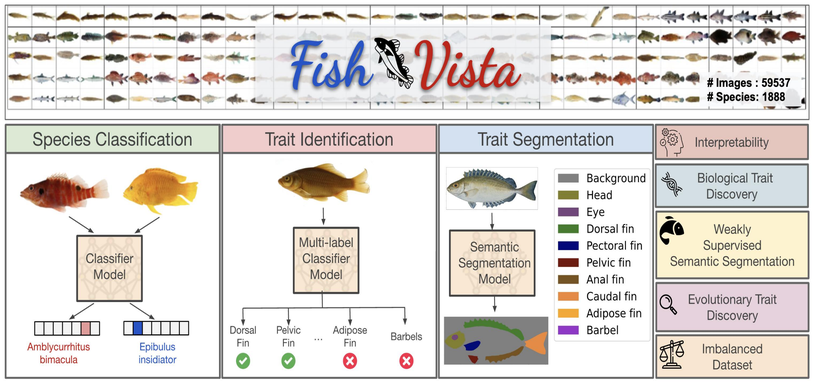
Fish-Vista: A Multi-Purpose Dataset for Understanding & Identification of Traits from Images

Kazi Sajeed Mehrab*[1], M. Maruf*[1], Arka Daw*[7], Harish Babu Manogaran[1], Abhilash Neog[1], Mridul Khurana[1], Bahadir Altintas[4], Yasin Bakis[4], Elizabeth Campolongo[5], Matthew J Thompson[5], Xiaojun Wang[4], Hilmar Lapp[6], Wei-Lun Chao[5], Paula Mabee[2], Henry Bart[4], Wasila M Dahdul[3], Anuj Karpatne*[1]

[1]Virginia Tech, [2]Battelle, [3]University of California Irvine, [4]Tulane University, [5]The Ohio State University, [6]Duke University, [7]Oak Ridge National Lab

Motivation

- The **availability of large datasets of organism images** combined with **advances in AI** have significantly enhanced **image-based studies of organisms**
- However **existing organism image datasets have several *limitations:***
  - Existing datasets often focus solely on species classification, **overlooking tasks related to visual traits of organisms.**
  - They lack detailed visual trait annotations, such as pixel-level segmentation, crucial for in-depth biological research
  - They feature organisms in natural habitats, making it difficult to study aquatic species like fish due to poor visual clarity in underwater images
- These limitations hamper the study of aquatic biodiversity and macroevolution

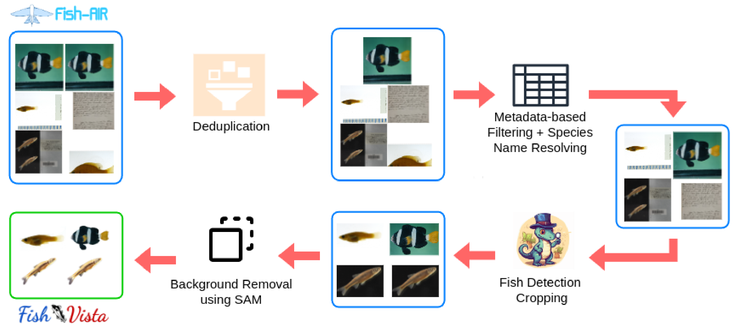


- To address this gap, we introduce the Fish-Visual Trait Analysis (Fish-Vista) dataset:
  - **60K annotated fish images** spanning **1900 species**
  - Supports **biologically relevant and novel tasks**, including species classification, trait identification, and trait segmentation
  - Curated through a **sophisticated data processing pipeline** applied to images sourced from **various museum collections**
  - Ensures clear visibility of visual traits, providing **fine-grained labels** for each image
  - Pixel-level annotations of 9 different traits are available for **4k fish images**, enabling trait segmentation and localization tasks.
- A comprehensive analysis of state-of-the-art deep learning techniques on Fish-Vista is provided to benchmark the dataset.

Related Datasets

Dataset	Organism	# Species	# Images	Background	Trait Information	Trait Segmentation
CUB-200-2011 [1]	Birds	200	11,788	Natural Habitat	✓	x
Birds 525 [2]	Birds	525	89,885	Natural Habitat	x	x
NABirds [16]	Birds	555	48,562	Natural Habitat	✓	x
Stanford dogs [7]	Dogs	120	20,580	Natural Habitat	x	x
Oxford Pet [8]	Cats, Dogs	37	7,349	Natural Habitat	x	x
FathomNet [22]	Marine Species	2244	84,454	Natural Habitat	x	x
Ulucan et al. [19]	Fish	9	9,000	Controlled	x	x
QUT Fish [4]	Fish	468	3,960	Natural Habitat + Controlled	x	x
DeepFish [9]	Fish	NA	39,766	Natural Habitat	x	x
Fish4Knowledge [23]	Fish	23	27,370	Natural Habitat	x	x
FishBase [5]	Fish	35,600	64,000	Natural Habitat	✓	x
iNaturalist-2021-Fish [18]	Fish	183	46,996	Natural Habitat	x	x
Fish-Vista (Ours)	Fish	1888	59,537	Controlled + Uniform	✓	✓

Data Processing Pipeline



Manual Filtering of Images

- We further manually filter images that do not have all visual traits visible

Fish-Vista Tasks and Stats

Fine-grained Species Classification

Task: Categorizing an image to its species, considering fine-grained differences

Dataset:

- **Main classification dataset:**
  - At least 20 images per species; **48k images, 419 species**
- **Rare species dataset**
  - 4-20 images per species; **12k images, 1660 species**

Challenges:

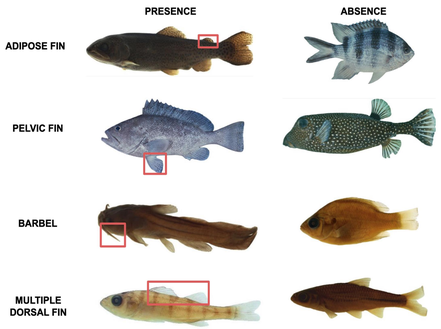
- Extreme long-tailed imbalance
- Subtle, fine-grained differences between species

Trait Identification

Task: Predicting the presence/absence of 4 visible traits that vary across species

Dataset:

- Species-level traits from knowledge bases like Phenoscape and FishBase
- **53k images, 682 species**
- 51 species are held-out of the training set to evaluate the out-of-species generalizability



Challenges:

- **Trait imbalance:** some traits are rare, like adipose, barbel and multiple dorsal fin
- The trait information is at the species-level, we want to predict on each image
- **Out-of-species generalizability** is more important

Trait Segmentation

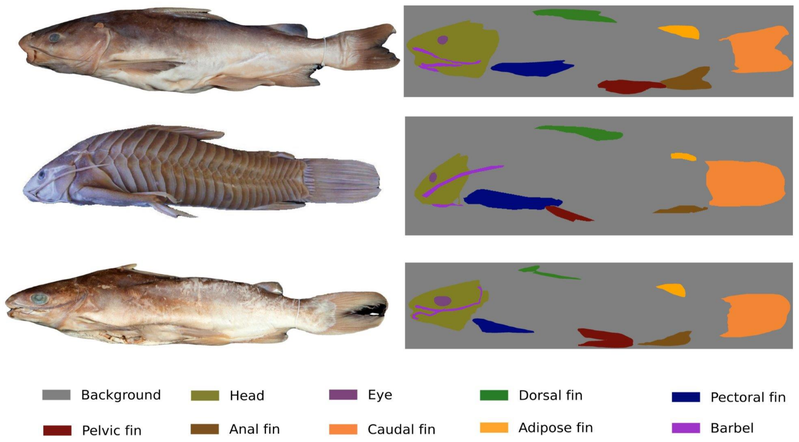
Task: Involves semantic segmentation of visual traits in fish images

Dataset:

- Manual segmentation annotation of pixel-level traits by expert biologists using the CVAT tool
- Annotation is expensive: 30 images/hour; requires manual inspection
- **4k images, 2.2k species; diverse set of species**

- **Challenges:**
  - Trait imbalance: adipose fin and barbel are rarely present
  - Some traits are rare and very small, like the barbel
  - Some traits are rare and similar looking to other traits, like the adipose fin

Examples of Manual Segmentation Annotations



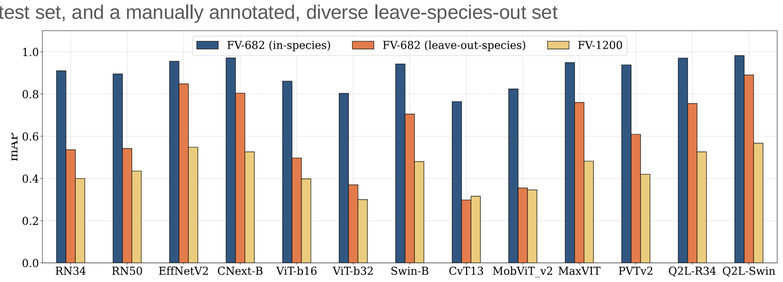
Benchmarking Results and Findings

Fine-grained Species Classification

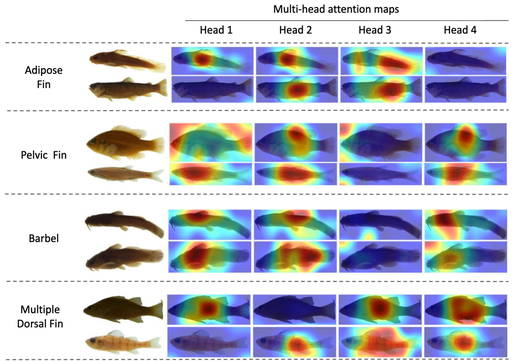
- SOTA baseline methods achieve poor accuracy on minority (~75%) and rare species (~50%)
- Different imbalanced classification methods only slightly improve performance minority performance, while lowering overall performance

Trait Identification

- We evaluate identification baselines on three sets: in-distribution test set, leave-species-out test set, and a manually annotated, diverse leave-species-out set



- SOTA methods fail to generalize on out of distribution species
- Models do not 'localize' traits in their predictions



Trait Segmentation

- SOTA methods fail to segment rare and small traits – adipose fin and barbel

Acknowledgement

This work is supported by the National Science Foundation (NSF) under Award No. 2118240.