Croatian Fish Dataset: Fine-grained classification of fish species in their natural habitat

Jonas Jaeger[1,2]
http://www.hs-fulda.de/index.php?id=11123
Marcel Simon[2]
http://www.inf-cv.uni-jena.de/simon.html
Joachim Denzler[2]
http://www.inf-cv.uni-jena.de/denzler
Viviane Wolff[1]
http://www.hs-fulda.de/index.php?id=5780
Klaus Fricke-Neuderth[1]
http://www.hs-fulda.de/index.php?id=12276
Claudia Kruschel[3]
ckrusche@unizd.hr

[1] Fulda University of Applied Sciences
Department of Electrical Engineering and Information Technology
D-36037 Fulda, Germany

[2] Friedrich Schiller University Jena
Computer Vision Group
D-07737 Jena, Germany

[3] University of Zadar
Department of Maritime Science
23000 ZADAR, Croatia

Abstract

This paper presents a new dataset for fine-grained visual classification (FGVC) of fish species in their natural environment. It contains 794 images of 12 different fish species collected at the Adriatic sea in Croatia. All images show fishes in real live situations, recorded by high definition cameras. Remote and diver-based videography is used by a growing number of marine researchers to understand spatial and temporal variability of habitats and species. The required large numbers of independent observations necessitate the development of computer vision tools for an automated processing of high volumes of videos featuring high fish richness and density. As baseline experiment, we are using CNN features [1] and a linear SVM classifier and achieve an accuracy of 66.78% on our dataset.

1 Introduction

The analysis of fish-communities is important to understand the influence of natural and anthropogenic effects like habitat loss, pollution, overfishing and climate change to marine life. Such knowledge is needed to develop effective protection and management tools for fish as they are important resources for the human population and important players in the global ocean system.

As an alternative to destructive and extractive methods, the use and development of remote and diver based underwater video technology, e.g. for recording fish in their natural habitats, have become a fast developing field. A base task in the analysis of the resulting

Pages 6.1-6.7
https://dx.doi.org/10.5244/C.29.MVAB.6

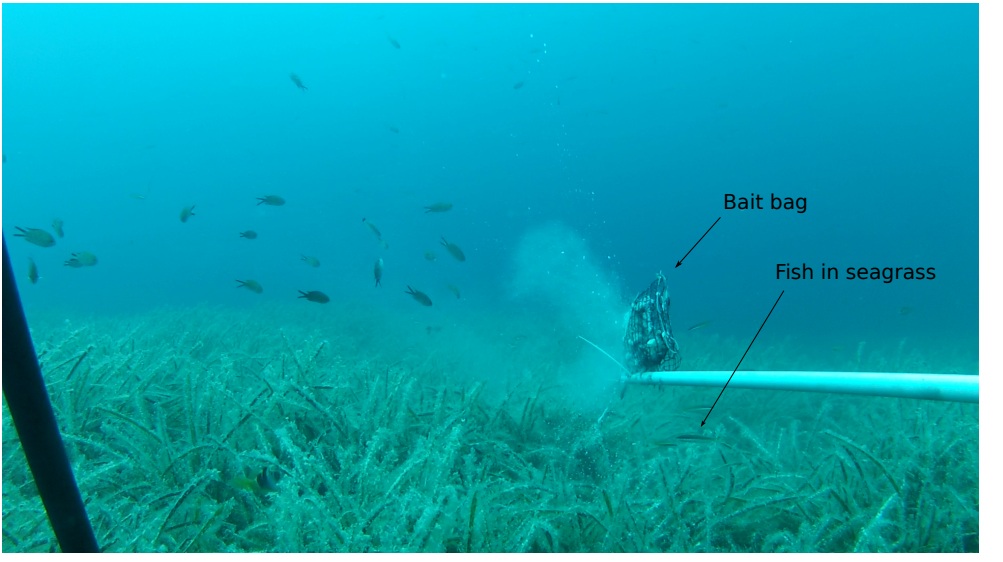

Figure 1: Sample video frame showing a common seagrass environment. Most fishes are swarming around the bait bag, while others are hiding in the seagrass.

large volumes of video is the identification, counting and size measurement of individual fish. At present these tasks are done by experts, which limits the throughput of video analysis and is time and cost inefficient. Automated methods with rapid video processing would allow for higher numbers of analysed samples in the field. This would result in larger data sets supporting a higher statistical power to recognize meaningful and significant variability in various spatial and temporal scales. For the development of the required computer vision methods, datasets provide a way to compare the accuracy and reliability of different approaches with respect to this specialised application.

We introduce a new dataset for fine-grained visual classification consisting of 12 fish species in their unconstrained natural habitat. The images present fishes in front of a challenging seagrass background (see Figure 1) and in rocky reefs. At the same time, most species show a high visual similarity. The dataset is publicly available at: `http://www.inf-cv.uni-jena.de/fine_grained_recognition.html#datasets`

Section 2 shows the relation of our dataset to other fine-grained recognition tasks. In section 3 the dataset is described in detail. Section 4 presents first experimental results and compares our dataset with the FishCLEF2015 [3] set by applying the same baseline method. The last part of this paper gives a short summary and describes the future work.

2 Related work

Research efforts in the field of fine-grained visual classification have lead to a compilation of many datasets consisting of birds [10], dogs [4], aircrafts [7], moths and butterflies [8]. Numerous specialised methods for these applications were developed and an astonishing increase in classification accuracy can be observed in the recent years. To evaluate the accuracy of these approaches for our application, fish specific datasets are required. A dataset which

provides fish in their real-life environment is FishCLEF2015 [3]. All videos were taken in coral reefs in Taiwan and manually annotated by experts with bounding boxes and species names. The training set consists of 20 videos and more than 20000 sample images of 15 fish species. The test set includes 73 videos.

While the dataset of [3] has little background clutter, our dataset includes images with strong clutter due to seagrass and the epiphytes covering the seagrass blades. Since some fish species look very similar to these background patterns, detection is a difficult task even for humans. Compared to the other dataset, the resolution of our videos is much higher. Another aspect is that our camera setup aims to record a wide underwater scene with many fishes at the same time(see Figure 1), while the coral reef setup in [3] has a narrow field of view. The high resolution of this broad scene makes it possible to record fishes in a wide range of distances to the camera, which has a strong influence to light condition and the sizes of the cropped fish images (pixels per fish). In comparison to the dataset of [3], our dataset also distinguishes a different set of fish species with high visual similarity between many of them. Hence, our dataset is also suitable as a benchmark for fish classification in unrestricted natural environments as in Huang *et al*. [2] or more general approaches for fine-grained recognition as presented by Simon and Rodner [9].

3 Dataset

We present a dataset for fine-grained classification containing 794 images of 12 fish species. The source of the dataset are high resolution videos with 1280×960 and 1920×1080 pixel and a frame rate of 25 frames per second. Each video was recorded either by a stationary "Baited Remote Underwater Video" (BRUV [6]) or a moving "Diver Operated Video" system (DOV), at the Adriatic sea in Croatia. A typical video frame, recorded by these systems, is shown in Figure 1. It captures fishes swimming near a bait bag within a seagrass environment. They are moving in and out of the structurally complex seagrass and are very difficult to detect even for humans.

At the moment, 10 sequences recorded at 3 different locations are used. These videos contain the 12 commonly observed fish species. Out of each video one frame per second was extracted. A selection of frames was annotated by an expert in fish identification (CK). The fishes were marked with a bounding box and the corresponding species name. Each image patch described by a bounding box was extracted and saved as a single dataset image. It is possible that in some cases multiple views of the same fish are within the dataset, because we did not use tracking for the selection of fishes. But it is unlikely since the annotator tried to select every fish only once.

Figure 2 shows example images of the 12 fish classes of the dataset. The animals are recorded in different poses and sizes. The size depends mainly on the distance between fish and camera as well as the size of the fish itself. The source videos were taken under varying light conditions. The recordings show two habitat types, sediment consolidated by seagrass and heterogeneous combinations of rocky and sandy bottoms. Table 1 shows the number of images per species. Due to the varying frequency of appearance of each species in our videos, the number of images varies between classes. For example, the most frequent species *Diplodus vulgaris* has 110 images and the least (*Sarpa salpa*) is presented 17 times.

The dataset is publicly available at: http://www.inf-cv.uni-jena.de/fine_grained_recognition.html#datasets

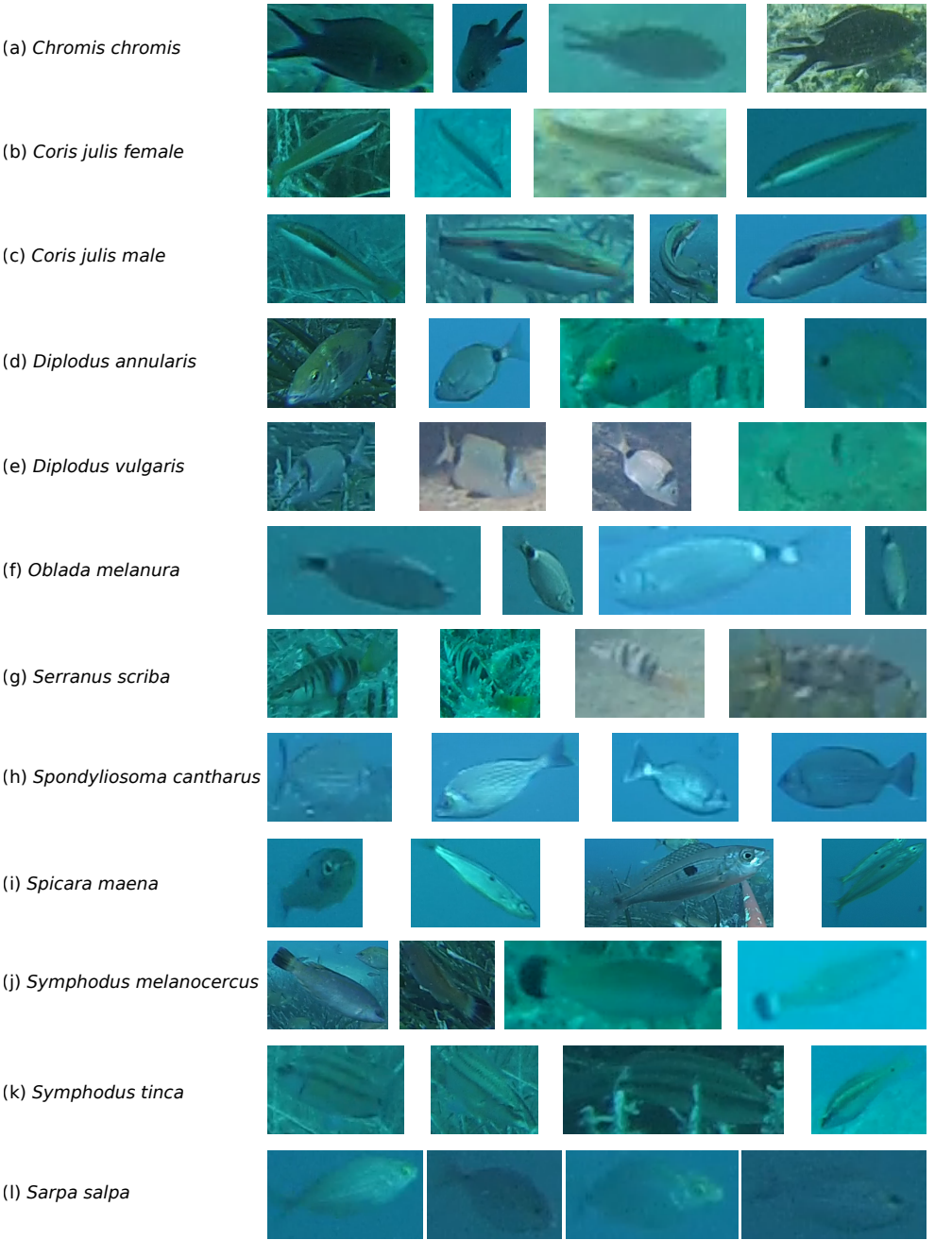

Figure 2: Resized example images of the 12 fish species in the dataset.

Species	Number of images
(a) Chromis chromis	106
(b) Coris julis female	57
(c) Coris julis male	57
(d) Diplodus annularis	94
(e) Diplodus vulgaris	111
(f) Oblada melanura	57
(g) Serranus scriba	56
(h) Spondyliosoma cantharus	51
(i) Spicara maena	49
(j) Symphodus melanocercus	105
(k) Symphodus tinca	34
(l) Sarpa salpa	17
Total	794

Table 1: Number of images per species.

4 Experiments

4.1 Evaluation protocol

We propose a random selection of 10 images per species to train the classifier. This selection is repeated 10 times and the mean accuracy over all splits is used as performance measure. Since our dataset is unbalanced in terms of the total number of images per species (see Table 1) the *mean average precision (mAP)* is used for the evaluation of each split.

4.2 Baseline

As baseline and for first experimental results on our dataset, we used intermediate activations of a CNN pre-trained on ImageNet as features and a linear SVM for classification. The CNN features were computed with the DeCAF framework [10], using the activations of the 7^{th} hidden layer. This deep convolutional neural network uses the architecture proposed in [5] by Krizhevsky *et al*. The SVM was trained in a one-vs-all manner.

With this pipeline we achieved an accuracy of 66.78% on our dataset. Figure 3 shows the corresponding averaged confusion matrix over all 10 splits of our experiment. The ground-truth classes correspond to rows and the predicted classes to columns. The matrix elements are stated in percent. We further applied this baseline method to the 20000 sample images of [3] and obtained a *mAP* of 82.21% by using the above mentioned evaluation protocol. This indicates that our fish dataset is more challenging, most likely due to the higher visual similarity of fish species and the strong background clutter.

4.3 Conclusion

We introduced a new real life fish dataset for fine-grained classification tasks. The dataset contains 794 images of 12 fish species in their unconstrained natural habitat. The fishes are presented in seagrass meadows and rocky reefs. The images were extracted from high resolution underwater videos used by marine ecologists to explore fish communities and observe biodiversity of marine life.

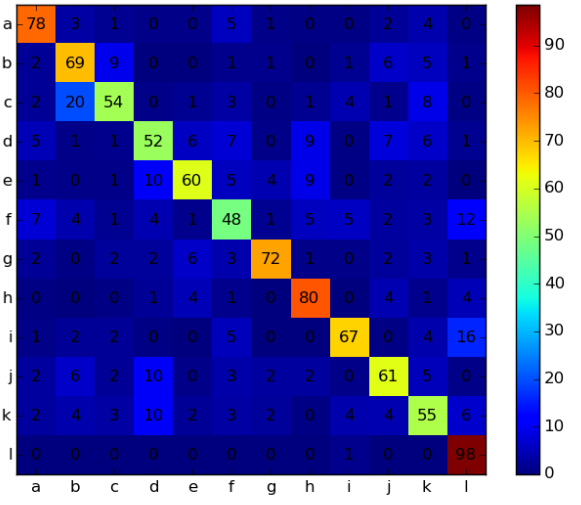

Figure 3: Averaged confusion matrix for the experiments on our dataset. The alphabetic labels are corresponding to the fish class labels in Figure 2.

First experiments obtained an accuracy of 66.78% on the presented dataset, by using global CNN features in combination with a linear SVM classifier. We also demonstrate that our dataset is a more challenging fine-grained recognition task than the FishCLEF2015 dataset [3], where an accuracy of 82.21% is achieved by applying the same baseline and evaluation protocol.

In the future, we plan to expand the presented dataset and to create a new one for detection, tracking and recognition of fishes, which maps the whole process of high resolution fish video examination. This will lead to further challenging computer vision tasks, because of the high similarity of the fishes to their environment and the large number of animals visible at the same time.

References

[1] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *CoRR*, abs/1310.1531, 2013.

[2] Phoenix X. Huang, Bastiaan J. Boom, and Robert B. Fisher. Hierarchical classification with reject option for live fish recognition. *Machine Vision and Applications*, 26(1): 89–102, January 2015.

[3] Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Robert Planque, Andreas Rauber, Robert Fisher, and Henning Müller. Lifeclef 2014: Multimedia life species identification challenges. In *Information Access*

Evaluation. Multilinguality, Multimodality, and Interaction, volume 8685 of *Lecture Notes in Computer Science*. Springer International Publishing, 2015.

[4] Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, and Li Fei-Fei. Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105. 2012.

[6] T. J. Langlois, E. S. Harvey, B. Fitzpatrick, J. J. Meeuwig, G. Shedrawi, and D. L. Watson. Cost-efficient sampling of fish assemblages: comparison of baited video stations and diver video transects. *Aquatic Biology*, 9(2):155–168, 2010.

[7] S. Maji, J. Kannala, E. Rahtu, M. Blaschko, and A. Vedaldi. Fine-grained visual classification of aircraft. Technical report, 2013.

[8] Erik Rodner, Marcel Simon, Gunnar Brehm, Stephanie Pietsch, J. Wolfgang Wägele, and Joachim Denzler. Fine-grained recognition datasets for biodiversity analysis. In *CVPR Workshop on Fine-grained Visual Classification (CVPR-WS)*, 2015.

[9] Marcel Simon and Erik Rodner. Neural activation constellations: Unsupervised part model discovery with convolutional networks. *CoRR*, abs/1504.08289, 2015.

[10] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.