FishNet: A Large-scale Dataset and Benchmark for Fish Recognition, Detection, and Functional Trait Prediction

Faizan Farooq Khan*,† Xiang Li* Andrew J. Temple Mohamed Elhoseiny
*equal contribution †project lead
King Abdullah University of Science and Technology (KAUST)
{faizan.khan, xiang.li.1, andrew.temple, mohamed.elhoseiny}@kaust.edu.sa

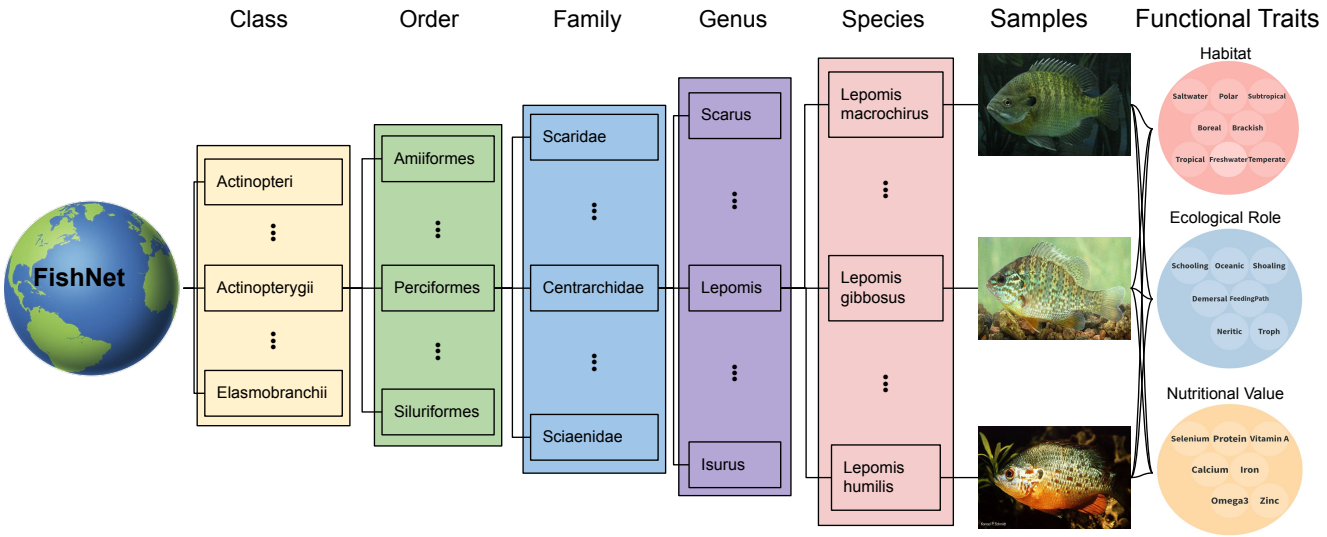

Figure 1: We present FishNet, a comprehensive benchmark for large-scale aquatic species recognition, detection, and functional trait identification. Our benchmark dataset is based on an aquatic biological taxonomy, consisting of 8 taxonomic classes, 83 orders, 463 families, 3,826 genera, 17,357 species, and 94,532 images. The dataset also includes bounding box annotations for fish detection. Additionally, the dataset encompasses 22 traits, grouped into three categories: habitat, ecological rule, and nutritional value. These traits facilitate the identification of the ecological roles of aquatic species and their interactions with other species.

Abstract

Aquatic species are essential components of the world's ecosystem, and the preservation of aquatic biodiversity is crucial for maintaining proper ecosystem functioning. Unfortunately, increasing anthropogenic pressures such as overfishing, climate change, and coastal development pose significant threats to aquatic biodiversity. To address this challenge, it is necessary to design an automatic aquatic species monitoring systems that can help researchers and policymakers better understand changes in aquatic ecosystems and take appropriate actions to preserve biodiversity. However, the development of such systems is impeded by a lack of large-scale diverse aquatic species datasets. Existing aquatic species recognition datasets generally have a limited number of species, nor do they provide functional trait data, and so have only narrow potential for application. To address the need for generalized systems that can recognize, locate, and predict a wide array of species and their functional traits, we present FishNet, a large-scale diverse dataset containing 94,532 meticulously organized images from 17,357 aquatic species, organized according to aquatic biological taxonomy (order, family, genus, and species). We further build three benchmarks, i.e., fish classification, fish detection, and functional trait prediction, inspired by ecological research needs, to facilitate the development of aquatic species recognition systems, and promote further research in the field of aquatic ecology. Our FishNet dataset has the potential to encourage the development of more accurate and effective tools for the monitoring and protection of aquatic ecosystems, and hence

take effective action toward the conservation of our planet's aquatic biodiversity. Our dataset and code will be released at https://fishnet-2023.github.io/.

1. Introduction

Aquatic biodiversity is of paramount importance in safeguarding the structure, stability, and overall health of aquatic ecosystems. Nonetheless, in recent decades, the escalating anthropogenic pressures from human activities, including fisheries, climate change, and coastal development [22, 24, 29], have made the conservation of aquatic biodiversity increasingly difficult and increasingly prominent in public attention. A key step in aquatic biodiversity protection is the ongoing monitoring, which calls for highly efficient species recognition and functional trait identification [36, 39, 42]. However, this process generally calls for a high level of expert knowledge due to complicated species taxonomy, which makes it time- and labor-consuming.

In recent years, deep learning methods have made significant breakthroughs in various computer vision tasks, presenting a promising solution for automatic and efficient species recognition. However, it is well-known that the accuracy of these AI-based models is heavily reliant on the scale and diversity of the training datasets. In the context of safeguarding aquatic biodiversity, there is an urgent need to develop large-scale and diverse datasets to facilitate AI-based aquatic species recognition systems. Several previous works have focused on building such datasets for fish recognition. For example, Fish4Knowledge [15] collected 27,370 fish images from 23 distinct species. WildFish++ [44] collected 103,034 fish images from 2,348 species along with text descriptions. Other works [3, 6, 16, 17, 35] have also accumulated fish images from specific regions, primarily focusing on species classification. However, these datasets are limited to a relatively small proportion of the >35,000 described species of fish, posing a challenge to their application in real-world scenarios that require the recognition of diverse species. Therefore, there is an urgent need to develop more extensive and diverse datasets that can enable more robust AI-based systems to accurately identify a broader range of aquatic species, supporting aquatic biodiversity conservation efforts.

In this work, we present a large-scale diverse dataset, called FishNet, to foster aquatic species recognition research. FishNet contains 94,532 images of aquatic species from 17,357 species, collected from different regions around the world. Please refer to Table 1 for a detailed comparison of our FishNet dataset with existing datasets. FishNet contains 94,532 images with different sizes, resolutions, and illumination, representing a diverse set of aquatic species. The dataset is meticulously organized based on the scientific classification of aquatic species, including 8 taxonomic classes[1], 83 orders, 463 families, 3826 genera, and 17,357 species, and is accompanied by manual-labeled bounding box annotations.

Our FishNet dataset also provides traits associated with each species of the dataset, which can facilitate a better understanding of the ecological roles of aquatic species. To the best of our knowledge, we are the first to use deep learning techniques to predict functional traits directly from fish images. The proposed dataset can serve as a valuable resource for training and evaluating deep learning models for aquatic species classification and detection, and can further support research efforts toward the conservation and protection of aquatic ecosystems.

Additionally, we establish three benchmarks inspired by ecological research needs - fish classification, fish detection, and functional trait prediction, to facilitate the advancement of aquatic species recognition systems, and promote further research in the field of aquatic ecology. We conducted fish recognition and functional trait prediction using state-of-the-art deep learning models on these three benchmarks. The results show that accurate fish recognition and functional trait prediction on our FishNet dataset remains challenging: (1) The best-performing model achieved a classification accuracy of 61.38% at the family levels, and the best-performing detection model achieved mAP_{50} of 56.5 at the order level. (2) For functional trait prediction, the best-performing model achieved a classification accuracy of 54.80% and 18.00% under supervised and zero-shot settings, respectively.

2. Related Work

Fish taxonomy has been a longstanding area of scientific inquiry, with early works dating back to Aristotle's taxonomic classification of 117 Mediterranean fish in the 4th century BC [18]. In the 18th century, Linnaeus developed a widely-used system for naming and classifying organisms, and subsequent taxonomists have further built upon his work using both morphological and genetic data [2, 32]. Despite the advances made, classifying fish species remains a time-consuming process [30, 43]. In recent decades, computer vision techniques have played a vital role in automating the fish identification process, resulting in faster and more cost-effective identification of fish species [43]. However, computer vision models targeted at fish classification and detection face several challenges, including illumination change, low contrast, frequent occlusion, turbid water, and complex background [4, 43]. To tackle the challenges of fish recognition, a large-scale high-quality dataset is desired to automatically recognize aquatic species for biodiversity monitoring and protection. In this section, we provide a summary of previous aquatic species datasets and delineate

[1]We use "taxonomic class" to avoid confusion with the term "class" in the computer vision field that denotes the categorization of images or objects into groups based on their visual characteristics.

Datasets	Properties				Tasks		
	Images	Species	Taxonomy Level	Functional Traits	Classification	Trait Prediction	Detection
Fish4-Knowledge-A [15]	27,370	23	1	✗	✓	✗	✓
SEAMPD21 [3]	28,328	130	1	✗	✓	✗	✗
Fish-gres [35]	3,248	8	1	✗	✓	✗	✗
Mediterranean Fish Species [16]	≈40,000	20	1	✗	✓	✗	✗
Fish Abundance [17]	4,909	50	1	✗	✓	✗	✗
Image Dataset [12]	33,805	30	1	✗	✓	✗	✗
NCFM [6]	16,915	8	1	✗	✓	✗	✗
iNaturalist$_{Fish}$ 1001[21]	54,006	369	1	✗	✓	✗	✗
WildFish++ [44]	103,034	2,348	1	✗	✓	✗	✗
Ours	94,778	17,357	5-1	22	✓	✓	✓

Table 1. Comparison with existing datasets for fish recognition. FishNet provides annotations following biological taxonomy from class(5) to species(1); while previous datasets only provide species-level annotations. FishNet covers two orders of magnitude more species categories than existing datasets (17k vs 130). More importantly, FishNet provides additional functional traits that enable trait identification.

the limitations that impede their applicability for detailed analysis of aquatic species.

Lack of species diversity. Fish4Knowledge [15] is one of the pioneering works to build a fish recognition dataset. This dataset comprises a vast collection of approximately 700,000 underwater video clips captured in open sea waters in the Adriatic Sea in Croatia. A subset of 27,370 images was extracted from these video clips and manually annotated with bounding boxes and the corresponding species name. Another work [12] collects 33,805 images of 30 different fish species along with information about the meteorological and oceanographic conditions from the region of Barcelona, Spain. Several other works like [3, 6, 16, 17, 35] also collect fish images for classification from specific regions. These datasets are limited to a small number of species, posing a challenge to their application in real-world scenarios that require the recognition of diverse species. In contrast, our FishNet dataset contains more than 17,000 aquatic species which is more suitable for real-world application.

Lack of taxonomic diversity. Taxonomic classification of fishes is a crucial task in ecology and conservation biology [40]. Although several datasets have been previously developed for this task [3, 12], most of them have focused on a narrow taxonomic range (i.e. containing species from a small number of higher-order taxonomic groups), limiting their utility for exploring the wider diversity of fish species. For instance, the SEAMPD-21 [3] dataset comprises over 23,000 images, but only encompasses 130 species, while another dataset [12] containing more than 30,000 images includes merely 30 species. While these datasets serve as valuable resources for studying specific species, they do not provide a comprehensive representation of the diverse range of fish species present, which is crucial for understanding the fish ecosystem and recognizing the unique properties of different fish species.

Lack of functional traits. Functional traits are characteristics of an organism that relate to its ecological role [31, 39]. Functional traits in fish, such as trophic ecology, swimming modality, habitat use, and reproductive strategies, provide crucial insights into their ecological role and interactions with other species [28]. Understanding these traits can aid in sustainable fisheries management and maintaining ecosystem balance [38]. Monitoring functional diversity, rather than species diversity, is a common method of understanding ecosystem health, and guiding ecosystem management. Moreover, predicting functional traits can facilitate the understanding of probable functions performed by unknown or novel species, which otherwise may take considerable research effort and time to ascertain [9].

Unlike most existing large-scale bio-related datasets (such as iNaturalist [21] and WildFish++ [44]) that primarily focuses on species classification, the Brackish dataset [33] provides underwater images with bounding box annotations of fish, crabs, and starfish. The BrackishMOT [34] dataset focuses on tracking schools of small fish and provides 98 video sequences captured in the wild. The Caltech fish counting dataset [23] consists of sonar videos, which makes it challenging for diverse species recognition. Similar to functional traits in our FishNet dataset, the NEWT dataset [41] offers functional attributes within broader domains such as behavior, age, and health. In contrast, FishNet places its emphasis on habitat-related functional traits, e.g., tropical area suitability. A notable distinction lies in the breadth of species categories covered: our FishNet dataset encompasses a significantly larger spectrum of categories compared to the NEWT dataset (17,000 vs. 183). We compare our FishNet dataset with other existing fish-related datasets and summarize the key differences in Table 1.

3. FishNet

The goal of our FishNet dataset is to provide a large-scale fish image dataset that benchmarks fish recognition, detection, and functional trait predictions. In this section,

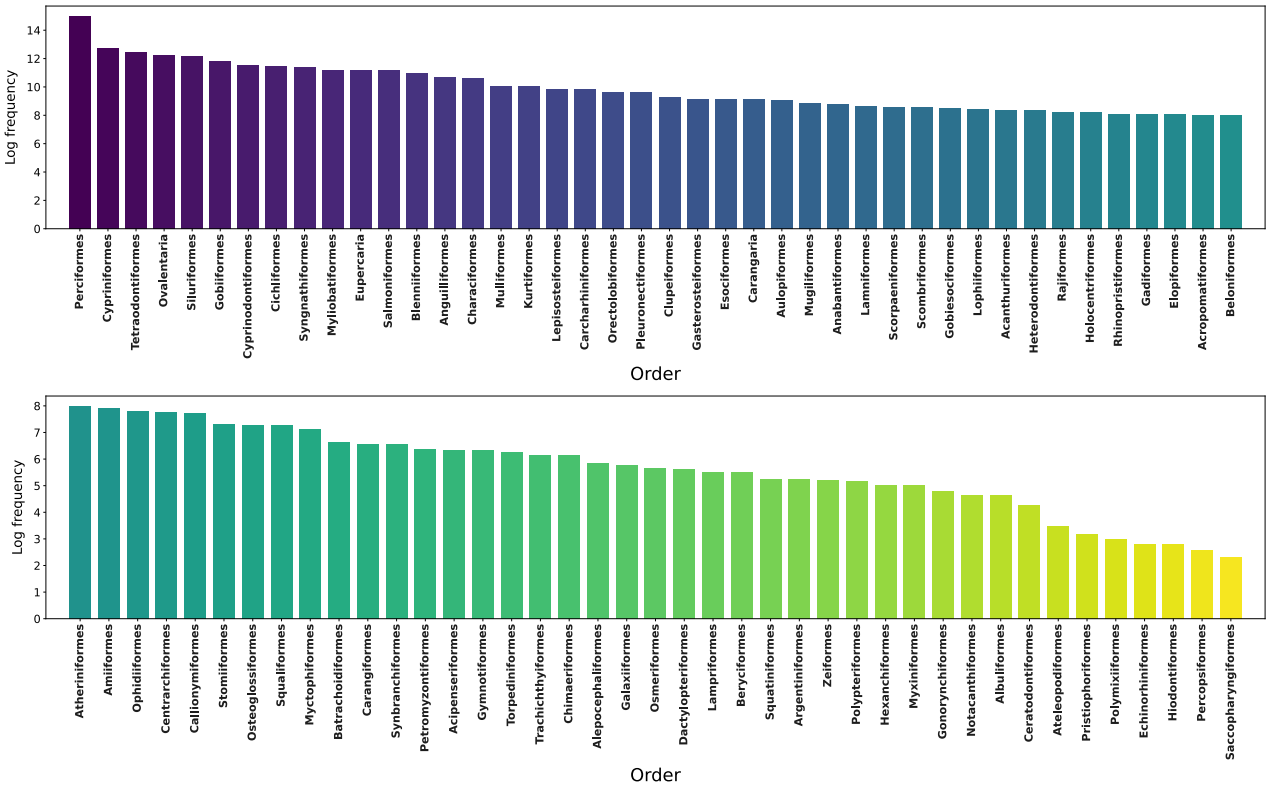

Figure 2. Order distribution. We plot the natural logarithm of the number of images per order. We show 42 more frequent orders at the top and 41 less frequent orders at the bottom.

we outline our dataset collection approach, including fish taxonomy construction, image and attribute collection, and detection annotation. To ensure the accuracy and consistency of our annotations, we implemented a quality control process during annotation collection. Finally, we present statistics of FishNet.

3.1. Fish taxonomy construction

The construction of fish taxonomy represents a critical preliminary step for the development of a fish recognition system. In order to facilitate the scientific and user-friendly utilization of our FishNet dataset, we constructed and organized the dataset in accordance with standard fish taxonomy protocols. To ensure the quality of the data and meet the ecological objectives, aquatic behavioral ecologists actively participated in this project and provided input on the design of the data collection for both taxonomical labels and functional traits. Initially, a comprehensive list of fish, amounting to approximately 35,024 distinct species, was sourced from FishBase [13]. Within FishBase, each individual fish is mapped to its respective position within the taxonomic hierarchy, consisting of taxonomic class, order, family, genus, and species. See Fig. 1 for a detailed structure of fish taxonomy.

3.2. Image and Functional Traits Collection

For each of the fish species in our taxonomic list, we collected available images from two key online databases, FishBase [13] and iNaturalist [21]. 51,433 images were scraped from FishBase using an updated version of the spiderfish package [37]. After we obtained the fish images from Fishbase [13], to ensure the quality and relevance of the data, we manually reviewed all the images and removed any instances that were not suitable for our analysis. This involved removing images that were not photographs of fish, such as images of fish sketches, coin images, stamp images, larval fish, and images of fish bone structures. By performing this manual filtering process, we were able to ensure that our dataset consisted only of high-quality images that were suitable for our research purposes. This process also helped to reduce the potential for errors or inaccuracies in our analysis, by removing any images that could introduce additional bias or noise into the data. Some of the removed instances are shown in Figure 3.

During this phase, 7,634 images were dropped from the dataset and the remaining 43,799 were added to our FishNet dataset. A further 51,006 images were then downloaded from the iNaturalist database. Combined, the compiled im-

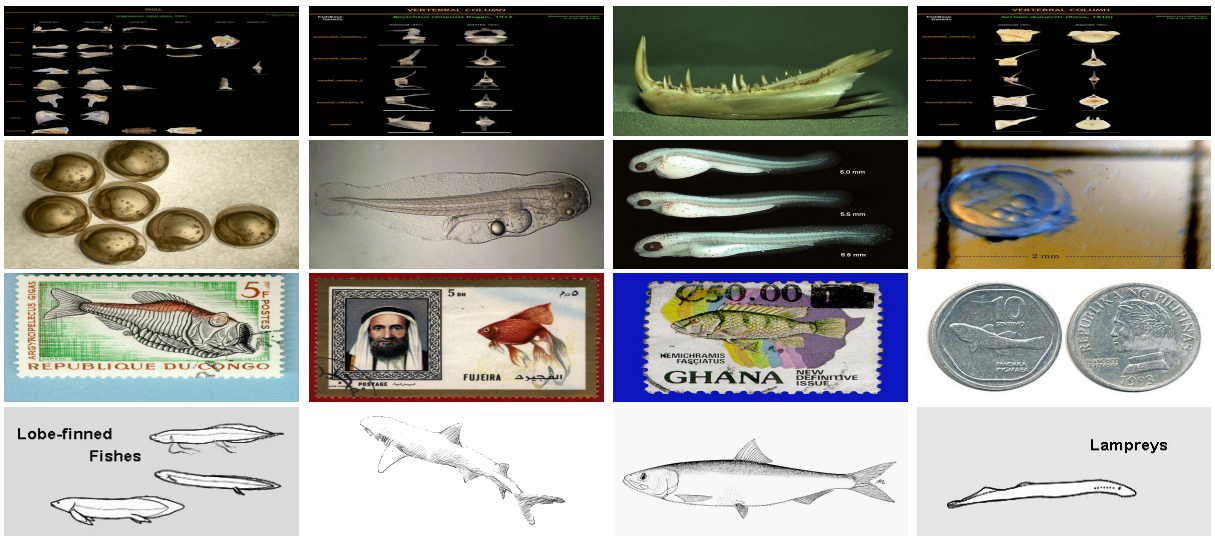

Figure 3. Examples removed from the dataset: The first row shows bone structures, the second row shows larval fish, the third row shows stamps and coins, and the fourth row shows fish sketches. All instances of these were identified and excluded.

age library encompasses 94,805 images from 17,518 species. To facilitate model training and evaluation, we dropped 273 images whose families have less than 5 images. For each species in our dataset, we gathered information on 22 traits (90,147 images have traits), which we believe to be important indicators of ecological roles and nutritional properties. These traits include the true functional traits of trophic level, feeding pathway, and habitat preferences based on temperature and salinity. In addition, we also included nutritional content "traits" for calcium, iron, omega-3 fatty acid, protein, selenium, vitamin A, and zinc content [20]. Table 2 provides detailed descriptions of the ten true functional traits.

3.3. Fish Detection Annotations

We employed the Amazon Mechanical Turk (AMT) [7] platform to collect bounding box annotations for all images in our dataset. Specifically, three independent annotators were tasked with providing locations for every instance in each image. To ensure accurate and reliable annotation of our image dataset, we have implemented a rigorous process for collecting annotations for object detection. Each image undergoes annotation by three different users on Amazon Mechanical Turk (AMT) [7]. We only accept annotations if all three users provide the same number of bounding boxes and if all the boxes have an Intersection over Union (IOU) value greater than 0.9. In cases where the annotations do not meet these criteria, we collect an additional annotation for the image and repeat the process of comparing the annotations reported by the different AMT users. We continue this process until we have at least three user annotations that agree with the number of bounding boxes and have an IOU greater than 0.9. For image categorization, we have already assigned labels to each image and asked users to select images that belong to a specific family by providing example images of that family. Through this meticulous process, we were able to calculate 114,375 object instances.

Trait	Value	Description
Trophic level	Real value	The position in the food chain or food web.
Feeding path	Benthic/ Pelagic	Species feeds primarily from which trophic pathway.
Tropical	True/ False	Whether the fish can live in tropical areas.
Temperate	True False	Whether the fish can live in temperate areas.
Subtropical	True/ False	Whether the fish can live in subtropical areas.
Boreal	True/ False	Whether the fish can live in boreal areas.
Polar	True/ False	Whether the fish can live in polar areas.
Freshwater	True/ False	Whether the fish can live in freshwater.
Saltwater	True/ False	Whether the fish can live in saltwater.
Brackish	True/ False	Whether the fish can live in brackish.

Table 2. The explanation for 10 selected functional traits in the FishNet dataset. Please check the supplementary for the rest.

3.4. FishNet Statistics

The FishNet dataset comprises a total of 94,532 images featuring 17,357 distinct species. The images were compiled

Backbone	Family Classification (%)				Order Classification (%)			
	Common	Medium	Rare	All	Common	Medium	Rare	All
ResNet-34 [19]	77.16	69.10	36.65	40.82	83.42	77.87	47.36	52.07
ResNet-50 [19]	76.82	70.27	35.99	40.37	82.46	75.93	42.84	47.94
ResNet-101 [19]	75.73	69.00	31.61	36.38	81.35	75.01	42.32	47.36
ViT-S [10]	76.08	67.02	33.67	37.97	84.14	74.78	44.09	48.79
ViT-B [10]	82.93	74.50	38.26	42.91	88.52	81.64	52.44	56.93
ViT-L [10]	85.51	77.05	44.18	48.40	89.02	83.89	55.94	60.26
BeiT [1]	86.09	77.67	50.78	54.26	91.41	88.24	38.16	45.97
ConvNeXt* [27]	42.63	26.56	12.57	14.53	63.58	40.67	17.82	21.12
ConvNeXt [27]	90.32	**85.13**	57.03	60.61	**94.07**	90.49	64.84	68.81
ConvNeXt [27] + FL [25]	88.28	82.02	51.22	55.16	87.60	81.11	22.22	31.37
ConvNeXt [27] + CB [8]	**90.53**	84.80	**57.94**	**61.38**	93.15	**90.56**	**71.41**	**74.38**
ConvNeXt [27] + FL [25] + CB [8]	84.89	77.92	48.99	52.71	85.14	57.25	39.28	41.76

Table 3. Fish Family/Order classification accuracy. * denotes training from scratch, 'FL' denotes using focal loss during training, and 'CB' denotes using class-balanced training.

from 8 taxonomic classes, 83 orders, 463 families, 3,826 genera, and 17,357 species, with each species having at least one image. We present the distribution of Orders within our dataset in Fig. 2. The figure displays the logarithmic frequency of each Order category, and it demonstrates a long-tail distribution within the FishNet dataset with respect to Order categories. We observed similar distribution for other taxonomy levels. In our FishNet dataset, each image in the dataset is accompanied by 22 functional traits. Additionally, the dataset includes instance-level bounding box annotations. In total, the FishNet dataset encompasses 114,375 object instances.

4. Experiments

Based on FishNet, we build three main challenges: 1) Fish classification. To identify family/order labels from input images; 2) Fish detection. Identifying the fish locations and family/order labels from input images; 3) Functional trait prediction. Predicting 10 functional traits from input images. For each challenge, we baseline several state-of-the-art models to show the capabilities of automatic fish identification and functional trait prediction using deep learning models. For all experiments, unless otherwise stated, we randomly select 80% for training and the remaining 20% for performance evaluation.

4.1. Fish Classification

Benchmark settings. We conducted Family/Order classification using both ResNet-based and ViT-based architectures, including ResNets (ResNet-34, ResNet-50, ResNet-101) [19], ViT (ViT-S, ViT-B, ViT-L) [10], BeiT [1], and ConvNeXt [27].We replaced the classification head using two fully connected (FC) layers, with a dropout of 0.5 for the first FC layer. For all comparing methods, the backbone networks were pretrained on ImageNet, and we finetuned the whole model using an Adam optimizer for 100 epochs. The initial learning rate was set to 3e-5, and 3e-4 for backbone networks and classification head, and we divided the learning rate by 2 every 20 epochs. In our experiments, we also explored recent techniques to address the long-tail classification problem. We explored focal loss [25] and class-balanced training [8] to improve the fish recognition performance in less frequent classes. Furthermore, we provided genus-level classification in the supplementary file.

Class split. To capture the effect of the long-tailed nature of the FishNet dataset, the FishNet dataset was partitioned into three distinct subsets, namely common, medium, and rare categories, based on their respective class frequencies. To categorize the classes according to their frequencies, we established a criterion for the Family/Order categories. Specifically, we selected those categories with a proportional representation of 25% of the total number of images as the common classes, those categories with subsequent 50% as medium classes, and the remaining 25% as rare classes. After splitting, we identified a total of 6, 52, and 405 common, medium, and rare family classes, respectively. In our experiments, we reported class-averaged accuracy for each split and also the overall accuracy among all splits.

Results. Table 3 shows the fish classification results at the Family and Order levels. From Table 3, it is evident that larger models consistently outperform smaller models in terms of classification accuracy. Specifically, the best-performing ConvNeXt model achieved an average classification accuracy of 60.61% and 68.81% at the Family and Order levels, respectively. We also observed that the performance

Figure 4. Selected visualization of FishNet Family classification. For each ground truth class, we showcase six examples in which the ConvNeXt model accurately predicts the correct Family label on the left. Additionally, we display two examples in which the ConvNeXt makes incorrect predictions on the right. For reference, we show the predicted Family class ('Cls') and confidence score ('Conf') above each image.

Level	Method	Common Classes/Medium Classes/Rare Classes					Average Per Class				
		mAP_{50}	mAP_{60}	mAP_{70}	mAP_{80}	mAP_{90}	mAP_{50}	mAP_{60}	mAP_{70}	mAP_{80}	mAP_{90}
Order	YOLOF [5]	77.3/65.2/39.2	72.7/62.1/38.0	61.6/53.4/34.7	38.8/34.9/25.4	11.7/9.6/11.8	45.0	43.4	39.3	27.8	11.7
	TOOD [11]	**84.8/76.8/50.3**	**80.1/73.2/48.5**	**69.4/64.4/43.6**	**46.1/42.6/32.7**	**15.4/14.3/13.6**	**56.5**	**54.3**	**48.5**	**35.4**	**14.1**
Family	YOLOF [5]	67.2/53.1/**27.1**	64.0/50.4/**26.6**	54.9/43.3/**24.4**	35.5/27.3/**18.8**	9.9/7.9/**7.7**	**30.6**	**29.8**	**26.9**	**20.0**	7.8
	TOOD [11]	**81.1/64.0**/22.5	**77.1/60.1**/21.9	**66.5/51.2**/20.1	**44.0/32.8**/15.7	**14.5/10.9/7.7**	27.9	26.9	24.2	18	**8.2**

Table 4. Fish detection performance on common, medium, and rare classes at the order and family taxonomic levels.

for rare classes was significantly worse than that for common classes, with an average classification accuracy of 57.03% and 64.84% at the Family and Order levels, respectively. This is because these rare classes have insufficient annotations, which poses great challenges for fish classification recognition.

Fig. 4 visualizes selected classification results on the Family level using the ConvNeXt model. We showed both good cases in the first six columns and bad cases in the last two columns. Based on the results presented in Figure 4, it is evident that the ConvNeXt model is capable of accurately identifying the Family categories of input images, even when they exhibit significant variations in illumination and shape. However, there are some instances where the model failed to perform optimally. These failures are primarily attributed to factors such as low contrast, challenging view-angles, or visual similarities between different families. For example, arhynchobatidae, urolophidae, and dasyatidae belong to cartilaginous fish meaning skeletons are made of cartilage, have flattened bodies, and have a venomous spine on their tail, which can cause confusion for our classification models, as shown in the bottom right of the figure. More examples can be found in the supplementary file.

Effect of pretraining. We investigated the effect of using pretraining on the ImageNet dataset. By comparing the performance between ConvNeXt and ConvNeXt*, using ImageNet pretraining can significantly improve the fish classification performance on our FishNet dataset.

Effect of balanced loss. From Table 4, using class-balanced loss can significantly improve classification performance, especially in rare classes; while using focal loss harms the performance. Specifically, for Order classification, using class-balanced loss improved the classification accuracy on rare classes from 64.84% to 71.41%. We emphasize that there exist several other techniques that can be employed to mitigate long-tail classification problems, which may potentially yield superior results. Nonetheless, our focus in this experiment was not to maximize performance but rather to raise awareness of the long-tail issue with our FishNet dataset, thereby encouraging further investigations.

4.2. Fish Detection

Benchmark settings. We conducted fish detection experiments using two well-known methods, YOLOF [5], and

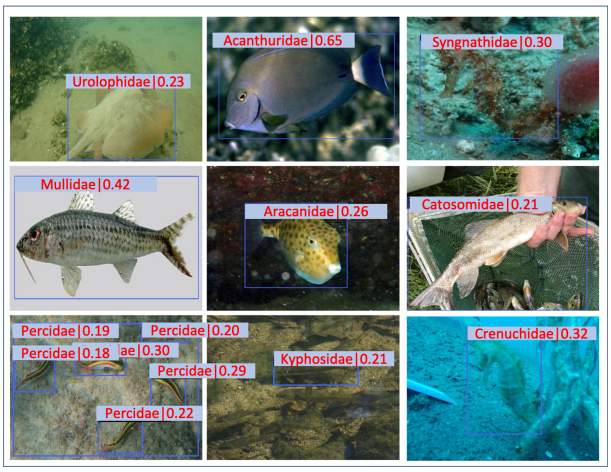

Figure 5. Visualization of fish detection results using family-level classification labels. We show successful cases in the first 2 rows and failure cases in the last row.

TOOD [11] on our FishNet dataset. For both methods, the backbone networks are pretrained on MS-COCO [26], and we finetuned the whole model using the stochastic gradient descent (SGD) optimizer with a learning rate of 2.5e-4, a momentum of 0.9, and a weight decay of 0.0001 for 100 epochs. For performance evaluation, we reported performance on two splits according to the Family/Order distribution, including common, medium, and rare classes, that contain 25%, 50%, and 25% of images.

Results. Table 4 shows the fish detection results at both the Family and Order levels. Our results demonstrated that for order-level classification, TOOD [11] outperformed YOLOF [5] in all the dataset splits. However, we observed that for family-level classification, YOLOF [5] exhibited better performance for Rare classes. Given the significantly higher number of Rare classes (405) as compared to Medium (52) and Common (6) classes, it is not surprising that the average Mean Average Precision (MAP) per class was lower for TOOD [11]. Therefore, we note that the evaluation of model performance can be influenced by the prevalence of different classes in the dataset, as shown in Table 4.

Fig. 5 visualizes several examples of fish detection results obtained by the TOOD model [11]. The model has proven successful in detecting fish from various species, with different scales, orientations, and backgrounds. However, the TOOD model may struggle to predict family labels for less frequent classes or those with low contrast. We provide more examples in the supplementary file.

4.3. Functional trait Prediction

Benchmark settings. In this section, we explored using deep learning models to predict fish functional traits based on input images. We tried ten functional traits in our experiments, namely trophic level, feeding path, and presence or absence in the following temperature and salinity zonations: tropical, temperate, subtropical, boreal, polar, freshwater, saltwater, and brackish. We used the same network architectures and training strategies as Section 4.1. For evaluating trophic level regression, we reported Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). For habitat classification, we used the average F1 score and overall accuracy to evaluate the performance. Additionally, we calculated the classification-all metric, which reflects the model's ability to accurately predict all functional traits simultaneously.

Backbone	Seen (%)		Unseen (%)	
	F1/Acc.	Acc-all	F1/Acc.	Acc-all
ResNet-34 [19]	74.83 / 87.20	40.04	54.38 / 78.48	15.84
ResNet-50 [19]	71.80 / 84.52	30.85	53.52 / 76.71	13.92
ResNet-101 [19]	73.80 / 85.67	34.45	54.37 / 77.47	15.12
ViT-S [10]	75.25 / 86.86	38.62	54.62 / 78.99	16.94
ViT-B [10]	78.65 / 88.94	46.91	55.72 / 79.93	18.44
ViT-L [10]	81.72 / 90.45	54.00	55.94 / **80.23**	**18.80**
BeiT [1]	78.15 / 88.93	49.97	54.60 / 78.82	17.26
ConvNeXt [27]	**83.30** / **91.03**	**54.80**	**56.34** / 79.50	18.00

Table 5. Habitat Classification performance in supervised and zero-shot settings.

Trophic level Regression. In our experiments, we found that different models showed similar performance for trophic level prediction, with an MAE of 0.45 and an RMSE of 0.60, respectively. We found that trophic level prediction is a challenging task when using common L1 or L2 loss functions. More advanced techniques, such as ordinal regression [14], could potentially remedy this challenge. However, we leave this for further research.

Habitat Classification Results. Table 5 shows the classification results for nine fish functional traits. Overall, the larger models outperformed the smaller ones. Specifically, the ConvNeXt model achieved the best results with an average f1-score of 81.72% and a classification accuracy of 90.45%. We also reported the classification accuracy when our model correctly predicts all nine functional traits simultaneously, marked as "Acc-all". It was surprising that the best-performing ConvNeXt model obtained an accuracy-all of 54.0% for correctly predicting all classes simultaneously. Fig. 6 shows selected examples of functional trait classification using the ConvNeXt model. From the figure, we can see that the ConvNeXt model can successfully predict most of the functional traits with large color and structure variations. In addition, Table 6 provides a detailed breakdown of the per-trait classification performance of the ConvNeXt model. The results show that the model achieved an f1-score of over 90% for four out of nine functional traits. However, the

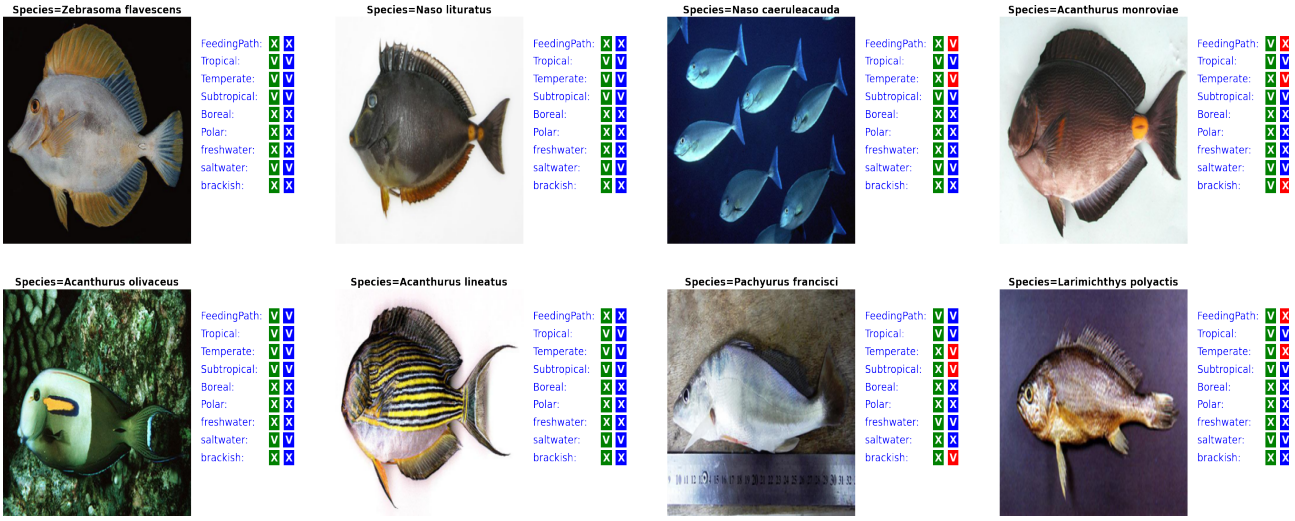

Figure 6. Selected visualization of functional trait classification. For each image, we show its species name on the top, and the "ground truth/predicted" functional trait labels on the right. The first two columns show success cases and the last two column show failure cases. Correct predictions are shown in blue color, false predictions are shown in red color.

Metric	FeedingPath	Tropical	Temperate	Subtropical	Boreal	Polar	Freshwater	Saltwater	Brackish	Avg.
Acc.	84.73	88.64	84.41	90.95	92.88	96.88	91.98	93.10	90.51	90.45
F1	68.07	92.97	88.03	94.49	76.99	59.91	91.02	95.59	68.40	81.72

Table 6. Per-trait classification performance. For each trait, we report both the F1 score (%) and classification accuracy (%).

model obtains less impressive performance for feeding path and polar classification.

Zero-shot Habitat Classification. We further evaluated the habitat classification performance in a zero-shot setting, where we evaluated habitat classification on unseen classes, which were not never seen during training. To achieve this, we trained the ConvNeXt model on 58 common and medium classes (according to their family classes) and evaluated the performance on 405 rare classes. From Table 5, although zero-shot habitat classification performance is worse than the supervised setting, the best-performed ViT-L model still shows a reasonable performance for zero-shot habitat classification, with a classification accuracy of 80.23% and 18.80% on accuracy and accuracy-all. This demonstration showcases the capacity of the deep learning model to predict the functional traits of previously unseen species, thereby establishing a valuable tool for the discovery of new species.

5. Conclusions and Future Work

This work presents FishNet, a large-scale diverse dataset consisting of 94,532 images from 17,357 species organized based on the scientific classification of aquatic species. The dataset is accompanied by 114,375 manually labeled bounding box annotations and 22 functional trait information. Furthermore, we establish three benchmarks, namely fish classification, fish detection, and functional trait prediction, to facilitate the advancement of aquatic species recognition and promote further research in the field of aquatic ecology. The experimental results demonstrate that accurate fish recognition and functional trait prediction are still challenging due to several factors, such as large species diversity, diverse backgrounds, low contrast, etc.

We release our data and associated code to encourage further research in developing more accurate and effective tools for monitoring and protecting aquatic ecosystems. Our FishNet dataset will also encourage research into evaluating the impact of climate change on aquatic ecosystems and identifying potential solutions to help maintain the health and resilience of aquatic ecosystems and contribute to mitigating the impacts of climate change.

6. Acknowledgements

We would like to thank Jun Chen and Habib Slim for providing their suggestions and Cameron Trotter for updating the spiderfish package to facilitate image collection from FishBase. This work is supported by KAUST, under Award No. BAS/1/1685-01-01. AJT is supported by KAUST Global Fellowship Program under Award No. ORA-2022-5001.

References

[1] Hangbo Bao, Li Dong, Songhao Piao, and Furu Wei. BEit: BERT pre-training of image transformers. In *International Conference on Learning Representations*, 2022. 6, 8

[2] Wilfrid Blunt. Linnaeus: the complete naturalist. 2001. 2

[3] Océane Boulais, Simegnew Yihunie Alaba, John E Ball, Matthew Campbell, Ahmed Tashfin Iftekhar, Robert Moorehead, James Primrose, Jack Prior, Farron Wallace, Henry Yu, et al. Seamapd21: A large-scale reef fish dataset for fine-grained categorization. In *Proceedings of the FGVC8: The Eight Workshop on Fine-Grained Visual Categorization, Online*, volume 25, 2021. 2, 3

[4] Guang Chen, Peng Sun, and Yi Shang. Automatic fish classification system using deep learning. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 24–29, 2017. 2

[5] Qiang Chen, Yingming Wang, Tong Yang, Xiangyu Zhang, Jian Cheng, and Jian Sun. You only look one-level feature, 2021. 7, 8

[6] The Nature Conservancy. The nature conservancy fisheries monitoring. *Kaggle Data*, 2017. 2, 3

[7] Kevin Crowston. Amazon mechanical turk: A research tool for organizations and information systems scholars. In Anol Bhattacherjee and Brian Fitzgerald, editors, *Shaping the Future of ICT Research. Methods and Approaches*, pages 210–221, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. 5

[8] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019. 6

[9] Victor E. L. da Silva, Luana P. S. Silva-Firmiano, Fabrício B. Teresa, Vandick S. Batista, Richard J. Ladle, and Nidia N. Fabré. Functional traits of fish species: Adjusting resolution to accurately express resource partitioning. *Frontiers in Marine Science*, 6, 2019. 3

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 6, 8

[11] Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R. Scott, and Weilin Huang. Tood: Task-aligned one-stage object detection, 2021. 7, 8

[12] Simone; ;Martínez Enoc;Del Río Joaquín;Toma Daniel M.;Nogueras Marc; Francescangeli, Marco; Marini and Jacopo Aguzzi. Image dataset for benchmarking automated fish detection and classification algorithms. *Scientific Data*, 117(24):13596–13602, 2023. 3

[13] R. Froese and D. Pauly. Fishbase, 2021. Accessed: March 6, 2023. 4

[14] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, and Dacheng Tao. Deep ordinal regression network for monocular depth estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2002–2011, 2018. 8

[15] J. Chen-Burger G. Nadarajan and R. Fisher. A knowledge-based planner for processing unconstrained underwater videos. *Proc. IJCAI-09 Workshop on Learning Structural Knowledge From Observations*, (37-44), 200-9. 2, 3

[16] Giannis Georgiou. Fish species. *Kaggle Data*, 2021. 2, 3

[17] Ben L. Gilby, Andrew D. Olds, Rod M. Connolly, Nicholas A. Yabsley, Paul S. Maxwell, Ian R. Tibbetts, David S. Schoeman, and Thomas A. Schlacher. Umbrellas can work under water: Using threatened species as indicator and management surrogates can improve coastal conservation. *Estuarine, Coastal and Shelf Science*, 199:132–140, 2017. 2, 3

[18] E. W. Gudger. The five great naturalists of the sixteenth century: Belon, rondelet, salviani, gesner and aldrovandi: A chapter in the history of ichthyology. 1934. 2

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6, 8

[20] Christina C. Hicks, Philippa J. Cohen, Nicholas A. J. Graham, Kirsty L. Nash, Edward H. Allison, Coralie D'Lima, David J. Mills, Matthew Roscher, Shakuntala H. Thilsted, Andrew L. Thorne-Lyman, and M. Aaron MacNeil. Harnessing global fisheries to tackle micronutrient deficiencies. *Nature*, 574, 2019. 5

[21] iNaturalist. inaturalist, 2021. Accessed: March 6, 2023. 3, 4

[22] S. Irfan and A Alatawi. Aquatic ecosystem and biodiversity: A review, 2019. 2

[23] Justin Kay, Peter Kulits, Suzanne Stathatos, Siqi Deng, Erik Young, Sara Beery, Grant Van Horn, and Pietro Perona. The caltech fish counting dataset: A benchmark for multiple-object tracking and counting, 2022. 3

[24] Schaffner-F.C. Abdelhadi A. Goosen M.F.A. Laboy-Nieves, E.N. *Environmental Management, Sustainable Development and Human Health*. CRC Press, 2008. 2

[25] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 6

[26] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 8

[27] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022. 6, 8

[28] David Mouillot, Nicholas A.J. Graham, Sébastien Villéger, Norman W.H. Mason, and David R. Bellwood. A functional approach reveals community responses to disturbances. *Trends in Ecology Evolution*, 28(3):167–177, 2013. 3

[29] Peter B. Moyle and Robert A. Leidy. *Loss of Biodiversity in Aquatic Ecosystems: Evidence from Fish Faunas*, pages 127–169. Springer US, Boston, MA, 1992. 2

[30] Erling Holm Nicholas E. Mandrak Eric Taylor Mary Burridge Douglas Watkinson Pierre Dumont Allen Curry Paul Bentzen Junbin Zhang Julien April Louis Bernatchez Nicolas Hubert, Robert Hanner. Identifying canadian freshwater fishes through dna barcodes. *PLOS ONE*, 2008. 2

[31] Tom H. Oliver, Matthew S. Heard, Nick J.B. Isaac, David B. Roy, Deborah Procter, Felix Eigenbrod, Rob Freckleton, Andy Hector, C. David L. Orme, Owen L. Petchey, Vânia Proença, David Raffaelli, K. Blake Suttle, Georgina M. Mace, Berta Martín-López, Ben A. Woodcock, and James M. Bullock. Biodiversity and resilience of ecosystem functions. *Trends in Ecology Evolution*, 30(11):673–684, 2015. 3

[32] Marta Paterlini. There shall be order. the legacy of linnaeus in the age of molecular biology. *EMBO reports*, 2007. 2

[33] Malte Pedersen, Joakim Bruslund Haurum, Rikke Gade, and Thomas B Moeslund. Detection of marine animals in a new underwater dataset with varying visibility. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 18–26, 2019. 3

[34] Malte Pedersen, Daniel Lehotskỳ, Ivan Nikolov, and Thomas B Moeslund. Brackishmot: The brackish multi-object tracking dataset. In *Scandinavian Conference on Image Analysis*, pages 17–33. Springer, 2023. 3

[35] Nanik; Fatichah Chastine Prasetyo, Eko; Suciati. Fish-gres dataset for fish species classification. *Mendeley Data*, 2021. 2, 3

[36] Julia Reiss, Jon R. Bridle, José M. Montoya, and Guy Woodward. Emerging horizons in biodiversity and ecosystem functioning research. *Trends in Ecology Evolution*, 24(9):505–514, 2009. 2

[37] spiderfish. spiderfish, 2021. Accessed: March 6, 2023. 4

[38] Rick D. Stuart-Smith, Amanda E. Bates, Jonathan S. Lefcheck, J. Emmett Duffy, Susan C. Baker, Russell J. Thomson, Jemina F. Stuart-Smith, Nicole A. Hill, Stuart J. Kininmonth, Laura Airoldi, Mikel A. Becerro, Stuart J. Campbell, Terence P. Dawson, Sergio A. Navarrete, Elisabeth M. A. Soler, German A.and Strain, Trevor J. Willis, and Graham J. Edgar. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*, 501, 2013. 3

[39] Nuria Teixidó, Maria Cristina Gambi, Valeriano Parravacini, Kristy Kroeker, Fiorenza Micheli, Sebastien Villéger, and Enric Ballesteros. Functional biodiversity loss along natural co2 gradients. *Nature Communications*, 2018. 2, 3

[40] Devis Tuia, Benjamin Kellenberger, Sara Beery, Blair R. Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W. Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D. Couzin, Grant van Horn, Margaret C. Crofoot, Charles V. Stewart, and Tanya Berger-Wolf. Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13, 2022. 3

[41] Grant Van Horn, Elijah Cole, Sara Beery, Kimberly Wilber, Serge Belongie, and Oisin Mac Aodha. Benchmarking representation learning for natural world image collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12884–12893, 2021. 3

[42] Lieven Wittebolle, Massimo Marzorati, Lieven Clement, Annalisa Balloi, Daniele Daffonchio, Kim Heylen, Paul De Vos, Willy Verstraete, and Nico Boon. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*, 458, 2009. 2

[43] Ling Yang, Yeqi Liu, Huihui Yu, Xiaomin Fang, Lihua Song, Daoliang Li, and Yingyi Chen. Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: A review. *Archives of Computational Methods in Engineering*, 28, 2021. 2

[44] Peiqin Zhuang, Yali Wang, and Yu Qiao. Wildfish++: A comprehensive fish benchmark for multimedia research. *IEEE Transactions on Multimedia*, 23:3603–3617, 2020. 2, 3