

# A contrastive learning approach for individual re-identification in a wild fish population

Ørjan Langøy Olsen<sup>1</sup>, Tonje Knutsen Sjørdalen<sup>2</sup>, Morten Goodwin<sup>1</sup>, Ketil Malde<sup>3</sup>,  
Kristian Muri Knausgård<sup>\*4</sup>, and Kim Tallaksen Halvorsen<sup>3</sup>

<sup>1</sup>Centre for Artificial Intelligence Research, University of Agder, Norway

<sup>2</sup>Centre for Coastal Research, University of Agder, Norway

<sup>3</sup>Institute of Marine Research, Ecosystem Acoustics Group, Bergen, Norway

<sup>4</sup>Top Research Centre Mechatronics, University of Agder, Norway

## Abstract

In both terrestrial and marine ecology, physical tagging is a frequently used method to study population dynamics and behavior. However, such tagging techniques are increasingly being replaced by individual re-identification using image analysis.

This paper introduces a contrastive learning-based model for identifying individuals. The model uses the first parts of the Inception v3 network, supported by a projection head, and we use contrastive learning to find similar or dissimilar image pairs from a collection of uniform photographs. We apply this technique for corkwing wrasse, *Symphodus melops*, an ecologically and commercially important fish species. Photos are taken during repeated catches of the same individuals from a wild population, where the intervals between individual sightings might range from a few days to several years.

Our model achieves a one-shot accuracy of 0.35, a 5-shot accuracy of 0.56, and a 100-shot accuracy of 0.88, on our dataset.

## 1 Introduction

Physical tagging, using external or internal markings for individual identification, is a widely used method for monitoring terrestrial and aquatic animal populations. Information from resightings or recapture of the same individuals can be used

to estimate population size, survival and movement patterns. However, most tagging methods are costly, intrusive, and labor-intensive. To our benefit, many animals have natural markings or morphological features that are unique to individuals that could be used for photo-identification and replace the need for physical tags [22, 19]. However, for ecologists, working with fish may mean keeping track of hundreds or potentially thousands of individuals in a population, which makes manual photo-identification challenging, if not impossible. For this reason, fully- or semi-automatic tools for re-identification of individuals would be immensely useful for ecologists.

Re-identification (re-ID) is different from normal classification in that it is a few-shot learning problem. Few-shot problems are characterised by having few samples per class, but there may be a large or indefinite number of classes. One way to solve such problems is a technique called metric learning, where data is transformed into embeddings of a lower dimension, that clusters points from the same class together. Classification can then be performed on the embeddings. Metric learning approaches have been proved to work well for re-identification of animal species [18]. A crucial advantage with metric learning approaches is that the network does not need to be retrained to be able to add new classes.

Contrastive learning is a technique that can be used to solve few-shot problems. Contrastive learning compares data and identifies whether they are similar or dissimilar. A siamese network [1] is

---

\*Corresponding kristianmk@ieee.org

the most basic form and takes two inputs through the same network with the same shared weights and gets an embedding for both. During training, it tries to predict whether they are of the same class or not. A major advantage here is that it does not need to know which class an input belongs to, nor how many classes there are. Triplet networks [10] are an improvement to the siamese network with three inputs.

The goal of this work was to test the applicability of image based re-ID analysis for a commercially and ecologically important fish species, the corkwing wrasse (*Symphodus melops*). The image dataset consists of standardized photos of captures and recaptures of individuals in a wild population, where the time between individual sightings spans from days to several years. The first step is to detect a fish in an image with an object detector, followed by a re-identification method. With high enough precision, computer vision re-ID has the potential to replace physical tagging for individual identification and may be applied in monitoring of survival rates, growth, movement, and population size, key knowledge for sustainable management and conservation [18] [4].

## 2 Related works

Advancements in machine learning have produced powerful techniques for extracting ecologically important information from image and video data. For instance, machine learning have successfully been utilized to detect fish wounds [5], count and categorize organisms in digital photos and real-time video [14], [12], identify species, [6], and discover, and count creatures from digital images, [4], and even quantify their behaviour [3].

Some work on the topic of re-identification of fish has been conducted, but work on wild teleost fish are lacking. Bruslund Haurum et al. [2] achieved an mAP of 99% on Zebrafish using metric learning with 15 samples per class of 6 classes. Meidell and Sjøblom [16] reports a true positive rate of 96% on 225 thousand images of salmon divided between 715 individuals. Li et al. [13] achieved an accuracy of 92% using 3412 images of 10 individuals using their novel FFRNet network. These studies have in common that they were carried out in captivity and are not using temporally indepen-

dent observations. In other words, the individuals did not change morphology through growth, maturation, senescence, or similar biological processes. Moskvayak et al. [17] used a metric learning approach on a dataset of 1730 images of 120 manta ray individuals and achieved an accuracy@1 of 62% and an accuracy@10 of 97%.

## 3 Method

### 3.1 Data collection

The study species, *S. melops*, is a commercially and ecologically important species in coastal ecosystems in the Northeastern Atlantic [7]. This species have two distinct male morphs, colourful large males that build nest and care for the eggs, and smaller sneaker males, with a more brown coloration resembling the female morphology (brown and gray) [21]. The dataset was collected in Austevoll, western Norway, 2018-2021, by catching corkwing wrasse by fyke nets left in the sea overnight and marking all captured individuals with uniquely coded passive integrated transponder (PIT) tags (11 mm tags, RFID Solutions). The tags were implanted in the abdominal cavity of the fish, see full sampling description in [8] and [9].

This method enabled us to collect independent observations of each individual across time and for the dataset to encompass changes in the fish's morphology. At each capture, a few images were taken of the fish on both sides and the images were tagged with an id based on the RFID. The images are captured with the dorsal side of the fish facing up. After some filtration, a dataset that could be used for the task was compiled. The final dataset consists of 2113 images from 513 unique individuals. As an added statistic, the mean between the first and last capture-date of all the individuals is 230 days. Samples from the dataset can be seen in Figure 1.

### 3.2 Individual re-identification

The re-identification system consists of a pipeline of different components, as illustrated in Figure 2. The components fall into two categories, a preprocessing part and a re-identification part. As part of a preprocessing step in the pipeline, the system takes an image as input and feeds it to a object de-

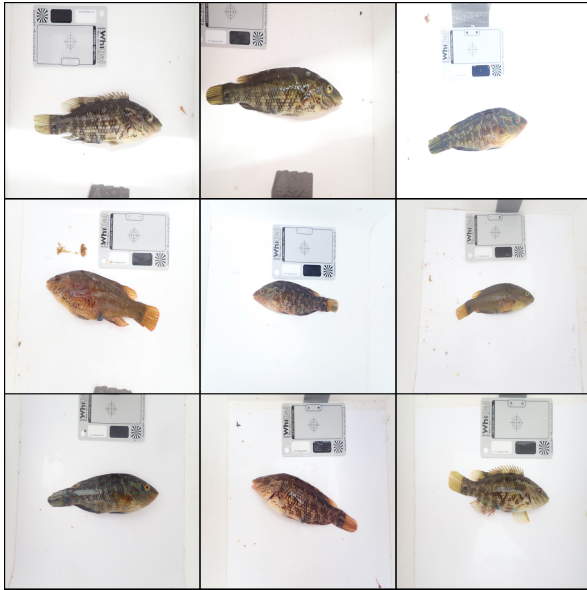


Figure 1: Samples from the unprocessed dataset.

tection network to get an image crop, only containing the fish in the frame. Then a different network, the direction component, classifies whether the fish is facing right or left and passes this as metadata. For the re-identification part, the preprocessed data is fed to a contrastive learning network that learns to group embeddings for the same individual together and different apart. Classification can then be performed on the embeddings. By storing the embeddings of all previously observed individuals, re-identification can be achieved by nearest neighbor methods.

The object detector uses YOLOv5 [11] with an image size of 416x416, a batch size of 32 and is trained for 50 epochs. During training, the network was provided with manually annotated bounding boxes enclosing the fish.

The direction network is an Inception v3 [20] model with all its weights frozen. A global average pooling layer, a ReLU activated layer with 32 neurons, and a sigmoid activated output have been appended to the network. The dataset used for the training is the images cropped to only contain the head. The dataset is manually annotated with the direction.

The embedding network consists of a CNN model with a projection head. Its constituent parts were

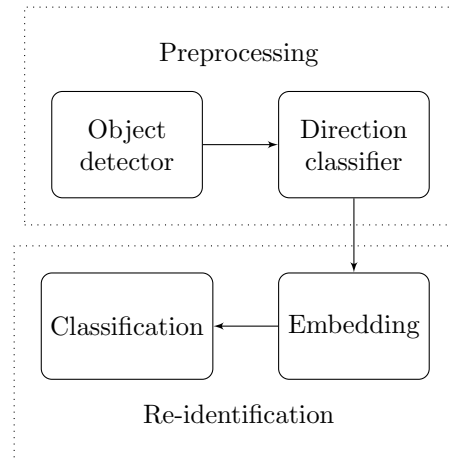


Figure 2: The network pipeline takes an image of a fish as input and outputs the id of the individual.

found experimentally. The CNN model is an Inception v3 model pre-trained on ImageNet, with the layers after the fourth concatenation layer (layer 46, or 132 if counting activation layers) removed. Appended at the end is a 2D global average pooling layer and a 128-dimensional linear projection that is normalized to the unit hypersphere. The network diagram is shown in Figure 3. The input size of the network is 224, and the images are resized accordingly before being fed into it. The network utilizes letter-boxing to maintain aspect-ratio. For the training of the embedding network, the dataset is split into a training and test set with a test set fraction of 0.3.

The training of the embedding network uses gradual unfreezing. The first 100 epochs have the layers before layer 29 frozen and a learning rate of 0.001, and the next 100 epochs have the layers before layer 18 frozen and a learning rate of 0.0001 for a total of 200 epochs. Layer 29 and layer 18 were selected because they are concatenation bottlenecks in the network architecture (green nodes in Figure 3). The loss function is online hard triplet mining with a margin of 1.0. Hard triplet mining is a technique where loss is only backpropagated for triplets where the negative is closer to the anchor than the positive. Thus, the use of online mining circumvents the need for three identical networks with shared weights. The training samples are randomly applied with image augmentations. A number between -20 and 20 is added to the hue

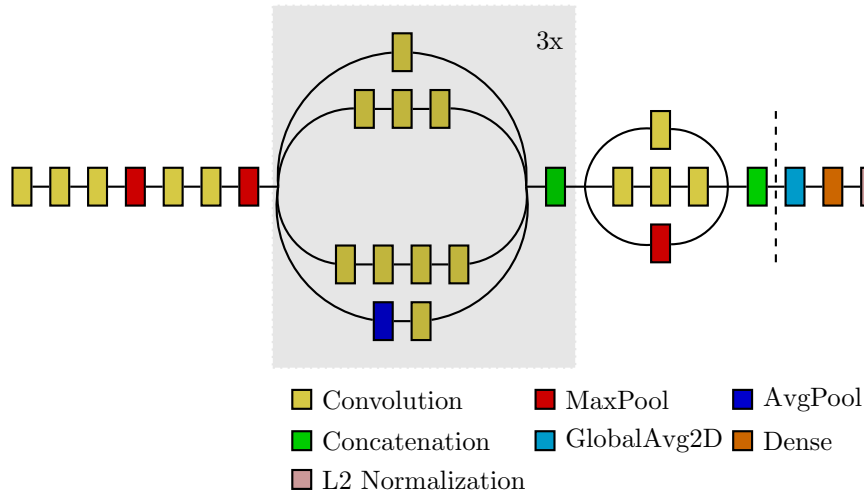


Figure 3: The embedding network utilizes the first part of Inception v3 [20] with a custom projection head. The dashed line marks where the Inception part ends and the custom part starts. The grey part is repeated three times.

and saturation. The image is rotated by a fraction between 0 and 0.1 in either direction, and a scale transformation between 0 and 0.1 is applied. The batch size used is 32.

Classification, and by extension re-identification, is done using a nearest neighbor approach. And in this case it is useful to define the training set as the support set and the test set as the query set. Nearest neighbor classification is non-parametric and does not need to be trained through optimization. The training step is simply to feed the support images through the embedding network and store the associated embeddings for the inference step. To classify an image, a query image is fed through the embedding network and then simply select the class of the nearest point of the query embedding to the support set embeddings. Source code for our implementation is available at GitHub<sup>1</sup>.

### 3.3 Method for experiments

The *Symphodus melops* have a distinct high-contrast pattern in the head region (particularly on the operculum). For this reason, it would be useful to explore whether the network performs better on head crops than on crops of the whole body. The experiment is performed by training and evaluating

the embedding network on images that are cropped to either part.

The system can also treat each side of the fish as different classes, and thus valuable information can be gained by doing inference on both, and then combining the results in an ensemble classifier. For this experiment, the dataset is split up into a left-sided section and a right-sided section such that there is a pairwise correspondance between the images. Two models are trained, where one is only for left-sided images and the other is only for right-sided images. The embeddings in the support set is sorted by the distance to the query image for each side. The predicted class is then the class which appears first when both sorted collections are taken into account.

An experiment to evaluate how well the system is able to distinguish between a re-sighted individual and an individual that has never been seen before was also conducted. A query embedding is considered a new individual if its distance is greater than a certain distance away from any support embedding. The query set was split into a test set and a validation set. A grid search was used to find a good distance threshold by maximizing the F1 score when evaluating the test set. The validation dataset for this experiment contains 317 samples.

<sup>1</sup><https://github.com/orilan93/SiameseFish>

## 4 Results

The metrics we use are accuracy@1, accuracy@5, and mAP@5. Accuracy@1 shows the correctness of the highest ranked category, i.e., the percentages of the highest predicted class are equal to the true class. Accuracy@5 shows the correctness of the five highest ranked categories, i.e., how many of the five highest classes contain the true class. mAP@5 similarly shows the precision of the five highest ranked categories, i.e., how many of the true categories are among the five highest ranked categories.

### 4.1 Re-identification

The re-identification system was evaluated against both the head and body crop datasets. Table 1 presents results from accuracy@1 and accuracy@5 and shows that the model performs best on the head crops. Figure 4 shows how the model performs as the number of accumulated attempts increase. This approach is essential in practice because, instead of having an unsorted catalog of images to go through, a professional biologist can go through a sorted catalog and expect to find the correct individual after inspecting the  $k$  most promising images sorted based on the distance measure. The larger  $k$  the higher accuracy, and as the number of attempts are approaching the number of images in the support set, the accuracy is approaching 100%.

Table 1: Results for re-identification on head and body crops.

Type	Accuracy@1	Accuracy@5	mAP@5
Head	0.3534	0.5647	0.4227
Body	0.2043	0.3892	0.2690

Table 2 shows four random image samples from the dataset, together with the image the trained model predicts is the same individual and the ground truth. The classification rank and the distance in the embedding space are also shown.

To gain insight into what the model focuses on when making its inferences, we present some test set samples and the accompanying SHAP plot [15] in Figure 5. The colored area shows that the model is indeed picking up on the pattern of the fish.

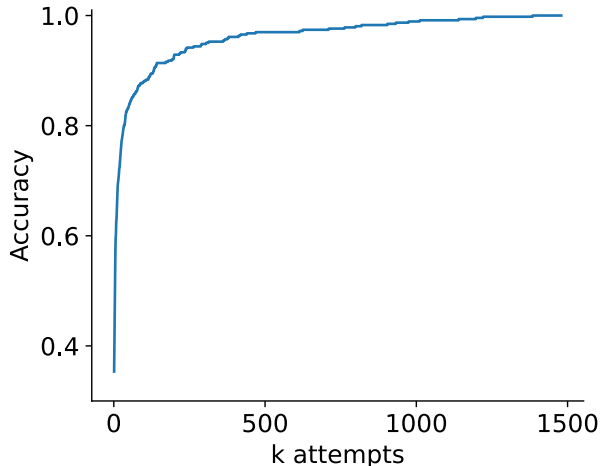


Figure 4: The number of accumulated attempts ( $k$ ) needed to attain a certain accuracy (accuracy@ $k$ ).

### 4.2 Ensemble classifier

This experiment shows the results of training a new model for each side of the fish and then combining their respective classifications. Table 3 shows that this strategy can significantly increase performance. Note that the direction component, that is required for the ensemble classifier, yielded an accuracy@1 of 99.38% on the validation set using the head cropped dataset.

### 4.3 New observations

As our previous experiments have shown, re-identification works relatively well. We aim at using this model for distinguishing new individuals from earlier observed individuals. To identify new individuals with the model, an embedding distance threshold needs to be decided. Note that this relates to the distance metric in Table 2. Using grid search, we found a threshold of 0.820 to yield the best performance score on the validation set. The system predicted 95 individuals as new sightings and got a 62.78% accuracy@1 at this task.

## 5 Discussion and conclusion

Our experiments, summarized in Table 4, indicate that the system performs better on the head crops of the fish than on the whole body. This is likely

Table 2: The retrieval rank and euclidean distance between the embedding of a query image and a correct image.






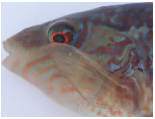

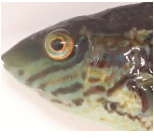
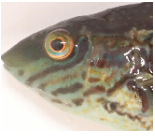



Query	Predicted	Ground truth	Rank	Distance
			1	0.65
			217	1.27
			1	0.61
			1012	1.46

Table 3: Ensemble classifier results.

Type	Accuracy@1	Accuracy@5	mAP@5
Left	0.3568	0.5463	0.4243
Right	0.4097	0.5595	0.4623
As pair	0.5286	0.7533	0.6140

Table 4: Summary of results.

Experiment	Result	Metric
Re-identification	0.3534	Accuracy@1
Object detector	0.9951	mAP@0.5
Direction classifier	0.9937	Accuracy@1
Ensemble classifier	0.5286	Accuracy@1
New observations	0.6278	Accuracy@1

because the pattern on the head is most distinct and thus an important feature, and this will appear at a higher resolution for the algorithm when resizing for the network input size. However, the drawback here is that the network is exposed to less information available in the data.

By utilizing the existing system in a new way by training separate models for each side of the fish, one can make an ensemble classifier. This method was tested and gained a considerable improvement from 35% to 53% accuracy. This shows how important it is to use all the information available to make good predictions.

The accuracy of this system is not high enough for a fully automated system with humans out-of-the-loop, which is required to replace the need for physical tags in ecological studies. However, we believe that continued collection of data can produce a dataset that is more temporally balanced to enable the model to account for the growth and ageing of the individuals.

Automatization can produce great benefits and is increasingly being adopted by many industries, and the field of ecology should be no different. A successful Re-ID algorithm with high precision can provide a new method with improved fish welfare, while also being cheaper (only a camera needed) and potentially more accurate (no tag loss). In the future, we envision that re-ID can be applied

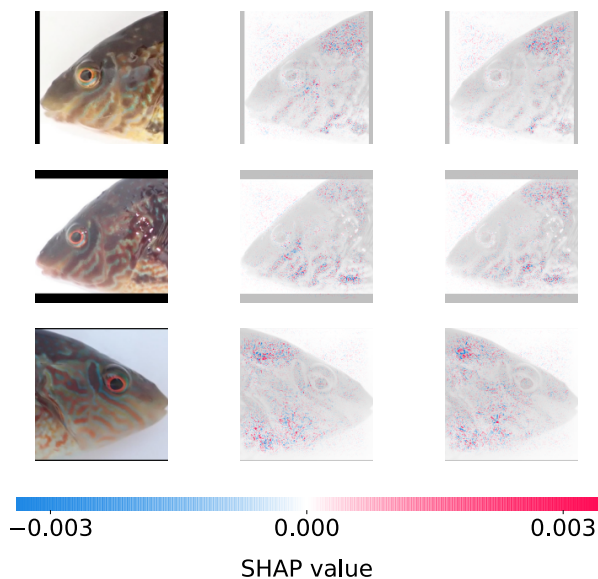


Figure 5: SHAP plot showing which areas in the images that are most influential for the decisions of the model.

directly on live streams from under-water video cameras, removing the need for capture and handling fish altogether. This would be a revolutionary method that can drastically change how we can collect key information for sustainable conservation and management of fish and other animals.

## Acknowledgements

We thank Torkel Larsen, Anne Berit Skiftesvik, Ovin Holm, Ylva Vik, Nicolai Aasen, Ben Ellis, Vegard Omestad Berntsen, and Steve Shema for assistance in collecting the photos and capture-recapture data in the field. This study received funding from Centre for Artificial Intelligence Research (CAIR), Centre for Coastal Research (CCR), and Top Research Centre Mechanics (TRCM) at University of Agder, the Institute of Marine Research (project 15638-01), and the Research Council of Norway (CoastVision, project number 325862, and CreateView, project number 309784).

## References

- [1] J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Säckinger, and R. Shah. Signature verification using a “siamese” time delay neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(04):669–688, 1993.
- [2] J. Bruslund Haurum, A. Karpova, M. Pedersen, S. Hein Bengtson, and T. B. Moeslund. Re-identification of zebrafish using metric learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 1–11, 2020. doi: 10.1109/WACVW50321.2020.9096922.
- [3] E. M. Ditria, E. L. Jinks, and R. M. Connolly. Automating the analysis of fish grazing behaviour from videos using image classification and optical flow. *Animal Behaviour*, 177:31–37, jul 2021. ISSN 00033472. doi: 10.1016/j.anbehav.2021.04.018.
- [4] M. Goodwin, K. T. Halvorsen, L. Jiao, K. M. Knausgård, A. H. Martin, M. Moyano, R. A. Oomen, J. H. Rasmussen, T. K. Sjørdalen, and S. H. Thorbjørnsen. Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. *ICES Journal of Marine Science*, 01 2022. ISSN 1054-3139. doi: 10.1093/icesjms/fsab255.
- [5] A. Gupta, E. Bringsdal, N. Salbuvik, K. M. Knausgård, and M. Goodwin. An accurate convolutional neural networks approach to wound detection for farmed salmon. In *International Conference on Engineering Applications of Neural Networks*, pages 139–149. Springer, 2022. doi: 10.1007/978-3-031-08223-8\_12.
- [6] A. Gupta, E. S. Kalhagen, Ø. L. Olsen, and M. Goodwin. Hierarchical object detection applied to fish species: Hierarchical object detection of fish species. *Nordic Machine Intelligence*, 2(1), 2022. doi: 10.5617/nmi.9452.
- [7] K. T. Halvorsen, T. K. Sjørdalen, C. Durif, H. Knutsen, E. M. Olsen, A. B. Skiftesvik, T. E. Rustand, R. M. Bjelland, and L. A. Vøllestad. Male-biased sexual size dimorphism

- in the nest building corkscrew wrasse (*Symphodus melops*): Implications for a size regulated fishery. *ICES Journal of Marine Science*, 73 (10):2586–2594, nov 2016. ISSN 10959289. doi: 10.1093/icesjms/fsw135.
- [8] K. T. Halvorsen, T. K. Sjørdalen, L. A. Vøllestad, A. B. Skiftesvik, S. H. Espeland, and E. M. Olsen. Sex- and size-selective harvesting of corkscrew wrasse (*Symphodus melops*)—a cleaner fish used in salmonid aquaculture. *ICES Journal of Marine Science*, 74 (3):660–669, mar 2017. ISSN 1054-3139. doi: 10.1093/icesjms/fsw221.
- [9] K. T. Halvorsen, T. Larsen, H. I. Browman, C. Durif, N. Aasen, L. A. Vøllestad, A. Cresci, T. K. Sjørdalen, R. M. Bjelland, and A. B. Skiftesvik. Movement patterns of temperate wrasses ( Labridae ) within a small marine protected area. *Journal of Fish Biology*, 99 (4):1513–1518, jul 2021. ISSN 0022-1112. doi: 10.1111/jfb.14825.
- [10] E. Hoffer and N. Ailon. Deep metric learning using triplet network. In *International workshop on similarity-based pattern recognition*, pages 84–92. Springer, 2015. doi: 10.1007/978-3-319-24261-3\_7.
- [11] G. Jocher, A. Stoken, A. Chaurasia, J. Borovec, et al. ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support, Oct. 2021. URL <https://doi.org/10.5281/zenodo.5563715>.
- [12] K. M. Knausgård, A. Wiklund, T. K. Sjørdalen, K. T. Halvorsen, A. R. Kleiven, L. Jiao, and M. Goodwin. Temperate fish detection and classification: a deep learning based approach. *Applied Intelligence*, 52(6):6988–7001, 2022. doi: 10.1007/s10489-020-02154-9.
- [13] D. Li, H. Su, K. Jiang, D. Liu, and X. Duan. Fish face identification based on rotated object detection: Dataset and exploration. *Fishes*, 7 (5):219, 2022. doi: 10.3390/fishes7050219.
- [14] V. Lopez-Vazquez, J. M. Lopez-Guede, S. Marini, E. Fanelli, E. Johnsen, and J. Aguzzi. Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. *Sensors*, 20(3):726, 2020. doi: 10.3390/s20030726.
- [15] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
- [16] E. Meidell and E. S. Sjøblom. Fishnet: A unified embedding for salmon recognition. Master’s thesis, NTNU, 2019.
- [17] O. Moskvayak, F. Maire, F. Dayoub, A. O. Armstrong, and M. Baktashmotlagh. Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. In *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2019. doi: 10.1109/DICTA52665.2021.9647359.
- [18] S. Schneider, G. W. Taylor, S. Linqvist, and S. C. Kremer. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, 10(4):461–470, 2019. doi: 10.1111/2041-210X.13133.
- [19] S. Schneider, G. W. Taylor, and S. C. Kremer. Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision workshops*, pages 44–52, 2020.
- [20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [21] I. Uglem, G. Rosenqvist, and H. S. Wasslavik. Phenotypic variation between dimorphic males in corkscrew wrasse. *Journal of Fish Biology*, 57(1):1–14, jul 2000. ISSN 00221112. doi: 10.1006/jfb.2000.1283.
- [22] B. G. Weinstein. A computer vision for animal ecology. *Journal of Animal Ecology*, 87(3): 533–545, 2018. doi: 10.1111/1365-2656.12780.