polígonos nas partições de treino e teste.

Treinamento				Teste			
Média	Desvio Padrão	Máximo	Mínimo	Média	Desvio Padrão	Máximo	Mínimo
78.5	4749.69	298508.0	0.5	62	4665.69	237729.5	1.0

3.2. Modelos e Parametrização

Nos últimos anos, as arquiteturas das R-CNNs da Família YOLO têm se destacado entre os diversos algoritmos de detecção de objetos devido ao equilíbrio que promovem entre velocidade e precisão. Estas arquiteturas viabilizam a identificação rápida e confiável de objetos em imagens e vídeos, e fornecem modelos leves e otimizados para utilização em dispositivos embarcados [Terven et al. 2023]. Ademais, em suas versões mais recentes, YOLOv7 [Wang et al. 2022] e YOLOv8 [Jocher et al. 2023], foram introduzidos modelos para a tarefa de segmentação de instância de objetos em imagens. Assim, por serem considerados rápidos e acurados, de baixo custo computacional e constituírem atualmente o estado da arte para este tipo de problema, com aplicações em diversos contextos [Michelucci 2019, Diwan et al. 2023], esses modelos foram escolhidos para abordar o problema de Aprendizado Supervisionado em questão. As versões dos modelos YOLO considerados no escopo deste trabalho encontram-se dispostas na Tabela 2.

Tabela 2: Modelos YOLO considerados.

Modelo	Versão	Parâmetros (Mi)
YOLOv8	Nano	3.4
YOLOv8	Medium	27.3
YOLOv7	Padrão	37.8

No contexto da segmentação, os modelos YOLO adotam o *backbone Cross Stage Partial Network* (CSPDarknet53) [Wang et al. 2020] como espinha dorsal para extração de características. Além disso, incluem não apenas cabeçalhos de detecção, mas também dois cabeçalhos de segmentação, os quais aprendem a prever padrões das máscaras de segmentação de instâncias correspondentes às imagens de entrada.

No âmbito experimental, cada modelo foi treinado utilizando o otimizador SGD com uma taxa de aprendizado de 10^{-2} , com *batch size* igual a 8, por até 2000 épocas, com uma paciência de 200 épocas para fins de *Early Stopping* e regularização do treino.

3.3. Avaliação de Desempenho

Para avaliação de desempenho, foi utilizada a técnica de validação cruzada *holdout*, com 70 % dos dados destinados ao treinamento, 10 % para validação e 20 % para testes, partição essa previamente definida pelos autores do *dataset*. As métricas de desempenho foram aferidas na partição de testes e contemplam:

- **Precisão.** Proporção de pixels da classe de lixo artificial (máscaras) corretamente identificados pelo modelo em relação ao total de pixels classificados;
- **Revocação.** Proporção de pixels de lixo artificial identificados pelo modelo em relação ao total presente na imagem;
- **mAP@0.5.** Considera a sobreposição (IoU – do inglês, *Intersection over Union*) entre as segmentações preditas e as máscaras de segmentação reais, fornecendo uma avaliação mais abrangente do desempenho do modelo ao considerar a forma e o tamanho das instâncias de lixo artificial [Padilla et al. 2020].

3.4. Ambiente de Desenvolvimento

A linguagem de programação Python, juntamente com os *frameworks* Tensorflow e Pytorch, foram as ferramentas primárias para treino e avaliação dos modelos. As implementações foram executadas em um servidor computacional equipado com processador Intel® Core™ i9-10900K com clock de 3,7 GHz, memória principal de 64 GB, 7 TB de memória secundária e 3 placas gráficas NVIDIA GTX 3060 com 12 GB VRAM cada, para promover a aceleração em hardware do treino dos modelos.

4. Resultados e Discussão

Os experimentos computacionais foram executados respeitando os passos da metodologia proposta e a síntese dos resultados obtidos encontra-se na Tabela 3.

Tabela 3: Resultados dos experimentos.

Modelo	Precisão	Revocação	mAP@0.5	Épocas	Tempo
YOLOv8 Nano	0.37	0.29	0.26	403	4 h 57 min
YOLOv8 Medium	0.37	0.31	0.27	251	8 h 13 min
YOLOv7	0.44	0.45	0.32	255	11 h 33 min

É possível observar que todos os modelos tiveram o treinamento interrompido de maneira precoce em razão da regularização por *Early Stopping*, um indicativo de que não houve aumento de generalização após um determinado número de épocas de treinamento. Isto pode ter decorrido em razão do baixo número de exemplos disponíveis para treinamento frente à complexidade de formato das máscaras a serem aprendidas pelos modelos.

Ao observar a métrica de referência mAP@0.5, é possível verificar uma baixa variação nos valores obtidos, apesar da grande diferença no número de parâmetros ajustáveis dos modelos, conforme disposto previamente na Tabela 2. Embora o tempo de treinamento tenha sido diferente em razão dos parâmetros de cada arquitetura, observou-se um limiar superior de

aprendizado pelos modelos YOLOv8 e YOLOv7 para esta tarefa de segmentação de instâncias. Essa dificuldade reside no grande número de máscaras por imagem, no formato arbitrário das mesmas e nas suas dimensões, consideradas pequenas, fatores estes já reportados na literatura como desafios para algoritmos de detecção de objetos [Lin et al. 2014].

O maior valor de mAP@0.5 observado deu-se na YOLOv7, ainda que este modelo não seja o mais recente nessa família de R-CNNs. As melhorias propostas no versionamento das redes YOLO não leva em conta tarefas de segmentação, mas sim um *benchmark* de detecção de objetos [Jocher et al. 2023], o que pode justificar um eventual maior viés neste tipo de tarefa, em desfavor da variância em tarefas distintas, como é o caso da segmentação. Exemplos de detecção de instâncias na tarefa considerada podem ser visualizados na Figura 5.

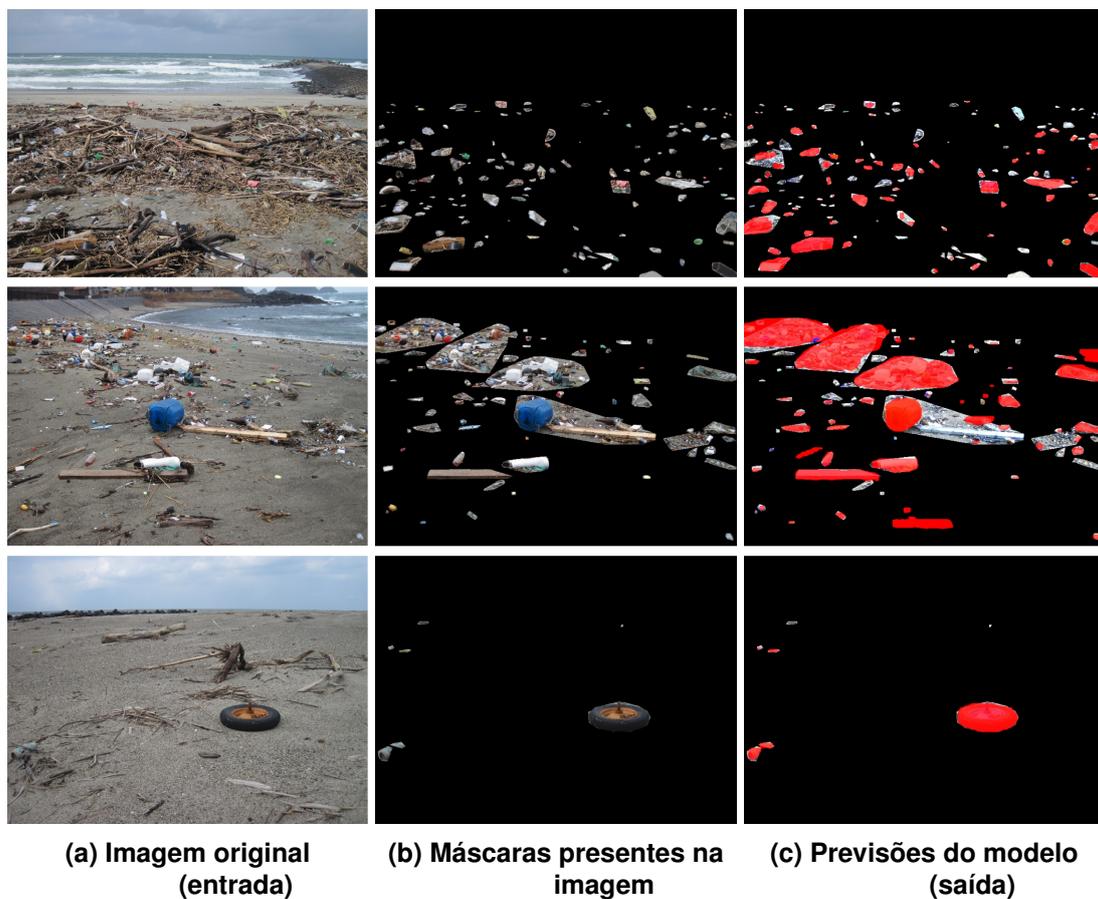


Figura 5: Exemplos de predição dos modelos YOLO.

No quesito de desempenho na tarefa, é importante destacar que os resultados aqui obtidos não podem ser diretamente comparados ao trabalho de Hidaka *et al.* [2022], cujos resultados foram previamente mencionados na Seção 2 e o qual utiliza segmentação semântica. Há uma diferença fundamental entre segmentação semântica e segmentação de instâncias na VC. Embora o trabalho dos referidos autores tenha alcançado métricas superiores, a segmentação de instâncias ainda é uma tarefa complexa, pois os modelos precisam identificar cada objeto individualmente, em vez de categorizar a imagem como um todo. Apesar da viabilidade de aprimorar o desempenho em ambas as tarefas por meio de um conjunto de dados expandido, tal incremento exige a rotulagem manual das imagens, o que representa um desafio adicional.

5. Considerações Finais

Com vistas a colaborar com um dos Objetivos de Desenvolvimento Sustentável da ONU, o presente trabalho considerou a tarefa de segmentar lixo artificial em regiões costeiras a partir de imagens, com vistas a fomentar soluções de monitoramento e despoluição. Para tanto, foi utilizada uma base de dados previamente disponível na literatura para a qual foi proposto um *pipeline* de preparação de dados com enfoque na representação dos rótulos para um domínio de segmentação de instâncias, até então não explorado no contexto desta tarefa. Modelos de R-CNNs da Família YOLO versões 7 e 8 foram considerados, dentre os quais o melhor desempenho foi observado na versão mais antiga, em decorrência da sua menor especialização para *benchmarks* considerados pelos seus proponentes.

Em trabalhos futuros, almeja-se expandir a avaliação de modelos de *Deep Learning* na tarefa em questão, contemplando soluções como Mask R-CNN, YOLOACT e U-Net. Uma outra perspectiva consiste também em aprimorar o *pipeline* de preparação de dados, experimentando outras técnicas de pré-processamento e aumento de dados com vistas a prover melhor generalização na preparação dos modelos. Por fim, há que se aferir a solução de segmentação proposta em tempo real, favorecendo sua adoção em contextos práticos, contribuindo assim para a preservação dos ecossistemas marinhos e a mitigação dos impactos ambientais.

Referências

- Andrades, R., Pegado, T., Godoy, B. S., Reis-Filho, J. A., Nunes, J. L. S., Grillo, A. C., Machado, R. C., Santos, R. G., Dalcin, R. H., Freitas, M. O., Kuhnen, V. V., Barbosa, N. D., Adelar-Alves, J., Albuquerque, T., Bentes, B., and Giarrizzo, T. (2020). Anthropogenic litter on brazilian beaches: Baseline, trends and recommendations for future approaches. *Mar. Pollut. Bull.*, 151(110842):110842.
- Bao, Z., Sha, J., Li, X., Hanchiso, T., and Shifaw, E. (2018). Monitoring of beach litter by automatic interpretation of unmanned aerial vehicle images using the segmentation threshold method. *Marine Pollution Bulletin*, 137:388–398.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms*. The MIT Press, 3 edition.
- Diwan, T., Anirudh, G., and Tembhurne, J. V. (2023). Object Detection Using YOLO: Challenges, Architectural Successors, Datasets and Applications. *Multimed. Tools Appl.*, 82(6):9243–9275.
- Hidaka, M., Matsuoka, D., Sugiyama, D., Murakami, K., and Kako, S. (2022). Pixel-level image classification for detecting beach litter using a deep learning approach. *Mar. Pollut. Bull.*, 175(113371):113371.
- Jocher, G., Chaurasia, A., and Qiu, J. (2023). YOLOv8 – real-time object detection. Disponível em <https://docs.ultralytics.com/models/yolov8/>. Acesso em 16 de maio de 2024.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *Proceedings of the 13th European Conference on Computer Vision (ECVV 2014)*, pages 740–755, Suíça. Springer.
- Martin, C., Parkes, S., Zhang, Q., Zhang, X., McCabe, M. F., and Duarte, C. M. (2018). Use of unmanned aerial vehicles for efficient beach litter monitoring. *Marine Pollution Bulletin*, 131:662–673.

- Michelucci, U. (2019). *Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection*. Apress, Suíça.
- ONU (2015). Objetivos de desenvolvimento sustentável – sdgs transform our world. Assembleia Geral da ONU.
- ONU (2023). Oceans. Disponível em <https://www.un.org/sustainabledevelopment/oceans/>. Acesso em 16 de maio de 2024.
- ONU (2024). Plastic Pollution and Marine Litter. Disponível em <https://www.unep.org/topics/ocean-seas-and-coasts/ecosystem-degradation-pollution/plastic-pollution-and-marine-litter-0>. Acesso em 16 de maio de 2024.
- Padilla, R., Netto, S. L., and da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 237–242, Niterói, Brasil.
- Romera-Castillo, C., Lucas, A., Mallenco-Fornies, R., Briones-Rizo, M., Calvo, E., and Pelejero, C. (2023). Abiotic plastic leaching contributes to ocean acidification. *Science of The Total Environment*, 854:158683.
- Simul Bhuyan, M., Venkatramanan, Selvam, Szabo, S., Maruf Hossain, M., Rashed-Un-Nabi, M., Paramasivam, Jonathan, and Shafiqul Islam, M. (2021). Plastics in marine ecosystem: A review of their sources and pollution conduits. *Reg. Stud. Mar. Sci.*, 41(101539):101539.
- Sugiyama, D., Hidaka, M., Matsuoka, D., Murakami, K., and Kako, S. (2022). The BeachLitter dataset for image segmentation of beach litter. *Data Brief*, 42(108072):108072.
- Tamin, O., Moug, E. G., Dargham, J. A., Yahya, F., Farzamnia, A., Sia, F., Naim, N. F. M., and Angeline, L. (2023). On-shore plastic waste detection with yolov5 and rgb-near-infrared fusion: A state-of-the-art solution for accurate and efficient environmental monitoring. *Big Data and Cognitive Computing*, 7(2).
- Terven, J., Córdova-Esparza, D.-M., and Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.*, 5(4):1680–1716.
- Veetil, B. K., Hong Quan, N., Hauser, L. T., Doan Van, D., and Quang, N. X. (2022). Coastal and marine plastic litter monitoring using remote sensing: A review. *Estuarine, Coastal and Shelf Science*, 279:108160.
- Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. Disponível em <https://arxiv.org/abs/2207.02696>. Acesso em 16 de maio de 2024.
- Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. (2020). Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop)*.