

Detecção de Espécies de Peixes no Bioparque Pantanal usando Aprendizado Profundo

Daniel Lima Grance

¹Universidade Federal de Mato Grosso do Sul (UFMS)
Campo Grande – MS – Brazil

daniel.grance@ufms.br

Abstract. *This paper describes an experiment using Computer Vision. In view of the great demand of visitors in the Bioparque Pantanal, the help of computational means has become necessary so that guided visits be more fluid, therefore, the idea of using Computer Vision is presented, making it possible that visitors be able to identify species of several fish by capturing pictures using their own mobile devices. This is accomplished with Convolutional Neural Networks, that receive an image, propose regions containing fish and identify their respective species. This article presents the results of the experiment, as well as specifications and conditions of the used dataset.*

Resumo. *Este artigo descreve um experimento utilizando Visão Computacional. Tendo em vista a grande demanda de visitantes no Bioparque Pantanal, tornou-se necessário a ajuda de meios computacionais para que as visitas guiadas sejam mais fluidas, sendo assim, surge-se a ideia de utilizar Visão Computacional para que seja possível que visitantes consigam identificar as espécies de diversos peixes ao capturar imagens em seus aparelhos celulares. Isso é feito através de Redes Neurais Convolucionais, que ao receber uma imagem, propõem regiões contendo peixes e identificam suas respectivas espécies. Este artigo apresenta os resultados do experimento, assim como especificações e condições da base de dados utilizada.*

1. Introdução

O Bioparque Pantanal, ou como também é chamado, Aquário do Pantanal, é um complexo turístico e ambiental brasileiro, sendo o maior aquário de água doce do mundo. Localizado na cidade de Campo Grande, no Mato Grosso do Sul, o complexo possui aproximadamente 19 mil metros quadrados e conta com 33 tanques que abrigam 220 espécies, sendo 151 espécies do Pantanal, 55 da Amazônia, 14 africanas e outras de diversas regiões. Apesar de ser considerado o maior ponto turístico da cidade de Campo Grande e um dos maiores do Brasil, o Bioparque Pantanal não apenas conta com o turismo como seu único pilar, mas também a educação e a preservação ambiental.

O aquário foi inaugurado no final do ano de 2020, e desde então, o número de visitantes se manteve em uma constante alta. Além da alta demanda, o elevado número de visitas diárias dificulta uma visita guiada adequada. Surge então, a necessidade de que o visitante tenha as informações das espécies e dos aquários em seu dispositivo móvel, ou que elas sejam embarcadas em dispositivos disponíveis para o uso do visitante, permitindo que uma visita ao Bioparque Pantanal seja mais fluida e confortável, com o

visitante podendo obter informações sobre o que é visto por conta própria, em seu tempo e seguindo o seu interesse.

Este artigo propõe o uso de Visão Computacional para que isso seja possível, reconhecendo automaticamente espécies de peixes a partir de uma imagem. Isso é realizado através de Redes Neurais Convolucionais, ou CNNs, que obtêm um conjunto de imagens rotuladas de treinamento, e através de um número de passos, vão aprendendo a propor possíveis regiões contendo peixes e também, características específicas que as ajudam a reconhecer e determinar a espécie de um determinado peixe analisado.

Os resultados obtidos por meio da aplicação de visão computacional para o reconhecimento automático de espécies de peixes nos aquários do Bioparque Pantanal foram promissores. Foram avaliados dois datasets correspondentes a dois aquários com diferentes espécies de peixes, obtendo-se mAP de 0.85 e 0.80, respectivamente. Além disso, um terceiro experimento foi realizado, combinando as espécies dos dois datasets, a fim de avaliar a capacidade de generalização do método. Nesse caso, o mAP obtido foi de 0.81. Esses resultados indicam que a abordagem baseada em redes neurais convolucionais possui um bom desempenho na identificação das espécies de peixes, tanto individualmente quanto em uma combinação de diferentes aquários.

Este artigo está organizado da seguinte forma: na Seção 2, será realizada uma revisão de literatura para discutir estudos anteriores relacionados ao reconhecimento de espécies de peixes por meio de visão computacional. Na Seção 3, serão apresentados os materiais e métodos utilizados, incluindo o dataset de imagens, as especificações dos datasets, o método de detecção e o protocolo experimental. Os resultados quantitativos e qualitativos serão apresentados na Seção 4, analisando os valores de mAP obtidos e destacando a capacidade das redes neurais convolucionais na identificação das espécies de peixes. Por fim, na Seção 5, será apresentada a conclusão deste estudo, resumindo os principais achados e apontando possíveis direções futuras de pesquisa.

2. Revisão de Literatura

A literatura apresenta alguns métodos de aprendizagem para detecção e classificação de peixes. [Ogunlana et al. 2015] propuseram o uso de um algoritmo de classificação baseado em SVM [Cortes and Vapnik 1995] analisando primordialmente algumas características físicas do peixe de forma individual, como tamanho e formato. O dataset utilizado possui imagens de características que foram extraídas de 150 peixes de 2 espécies diferentes. Apesar da acurácia obtida de 78,59%, a metodologia é complexa e difícil de ser replicada em larga escala.

Para facilitar o reconhecimento posterior de espécies de peixes, [Rathi et al. 2017] apresentaram um pré-processamento das imagens para a remoção de ruídos como aspectos do ambiente, obstáculos e objetos que não são peixes. Essa abordagem é realizada por meio de CNNs para a classificação. Neste estudo foi utilizado o dataset Fish4Knowledge [Hoashi et al. 2012], contendo 27142 imagens de peixes de diversas espécies. O resultado obtido foi uma acurácia de 96,29%.

[B S et al. 2020] propuseram a classificação dos peixes com uma única CNN inicialmente, porém os resultados obtidos não foram satisfatórios. Após perceberem a dependência do ambiente, decidiram utilizar uma fase de detecção antes da classificação,

utilizando CNNs derivadas do VGGNet [Simonyan and Zisserman 2014], obtendo um resultado de aproximadamente 90% de acurácia. Eles ressaltam, entretanto, que o processo de detecção é lento, levando cerca de 5 segundos para que uma imagem seja processada.

[Knausgård et al. 2020] apresentaram uma abordagem de detecção e classificação para peixes de água temperada utilizando YOLOv3 [Redmon and Farhadi 2018] para a detecção dos peixes e a arquitetura CNN-SENet [Hu et al. 2017] para a classificação das espécies. Neste caso, foi utilizado *Transfer Learning* [Pan and Yang 2010] com os pesos obtidos do pré-treinamento com o ImageNet [Deng et al. 2009]. A solução alcançou precisão de 99,27% com o modelo de pré-treinamento e 87,74% com o modelo de pós-treinamento, apontando que a solução é viável com um conjunto de dados mais extenso.

Em um estudo mais recente, [Prasetyo et al. 2021] propuseram o Multi-Level Residual (MLR) como estratégia de rede residual visando manter características de baixo e médio nível no último bloco de convoluções. O VGGNet foi utilizado como base da arquitetura de CNN, que levou o nome de MLR-VGGNet. Utilizando o peso inicial da VGGNet pré-treinada, a MLR-VGGNet obteve precisão de 99,69% nos conjuntos de dados Fish-gres e Fish4Knowledge.

Autor	Modelo Utilizado	Precisão (%)	Acurácia (%)
[Ogunlana et al. 2015]	SVM	-	78,59
[Rathi et al. 2017]	CNN (Não especificada)	-	96,29
[B S et al. 2020]	VGGNet	-	90
[Knausgård et al. 2020]	YOLOv3 + CNN-SENet	87,74	-
[Prasetyo et al. 2021]	MLR-VGGNet	99,69	-

Table 1. Comparação de Resultados e Modelos Utilizados

3. Materiais e Métodos

3.1. Dataset de Imagens

As imagens dos conjuntos de dados utilizados neste estudo foram capturadas por alunos da Universidade Federal de Mato Grosso do Sul, em uma excursão ao Aquário no dia 4 de junho de 2022. Ao todo, seis alunos participaram da captura que resultou na base de dados principal. Cada imagem passou por um processo de rotulação, onde foram anotados retângulos envolventes (bbox) para cada peixe, assim como suas respectivas espécies.

A Figura 1 apresenta o número de Bounding Boxes em intervalos de tamanhos em pixels para que possamos ter uma base do tamanho médio dos bbox anotados. A maioria dos Bounding Boxes (11493) possui até 300 mil pixels, isso representa que muitas imagens foram capturadas com uma visão mais ampla dos aquários, fazendo com que um grande número dos Bounding Boxes possuam tamanho pequeno ou médio. Entretanto, existem alguns deles com tamanho maior do que 5 milhões de pixels, indicando que algumas imagens foram capturadas de forma individual. Como o peixe ocupa um grande espaço da imagem, o seu retângulo envolvente tende a ser muito grande.

O dataset utilizado é consideravelmente representativo, contendo imagens capturadas por dispositivos de diferentes qualidades em diferentes condições de iluminação. Como exemplo, um deles sendo uma câmera fotográfica com alta resolução e um outro

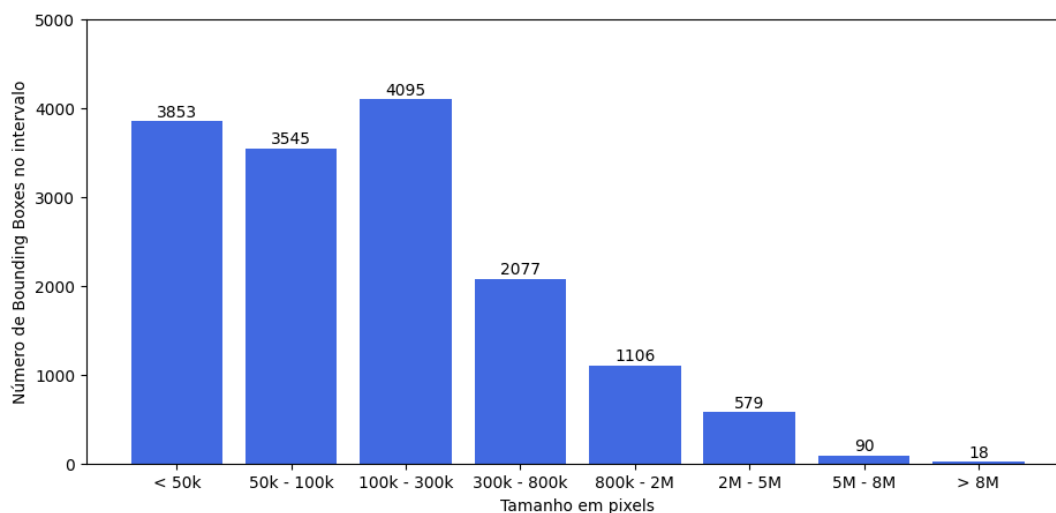


Figure 1. Tamanho dos Bounding Boxes em pixels

sendo um dispositivo de baixa qualidade. Ao todo, foram utilizados sete dispositivos de captura, sendo eles, seis aparelhos celulares e uma câmera fotográfica: Iphone 7, Iphone 8, Samsung Galaxy A22, Rog Phone 2, Samsung Galaxy S21 FE, um dispositivo desconhecido de baixa qualidade e Nikon d3200.

Foram fotografados dois aquários, chamados "Rios Grandes", contendo as espécies Pacu, Dourado e Jaú, e "Rios de Bonito", contendo as espécies Cachara, Curimbatá, Piau e Piraputanga.

3.1.1. Especificações do Dataset

As imagens capturadas foram organizadas em 3 bases de imagens. A base de imagens principal, denominada "Aquário 7 Classes", refere-se aos dois aquários observados e contém 3065 imagens que, em sua grande maioria, possuem diversos peixes nelas, resultando em mais duas subdivisões. Isto foi feito para que fosse realizada posteriormente uma comparação entre os resultados e uma análise se os resultados seriam melhores com mais ou menos espécies incluídas de uma vez no aprendizado. Os peixes contidos nas imagens estão divididos em 7 classes, como indica o nome da base: Cachara, Curimbatá, Dourado, Jaú, Pacu, Piau e Piraputanga. São apresentados exemplos de peixes dessas espécies na Figura 2.

A base "Rios de Bonito", proveniente da primeira subdivisão refere-se ao aquário "Rios de Bonito" e contém 1745 imagens. Os peixes contidos nas imagens estão divididos em 4 classes: Cachara, Curimbatá, Piau e Piraputanga. A base "Rios Grandes", proveniente da segunda subdivisão, refere-se ao aquário "Rios Grandes" e contém 1320 imagens. Os peixes contidos nas imagens estão divididos em 3 classes: Dourado, Jaú e Pacu.

O número de Bounding Boxes anotados por espécie no total pode ser observado na Figura 3. Observa-se que a espécie mais presente é a Piraputanga, isso ocorre pela maior presença de peixes da espécie em seu aquário. As imagens capturadas com uma



Figure 2. Espécies de peixes observadas

visão mais ampla do aquário possuíam muitos peixes da espécie, resultando em muitas rotulações. As espécies menos presentes, Curimbatá e Jaú, ambas possuíam apenas 1 peixe de sua espécie em seus respectivos aquários no momento da captura, resultando em um número menor de rotulações. O Jaú, inclusive, ficou grande parte da captura escondido, o que dificultou que houvessem muitas imagens dele. Esse desbalanceamento do número de peixes de cada espécie gera um desafio. O Jaú por exemplo, sendo um peixe parecido com o Pacu, requer um número maior de imagens rotuladas, o que geraria uma representatividade maior de ângulos e iluminações, o que não ocorre, visto seu baixo número de imagens.

3.2. Método de Detecção de Objetos

Para a detecção dos peixes, o método Faster R-CNN foi utilizado [Ren et al. 2016]. Este método surge com a intenção de resolver problemas do método Fast R-CNN [Girshick 2015], principalmente em relação ao tempo de processamento, que era razoavelmente alto pelo fato de utilizar o Selective Search [Uijlings et al. 2013] para propor regiões contendo objetos na imagem original para depois transpor os retângulos envolventes para o Feature Map (Mapa de Características) gerado pelas Redes Neurais Convolucionais.

O Faster R-CNN introduz então, a ideia de realizar o processamento sem o uso do Selective Search, visando inserir uma Rede de Proposição de Regiões (RPN) após a última camada de convolução. Com a proposta de proposição de regiões sendo incluída

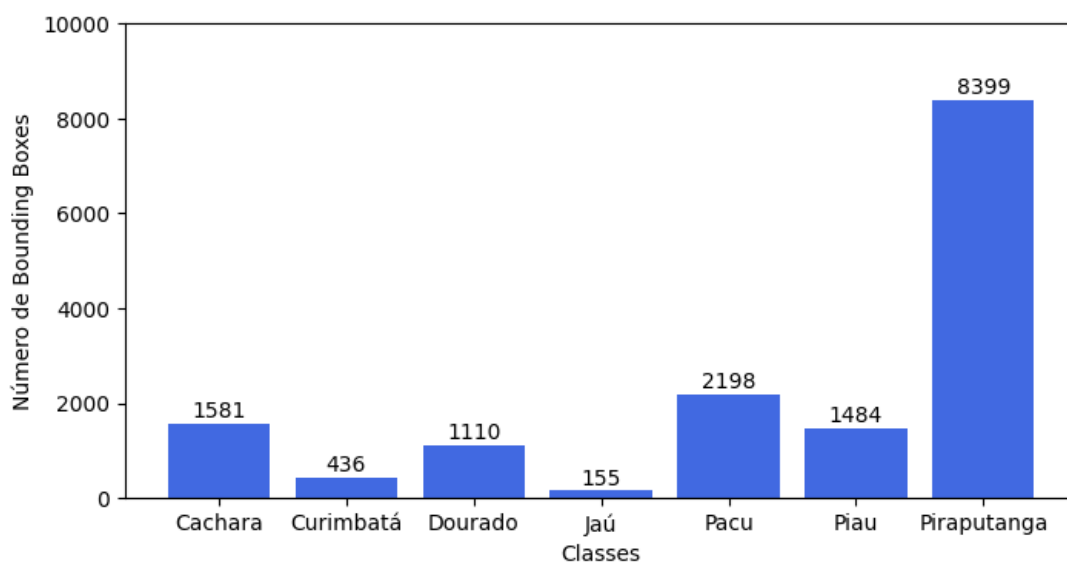


Figure 3. Números de Bounding Boxes por espécie

na própria Rede Neural Convolutacional, o tempo de processamento tende a ser consideravelmente menor.

Inicialmente, passa-se uma imagem de entrada por Camadas de convolução (CNN), onde as Redes Neurais Convolucionais irão aprender características sobre a imagem, obtendo-se então o Feature Map. O grande diferencial deste método é que depois de se obter o Feature Map, existe uma rede conectada à CNN para propor retângulos contendo objetos, esta é a Region Proposal Network (RPN). A RPN trata-se de uma janela deslizante que, utilizando-se de uma pequena Rede, propõe a probabilidade desta região contida na janela possuir um objeto e corrige as dimensões do retângulo com um Regressor Linear. Com os retângulos envolventes propostos e o Feature Map gerado pela CNN, os recortes destas regiões são realizados e passados para um Classificador.

A sequência de passos após a proposição de retângulos envolventes é semelhante aos passos finais do Fast R-CNN. Os retângulos envolventes que não contém objetos são descartados pela RPN. Os retângulos contendo objetos são passados pela Rol Pooling, é então realizada a classificação destes objetos, gerando as probabilidades para cada classe possível. Posteriormente, são realizadas correções das dimensões do retângulo, utilizando um Regressor Linear, visando que o retângulo tenha medidas condizentes com seu objeto em questão.

3.3. Protocolo Experimental

Os três datasets foram divididos em 70% para treinamento, 20% para teste e 10% para validação. O conjunto de treino é usado durante o aprendizado do Faster R-CNN, enquanto o conjunto de validação é utilizado para avaliar o desempenho do modelo durante o treinamento. Por fim, usamos o conjunto de teste após o treinamento para que sua eficácia seja avaliada, passando imagens não vistas no treinamento para que o modelo as classifique. A relação da quantidade de imagens em cada conjunto e dataset pode ser vista na Tabela 2.

Aquário	Quantidade de imagens	Treinamento	Validação	Teste
Aquário "Rios De Bonito"	1320	923 (70%)	132 (10%)	265 (20%)
Aquário "Rios Grandes"	1745	1221 (70%)	174 (10%)	350 (20%)
Aquário "7 Classes"	3065	2145 (70%)	306 (10%)	614 (20%)

Table 2. Relação do número de imagens nos conjuntos e datasets

Para avaliar o desempenho dos modelos, foi utilizada a taxa de acerto, que foi obtida por meio da análise das Matrizes de Confusão geradas para cada modelo. Essas matrizes permitem verificar os Falsos Negativos, bem como outras métricas relevantes para a avaliação do modelo. Além da taxa de acerto, consideramos também o Índice de Intersecção sobre União (IoU, do inglês Intersection over Union). Este índice mede a sobreposição entre a área indicada pelo modelo e a área real do Bounding Box que foi previamente anotado. Neste experimento, são relevantes dois valores de IoU: 50%, que considera como acerto sobreposições que excedem 50% da área total e 75%, que de forma mais rigorosa, considera como acerto sobreposições que excedem 75% da área total.

O modelo foi treinado por 24 épocas utilizando o otimizador SGD [Ruder 2017] com uma taxa de aprendizado de 0,001 e o tamanho do lote definido como 64. Neste experimento utilizamos o MMDetection [Contributors 2018] em um sistema com a seguinte configuração: AMD Ryzen 7 5800X 8-Core Processor, 64GB de RAM e duas GPUs NVIDIA GeForce RTX 3060.

4. Resultados

4.1. Resultados Quantitativos

Os resultados obtidos para cada dataset individualmente estão apresentados na Tabela 3. Observamos que o Aquário Rios Grandes obteve um desempenho significativamente melhor em relação aos outros modelos, com uma pontuação de 0.858 em relação ao IoU de 50%. Isso indica que o modelo treinado exclusivamente com as três classes presentes no Aquário Rios Grandes apresentou uma capacidade de detecção mais precisa e acurada para essas espécies específicas de peixes.

Por outro lado, o Aquário Rios de Bonito, que contém quatro espécies de peixes, obteve um resultado ligeiramente inferior, com mAP de 0.803. A adição de uma classe adicional ao modelo parece ter impactado negativamente a precisão da detecção. Esse resultado pode estar relacionado a um aumento da complexidade do problema devido à maior diversidade de espécies envolvidas.

Surpreendentemente, o Aquário 7 Classes, que combina as espécies dos dois datasets anteriores, apresentou um desempenho próximo ao Aquário Rios de Bonito, com mAP de 0.810. Esses resultados indicam que a adição das classes do Aquário Rios de Bonito ao Aquário Rios Grandes não resultou em uma melhora significativa na detecção das espécies de peixes. Entretanto, não houve uma degradação considerável.

Para avaliar detalhadamente os resultados de cada aquário, nós apresentamos as matrizes de confusão na Figura 4. Podemos observar na Matriz de Confusão 4(a) que no "Aquário Rios de Bonito", a espécie com melhor resultado foi o Cachara, com uma taxa de acerto de 89%, devido aos seus traços muito específicos que o diferencia das outras espécies em seu aquário. A espécie com piores resultados foi o Piau, apresen-

Aquários	IOU 50%	IOU 75%
Aquário Rios Grandes	0.858	0.630
Aquário Rios de Bonito	0.803	0.418
Aquário 7 Classes	0.810	0.494

Table 3. Resultados de mAP para os aquários observados

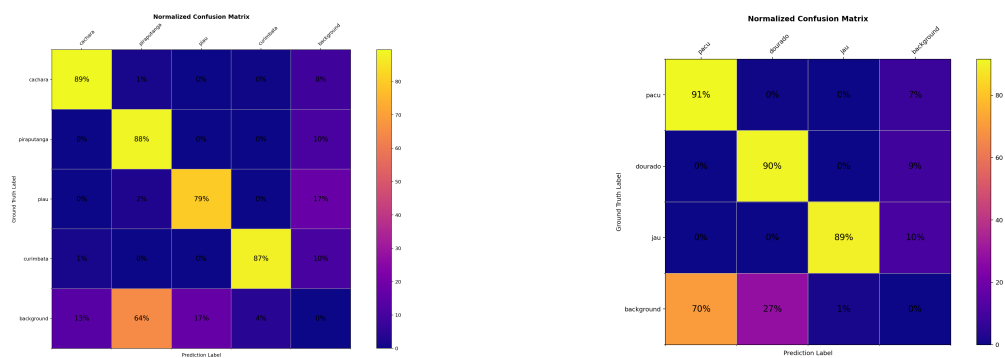
tando taxa de acerto de aproximadamente 79%, o que é significativamente menor em comparação com as outras. Essa espécie teve uma taxa de 17% de Falsos Negativos. Como muitas imagens possuem uma certa distância do aquário, o Piau não é percebido em um número considerável de vezes devido ao seu tamanho pequeno, sendo considerado erroneamente como fundo da imagem. As espécies que mais tiveram confusão foram Piau e Piraputanga, contudo, representando um número bem baixo. Podemos concluir que as espécies neste aquário não tiveram um valor significativo de confusão entre si.

Já na Matriz de Confusão 4(b) notamos que para o "Aquário Rios Grandes", o Pacu foi a espécie com melhor resultado, obtendo uma taxa de assertividade de 91%. Apesar de ser parecido com o Jaú, ele é mais perceptível por seu tamanho e cor. A espécie com pior resultado e mais Falsos Negativos foi o Jaú com aproximadamente 89% de taxa de acerto e 10% de Falsos Negativos. Isso ocorre pois o Jaú tem preferência por regiões do aquário com baixa iluminação, como o fundo do aquário. Portanto, a falta de iluminação prejudica a visibilidade do peixe e dificulta sua detecção nas imagens analisadas. As espécies neste aquário também não tiveram um valor significativo de confusão entre si. Podemos concluir, portanto, que em ambos os aquários, os principais erros foram de falsos negativos, isto é, quando um peixe deixa de ser detectado.

Quando analisamos a Figura 4(c) que se refere ao "Aquário 7 Classes", podemos ver que algumas espécies que se assemelham possuem piora nos resultados, como as espécies Dourado e Piraputanga. Analisando as matrizes confusões dos aquários "Aquário 7 Classes" e "Rios Grandes" por exemplo, nota-se que a assertividade quanto à espécie Dourado diminui no caso do dataset unificado. No "Aquário Rios Grandes", a taxa de acerto para essa espécie é de 90%, enquanto no dataset unificado essa taxa diminui para 80%. Isso ocorre pois no aquário "Rios Grandes", nenhum peixe se assemelha fisicamente ao Dourado. Entretanto, ao ser analisado com o aquário "Rios de Bonito", existe uma certa confusão na classificação pois o peixe Piraputanga é considerado. Por outro lado, a única espécie que teve melhora em seus resultados quanto a Verdadeiros Positivos foi o Jaú, com sua taxa de acerto aumentando de 89% para 96%. O Piraputanga se manteve na mesma porcentagem de assertividade e as demais espécies resultaram em taxas piores.

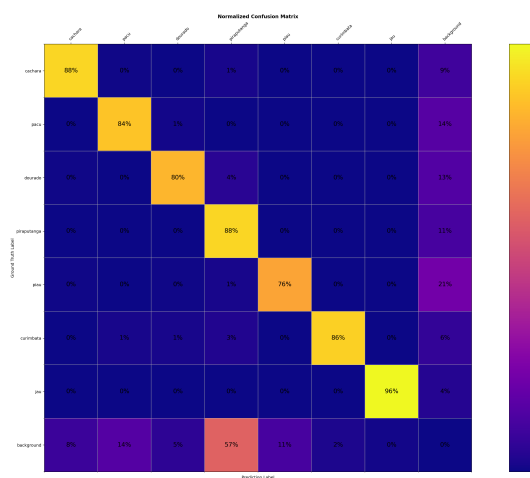
Com a interpretação das matrizes de confusão fica evidente que alguns aquários podem possuir espécies parecidas, resultando em uma confusão entre espécies e consequentemente, piora na assertividade. Assim como é observado que quando são analisadas muitas espécies de uma vez, o número de Falsos Negativos é maior. No caso deste experimento, a única espécie que teve melhora no número de Falsos Negativos quando analisados os resultados do treinamento dos aquários de forma unificada foi o Jaú.

Logo, torna-se mais viável que haja um modelo para cada aquário individualmente, tendo em vista a melhora dos resultados nestes casos. Desta forma, também



(a) Confusion Matrix - "Rios de Bonito"

(b) Confusion Matrix - "Rios Grandes"



(c) Confusion Matrix - "Aquário 7 Classes"

Figure 4. Matrizes Confusão dos datasets observados

é proposta a ideia de que um sistema de GPS seja implantado junto com o sistema de detecção, determinando a localização do usuário para a identificação do aquário que está sendo analisado ou até mesmo um módulo que reconheça o aquário.

4.2. Resultados Qualitativos

Quando analisamos os resultados de modo qualitativo, é possível observar alguns aspectos dos resultados das detecções geradas pelo modelo analisado. Discutiremos nesta seção, o desempenho do modelo, especialmente em possíveis desafios de casos reais, como diferenças de tamanhos das espécies dos peixes observados, imagens com iluminação que não favorece a identificação dos peixes, entre outros.

Como mencionado anteriormente sobre a semelhança de algumas espécies, o modelo se mostrou efetivo mesmo em casos onde não estão claros os traços que diferenciam certas espécies. Podemos observar na Figura 5(a) como o modelo se saiu ao ter que classificar um peixe da espécie Pacu, que se assemelha ao peixe da espécie Jaú. Entretanto, alguns casos tornam a classificação mais desafiadora, como casos onde o peixe está em um ângulo que não favorece a identificação dos traços que o diferenciam de uma espécie parecida. É possível observar na Figura 5(b) um erro de classificação devido a posição em que um peixe da espécie Jaú foi analisado, fazendo com que o modelo o interpretasse como da espécie Pacu.



(a) Classificação correta Jaú



(b) Classificação errada Jaú

Figure 5. Classificação de espécies de peixes parecidas

Em relação aos diferentes tamanhos dos peixes, o modelo se manteve consistente nas classificações, obtendo ótimos resultados na classificação dos peixes, o que poderia ser dificultoso. O peixe Piau, por exemplo, possui um tamanho muito pequeno, enquanto o Jaú e o Cachara possuem tamanhos maiores. Os retângulos envolventes se mantiveram apropriados, mesmo nesses casos. Observamos na Figura 6 os retângulos envolventes de diferentes tamanhos para as diferentes espécies de peixes.



(a) Bounding Box pequeno



(b) Bounding Box grande

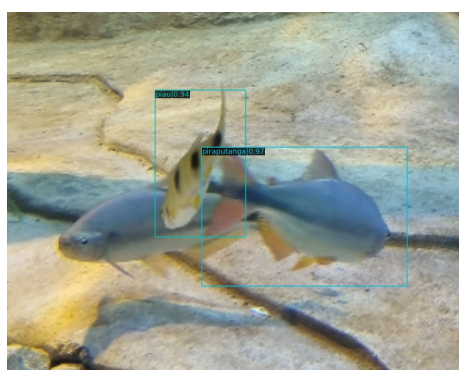
Figure 6. Bounding Boxes de diferentes tamanhos

Um outro ponto a se destacar é que, como dito anteriormente, foram usados dispositivos de diferentes qualidades, inclusive um dispositivo de baixa qualidade, fazendo com que os datasets sejam representativos e resultando em uma boa assertividade quando o modelo precisa realizar suas detecções em imagens de baixa qualidade ou com condições de iluminação não adequadas. Um exemplo disso é apresentado na Figura 7.

Por fim, a Figura 8(a) mostra um exemplo de erro de detecção por causa da sobreposição de peixes. Em algumas imagens em que esse caso ocorre, o modelo apresentou resultados não tão bons, algumas vezes apenas detectando o peixe por cima, e em outras vezes deixando de detectar os peixes. Como os peixes estão sobrepostos, o modelo acaba encontrando uma dificuldade em identificar características que foram aprendidas. Todavia, como foram anotados peixes sobrepostos no processo de rotulação, o modelo consegue detectar os peixes em alguns casos, como mostra a Figura 8(b).



Figure 7. Resultado das detecções em imagens de baixa qualidade



(a) Erro de detecção em peixes sobrepostos



(b) Acerto de detecção em peixes sobrepostos

Figure 8. Exemplos de detecção com peixes sobrepostos

5. Conclusão

Neste trabalho, apresentamos e avaliamos os resultados para a detecção e classificação de peixes do Bioparque Pantanal utilizando o método Faster R-CNN. Os resultados obtidos foram altamente promissores mesmo com os desafios apresentados por algumas imagens. O método se mostrou eficiente, resultando em uma taxa de acertos média de 90% para o aquário com melhor desempenho. Acreditamos que este trabalho contribui para o avanço da pesquisa e aplicação de técnicas avançadas de análise de imagens no contexto do Bioparque Pantanal e esperamos que os resultados incentivem outras pesquisas e aplicações para a ampliação do nosso conhecimento destes ambientes.

References

- B S, R., G N, D. S., Reddy, S., Kakwani, D., and Bhattad, N. (2020). *Fish Detection and Classification Using Convolutional Neural Networks*, pages 1221–1231.
- Contributors, M. (2018). MMDetection: Open MMLab Detection Toolbox and Benchmark. <https://github.com/open-mmlab/mmdetection>.
- Cortes, C. and Vapnik, V. (1995). Support vector machines: History, theory, and application. *Journal of Machine Learning*, 20(3):273–297.

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.
- Girshick, R. (2015). Fast r-cnn.
- Hoashi, K., Nakamura, Y., Suzuki, R., and Ogawa, H. (2012). Fish4knowledge: Collecting and analyzing massive coral reef fish video data. In *International Conference on Distributed Computing and Networking*, pages 157–169. Springer.
- Hu, J., Shen, L., and Sun, G. (2017). Squeeze-and-excitation networks. *CoRR*, abs/1709.01507.
- Knausgård, K. M., Wiklund, A., Sjørdalen, T. K., Halvorsen, K., Kleiven, A. R., Jiao, L., and Goodwin, M. (2020). Temperate fish detection and classification: a deep learning based approach. *CoRR*, abs/2005.07518.
- Ogunlana, S. O., Olabode, O., Oluwadare, S. A., and Iwasokun, G. B. (2015). Fish classification using support vector machine.
- Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.
- Prasetyo, E., Suciati, N., and Fatichah, C. (2021). Multi-level residual network vggnet for fish species classification. *Journal of King Saud University - Computer and Information Sciences*, 34.
- Rathi, D., Jain, S., and Sreedevi, I. (2017). Underwater fish species classification using convolutional neural network and deep learning. pages 1–6.
- Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *CoRR*, abs/1804.02767.
- Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks.
- Ruder, S. (2017). An overview of gradient descent optimization algorithms.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Uijlings, J., van de Sande, K., Gevers, T., and Smeulders, A. (2013). Selective search for object recognition. *International Journal of Computer Vision*.