

Fish Detection and Measurement based on Mask R-CNN

Wan Song Rocha

Departamento de Ciência da Computação
Universidade Federal de Rondônia
Porto Velho, Brasil
Email: wansong@unir.br

Carolina Rodrigues da Costa Doria

Laboratório de Ictiologia e Pesca
Universidade Federal de Rondônia
Porto Velho, Brasil
Email: carolinarcdoria@unir.br

Carolina Yukari Veludo Watanabe

Departamento de Ciência da Computação
Universidade Federal de Rondônia
Porto Velho, Brasil
Email: carolina@unir.br

Abstract—Medidas morfológicas de peixes, extraídas de seu contorno, são indicadores importantes com aplicações na indústria pesqueira bem como para monitoramento e preservação das espécies. Assim, técnicas automáticas da extração do contorno tem sido exploradas e o desenvolvimento de métodos de segmentação de imagens de peixes são necessárias. Portanto, o objetivo deste trabalho foi desenvolver um método para medir automaticamente o comprimento do peixe, usando a rede Mask R-CNN na tarefa de segmentação de peixes. Comparando as medidas do comprimento do peixe do método proposto com a medida realizada manualmente, foi obtido um erro relativo médio de apenas 2,26%. Assim, um sistema para a automatização desse processo pode ser mais rápido, mais eficaz e mais escalável.

Abstract—Morphological measurements of fish, extracted from their contour, are essential indicators, with applications in the fishing industry as well as for monitoring and preserving species. Thus, automatic contour extraction techniques have been explored, and the development of methods for segmenting fish images is necessary. Therefore, the goal of this work was to develop an approach to automatically measure the length of the fish, using the Mask R-CNN network in the fish segmentation task. Comparing the fish length measurements of the proposed method with the size performed manually, an average relative error of only 2.26% was obtained. Thus, a system for automating this process can be faster, more productive, and more scalable.

I. INTRODUÇÃO

Devido a ação do homem sobre a comunidade de peixes, são necessários estudos sobre impactos ambientais. Um dos principais indicadores na maricultura, bem como em bacias de água doce, são características morfológicas dos peixes, como comprimento e largura do corpo, comprimento do pedúnculo caudal, diâmetro da pupila e do olho [1] [2].

Atualmente, a maioria dos métodos de medição de peixes dependem de trabalho manual, utilizando fita métrica e um operador humano para a atividade. Esse tipo de operação requer um alto nível técnico, possui alta intensidade de trabalho e baixa eficiência. Ademais, o resultado da medição pode ser facilmente afetado e a medição demanda de um longo período de tempo [1].

Portanto, a coleta das medidas de um peixe de forma manual, em um ambiente real, pode acarretar na limitação da sua qualidade e quantidade. Comparado com a classificação e medição manual, um sistema de processamento automático pode ser mais rápido, menos propenso a erros, mais escalável e mais vantajoso para indivíduos sem treinamento especializado

[3]. Neste sentido, as técnicas de processamento e análise de imagens têm chamado cada vez mais atenção da indústria e da ciência da aquicultura [1], pois permitem uma abordagem não-extrativa e não-letal para a coleta de dados da pesca, como medição do tamanho dos peixes, estimativa de capturas, conformidade regulatória, reconhecimento de espécies e contagem populacional [3]. A extração de contorno a partir de imagens é a base de muitas aplicações de imagem de peixes, como diagnóstico e alerta precoce de doenças, comportamento animal, processamento de produtos aquáticos [4].

Contudo, para automatizar a medição dos peixes é necessário um método que possibilite a segmentação das imagens que contenham esses peixes. Uma técnica que tem se mostrado muito promissora na tarefa de segmentação de imagens é o uso de redes neurais convolucionais (do inglês, Convolutional Neural Networks – CNN) [5]. Um dos pontos fortes dessa técnica é o fato de que nenhuma identificação explícita de recurso é necessária, não é preciso especificar as características do item para que a CNN aprenda a reconhecê-lo [6].

No trabalho apresentado por [1] é desenvolvido um método utilizando a técnica do Mask R-CNN para efetuar a segmentação do peixe e por meio dessa segmentação realizar a medição de algumas características morfológicas do peixe, tais como: tamanho do corpo do peixe, largura do corpo, tamanho do pedúnculo caudal, largura do pedúnculo caudal, diâmetro da pupila e diâmetro do olho. A extração dessas informações são utilizadas para acompanhar o crescimento do peixe. O método conseguiu bons resultados em imagens com fundos puros, mas em fundos complexos a acurácia começou a decair.

Já no artigo de [7] utiliza-se a técnica do YOLO (You Only Look Once) para efetuar a detecção de peixes em tempo real no fundo do mar. Em contraste com outras arquiteturas CNN, o YOLO é uma rede unificada que prevê simultaneamente a posição, o tamanho e a probabilidade da classe do objeto, isso o torna um método veloz, possibilitando a sua utilização para detecção em tempo real. Visto que os peixes possuem cores protetoras, se misturando com o ambiente, em muitos casos o fundo do mar foi classificado erroneamente como um peixe, diminuindo a precisão do método.

Em revisão da literatura, observa-se, como os trabalhos citados anteriormente, que as técnicas desenvolvidas consideram

apenas peixes marítimos, com resultados promissores na tarefa de segmentação de peixes marítimos usando CNN.

Como peixes marítimos possuem características diferentes dos de água doce, como cor, textura e forma, técnicas de segmentação de imagens para este tipo de nicho precisam ser desenvolvidas, visando o monitoramento de espécies com importância reconhecida na pesca comercial e de subsistência.

Assim, o objetivo deste trabalho foi implementar método de rede neural convolucional para segmentar imagens de peixes de água doce, com base na técnica Mask R-CNN, a fim de extrair medidas de comprimento de maneira automática. As imagens dos peixes foram adquiridas junto ao Laboratório de Ictiologia e Pesca, de peixes da bacia do Rio Madeira, Porto Velho, Rondônia, que faz parte da Amazônia Ocidental. As medidas extraídas foram comparadas com as medidas realizadas por biólogos especialistas da área.

Este trabalho está estruturado da seguinte maneira. A seção 2 apresenta o conceito de Redes Neurais Convolucionais, bem como a técnica Mask R-CNN, escolhida para este trabalho. A seção 3 apresenta o método desenvolvido. A seção 4 mostra os experimentos e resultados identificados na efetivação da técnica empregada. Por fim, a seção 5 traz as conclusões.

II. CONCEITOS RELACIONADOS

A. CNN

Uma CNN (do inglês, *Convolutional Neural Network*) é um modelo específico de rede neural utilizada normalmente para classificação de imagens. Possui esse nome por utilizar a operação de convolução.

Como dito, a camada de convolução desempenha um papel vital na forma como uma CNN funciona, uma CNN pode conter o número variado de camadas de convolução, dependendo da complexidade da imagem e do modelo proposto. O resultado obtido quando uma imagem é inserida em uma camada de convolução é denominado mapa de características. Esse mapa comporta as principais características da imagem, possibilitando que a rede neural consiga reconhecer o objeto contido na mesma de forma simplificada e com uma maior performance [8].

B. Mask R-CNN

A rede Mask R-CNN foi desenvolvida por [9], que por sua vez foi adaptada e utilizada neste artigo para efetuar a segmentação e detecção do peixe na imagem. Mask R-CNN é um algoritmo que pode realizar detecção, classificação e segmentação instantânea simultaneamente em uma rede neural.

O comportamento dessa rede neural é dividida em etapas, como apresentado na Figura 1. Primeiramente, a imagem é inserida em uma rede ResNext101 e FPN (Feature Pyramid Network) treinada para obter o seu mapa de características. A ResNext101 é uma rede neural convolucional com 101 camadas que pode melhorar a precisão da extração das informações contidas na imagem. Segundo, um número fixo de ROIs (*Region of Interest*) é predefinido para cada imagem, no caso deste trabalho foi utilizado 200 ROIs. ROI é uma

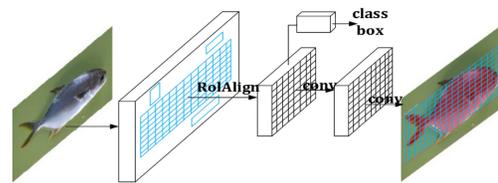


Fig. 1. Fluxograma do algoritmo de Mask R-CNN. Fonte: [1].

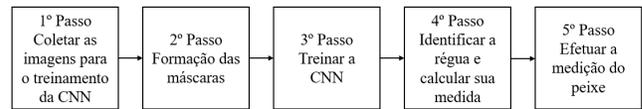


Fig. 2. Fluxograma do método proposto



Fig. 3. Exemplo de imagem contendo o peixe e sua respectiva máscara.



Fig. 4. Formação da máscara do peixe.

caixa delimitadora de um objeto candidato, se trata de um retângulo ao redor da área que a rede considera ter um objeto, no caso, o peixe. Terceiro, essas ROIs são inseridas na RPN (*Region Proposal Network*), a qual irá determinar se em cada ROI presente existe o objeto desejado, e retorna uma caixa delimitadora contendo o objeto localizado. Essas duas etapas excluem as ROIs que não possuem objetos. Após, a operação de ROIAlign é executada nas ROIs restantes para alinhar as características extraídas, possibilitando maior acurácia nas caixas delimitadoras. Por fim, é realizada a classificação e retorno das ROIs [1].

III. METODOLOGIA

Nesta seção estão apresentadas as etapas para a execução do método proposto, conforme mostra a Figura 2.

Primeiramente é necessário coletar as imagens dos peixes para o treinamento da rede. A base de imagens foi obtida no Laboratório de Ictiologia e Pesca da Universidade Federal de Rondônia (UNIR). Os peixes foram retirados de ambiente de armazenamento com formol e colocados em um recipiente branco. Foram medidos com fita métrica e régua pelo especialista biólogo e essas medidas foram armazenadas em



Fig. 5. Categorias de pescados utilizados.

tabela para fins de comparação com os resultados obtidos pela técnica automatizada. Antes de tirar as fotografias, é adicionada, abaixo do peixe, uma régua de 30 cm da cor verde. Portanto, nas fotografias, além do peixe a ser medido, contém uma régua para servir como fator de escala. A imagem foi então adquirida por meio de câmera de smartfone, de 12 megapixels, resolução de 1280 por 960. No total, foram adquiridas 06 imagens, sendo uma de cada uma das categorias de pescado de peixes da bacia do Rio Madeira, Porto Velho, Rondônia, na Amazônia Ocidental, importantes tanto para o comércio quanto para subsistência: bagre, cascudo, jaraqui, pacu, piau e tucunaré. A Figura 5 mostra um exemplo de cada uma das categorias de pescado consideradas.

Como pode-se observar da Figura 5, as características de cor, textura e forma dos peixes são diferentes. Como apenas 06 imagens foram inicialmente adquiridas, sendo uma de cada categoria de pescado diferente, esse número foi insuficiente para efetuar o treinamento da rede neural de forma aceitável, de maneira a garantir que a rede Mask R-CNN conseguisse capturar todas as características essenciais dos peixes para reconhecê-los. Portanto, fez-se necessário obter mais imagens de outras bases. Para isso, foi utilizado o aplicativo online de gerenciamento e compartilhamento, *Flickr*, do qual foram retirados mais de 190 imagens de peixes diversos. Após um processo de filtragem, tendo como referência de seleção das amostras a iluminação e resolução, ficou-se com um montante total de 190 imagens. Destas 190 imagens, 150 imagens ficaram para o conjunto de treinamento e 40 imagens para o conjunto de teste.

Na segunda etapa é formada a máscara de cada imagem do conjunto de treinamento. Como se trata de uma rede supervisionada, é necessário, durante a fase de treinamento, fornecer o resultado procurado, que nesse caso seria a posição

do peixe na imagem. Como na metodologia proposta, é preciso somente medir o tamanho do peixe, a máscara consiste de um retângulo contendo o peixe, uma caixa delimitadora, como pode ser observado na Figura 3. O software utilizado para a criação das máscaras foi o *LabelImg* [10], a formação das máscaras é realizada manualmente, portanto fez-se necessário criar uma máscara para cada uma das 190 imagens. A Figura 4 demonstra como é executada a tarefa.

Durante a fase de treinamento foi utilizado somente uma classe para classificação das imagens, agrupando todas as categorias de pescado como uma única classe, peixe. Nesse primeiro estágio do projeto ainda não se tornou necessário efetuar a distinção de cada categoria de pescado, por isso foi tomada a decisão de agrupar as 06 categorias como uma única classe. Após o treinamento, o modelo treinado está pronto para utilização. É então selecionada cada uma das 06 imagens contendo o peixe e a régua. O próximo passo consiste em identificar a régua a partir da cor da mesma e calcular sua medida, por fim, a imagem é colocada no modelo treinado, que retorna a caixa delimitadora com a posição do peixe, e por meio do tamanho da régua e a resposta obtida do modelo treinado é calculado o tamanho do peixe.

IV. EXPERIMENTOS E RESULTADOS

Para a execução do método proposto neste artigo, foi utilizada a plataforma *Google Colab*, que permite implementar algoritmos na linguagem *Python*, linguagem escolhida para a implementação da técnica empregada, pelo navegador, disponibilizando até mesmo GPU de forma gratuita. Na maioria das vezes, a GPU disponibilizada era a Tesla K10.

Como métricas de avaliação dos resultados, foram utilizados o erro relativo (eq. 1) e o erro relativo médio (eq. 2).

TABLE I
RESULTADO ATINGIDO

Categoria de pescado	Tamanho Real	Tamanho Obtido	Diferença	Erro relativo (%)
Piau	25,5 cm	25,32 cm	0,18 cm	0,70
Jaraqui	24,0 cm	23,66 cm	0,34 cm	1,41
Tucunaré	30,5 cm	33,11 cm	2,61 cm	8,55
Bagre	33,5 cm	34,18 cm	0,68 cm	2,02
Cascudo	32,0 cm	32,12 cm	0,12 cm	0,375
Pacu	24,0 cm	24,12 cm	0,12 cm	0,5
Erro relativo médio				2,26

$$\text{erro relativo} = \frac{|v_A - v_E|}{v_E} \quad (1)$$

$$\text{erro relativo médio} = \sum_{i=1}^n \frac{CV_i}{n} \quad (2)$$

Na equação 1, erro relativo, v_A representa o valor aproximado do peixe, o valor medido, e v_E representa o seu valor exato. Para se obter a porcentagem do erro relativo é necessário somente multiplicar o seu resultado por 100. Na equação 2, erro relativo médio, CV é o valor do erro relativo e n a quantidade de valores.

O tamanho real do peixe, para ser utilizado como comparação com a medição obtida do método proposto, foi adquirido junto ao Laboratório de Ictiologia e Pesca da Universidade Federal de Rondônia, na Figura 6 é possível visualizar a resposta obtida da rede treinada na imagem do peixe Cascudo. Pode-se observar na Figura 6 que a caixa delimitadora prevista é um pouco maior que o peixe, o que acarreta na falha ao estimar o tamanho exato do peixe.

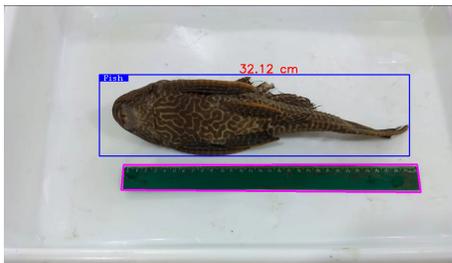


Fig. 6. Medição obtida para o peixe Cascudo.

Com os dados calculados, que podem ser observados na Tabela I, é possível chegar a resposta que o erro relativo médio é de aproximadamente 2,26%.

V. CONCLUSÃO

Neste trabalho foi explorado o uso de redes neurais convolucionais na tarefa de segmentação de peixes de água doce, usando a técnica Mask R-CNN. A partir dos resultados do experimento, pode-se observar que esta técnica apresentou resultados muito promissores para a medição do comprimento de peixes, já que o erro relativo médio ficou em torno de 2,26%.

A utilização do método proposto para a medição de peixes de maneira automática irá tornar bem mais veloz a coleta de dados e possibilitando a aceleração na obtenção desses dados, contribuindo em pesquisas envolvendo os mesmos.

Como trabalhos futuros, será adquirida uma base de dados maior, tanto em ambiente controlado (laboratório), como em ambiente real (no mercado de peixes, onde os biólogos fazem suas medições atualmente). Com relação à técnica usada, adaptá-la para segmentar somente o peixe, e não usar um retângulo como máscara, contendo a posição do peixe, a fim de possibilitar uma entrada menor de ruídos, pois assim a rede neural estaria manipulando apenas o peixe. Por último, treinar a rede para também detectar a régua seria uma forma de tirar a limitação da utilização da régua de uma determinada coloração, visto que a detecção por cor pode ocasionar imprecisões muito facilmente, dependendo do fator de luminosidade.

REFERÊNCIAS

- [1] C. Yu, X. Fan, H. Zhuhua, X. Xia, Y. Zhao, R. Li, and Y. Bai, "Segmentation and measurement scheme for fish morphological features based on mask r-cnn," *Information Processing in Agriculture*, January 2020.
- [2] R. E. X. Meante and C. R. d. C. Dória, "Caracterização da cadeia produtiva da piscicultura no estado de Rondônia: Desenvolvimento e fatores limitantes," *Revista de Administração e Negócios da Amazônia*, vol. 9, no. 4, pp. 164–181, 2018.
- [3] G. Wang, J.-N. Hwang, F. Wallace, and C. Rose, "Multi-scale fish segmentation refinement and missing shape recovery," *IEEE Access*, vol. 7, pp. 52 836–52 845, January 2019.
- [4] H. Yao, Q. Duan, D. Li, and J. Wang, "An improved k-means clustering algorithm for fish image segmentation," *Mathematical and Computer Modelling*, vol. 58, p. 790–798, August 2013.
- [5] T. Bouwmans, S. Javed, M. Sultana, and S. Jung, "Deep neural network concepts for background subtraction: A systematic review and comparative evaluation," *Neural Networks*, vol. 1, p. 40, May 2019.
- [6] R. Mieloszyk and P. Bhargava, "Convolutional neural networks: The possibilities are almost endless," *Current Problems in Diagnostic Radiology*, vol. 47, January 2018.
- [7] M. Sung, S. Yu, and Y. Girdhar, "Vision based real-time fish detection using convolutional neural network," in *OCEANS 2017 - Aberdeen*, 2017, pp. 1–6.
- [8] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *CoRR*, vol. abs/1511.08458, 2015.
- [9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [10] "Software LabelImg." [Online]. Available: <https://github.com/tzutalin/labelImg>
- [11] A. Tiwari, S. Srivastava, and M. Pant, "Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019," *Pattern Recognition Letters*, vol. 131, November 2019.