Automatic visual segmentation of fish in RAS environment using foundation models

Hilla Fred[1], Mogens Agerbo Krogh[2], Britt Bang Jensen[3], Paw Petersen[4], Laura Ruotsalainen[5], Tapio Kiuru[1], Matti Pastell[1*]

[1]Natural Resources Institute Finland (Luke), Helsinki, Finland

[2]Aarhus University, Aarhus, Denmark

[3] Norwegian Veterinary Institute, Oslo, Norway

[4]OxyGuard International AS, Farum, Denmark

[5]University of Helsinki, Helsinki, Finland

Abstract

Computer vision based methods in RAS farming are under development, as the automatic tracking of the animals using cameras would allow for cost-efficient monitoring and precise collection of data on fish behavior. Changes in behavior are an indicator of change in fish health or welfare. Accurate detection of fish in images is an important building block in a system that automatically tracks the movement of the animals.

The conditions in a RAS production setting are challenging for visual monitoring, so it is very useful to develop and test the methods on real-life data. Our video data was recorded on two commercial RAS farms, one farm cultivating Rainbow trout and one focusing on Atlantic salmon, using four RGB cameras positioned over the fish tanks. For the dataset, 2654 images were extracted at different times over several days' timespan to ensure a varied dataset, and the fish individuals were manually annotated in the images using bounding boxes.

Supervised solutions for detecting fish in RAS environment have been developed that work well in a research setting, but methods that do not require training data in the target environment have not been published. This is crucial, as the amount of manual work needed to create training datasets for supervised learning limits the use of machine learning models in real-life scenarios. We aim to leverage the recent advances in instance segmentation foundation models, namely the state-of-the-art benchmark performance reaching Segment Anything model (SAM). In instance segmentation, individual objects are identified in the image, resulting in segmentation masks that cover the pixels belonging to the corresponding objects.

We study how SAM performs when prompted with hand-annotated images of fish, and how it competes against automatic mask generation. Both experiments were conducted without any task-specific training of the model. Additionally, we looked into post-processing methods for filtering out bad quality masks using properties of the mask contours. We plan to use these masks to train a faster segmentation model to work in real-time and evaluate the performance.

Acknowledgements: The research was funded by Nordic Research and Innovation Programme for Sustainable Aquaculture under the NordForsk umbrella as part of the IntelliRAS project. Annotation work was conducted by Lea Ala-Ilomäki[1], Mille Stenholt Jensen[2] and Sofie Bælum Bugge[2].

*matti.pastell@luke.fi