COSE474-2024F: Final Project
Classification of Commonly Caught Fish using CLIP Fine-Tuning

Sung Jin Kim 2022320128

1. Introduction

1.1. Motivation

The automated classification of fish presents new possibilities to advance fish conservation efforts and economic activities. It could aid sustainable fishing practices by enabling automated fishing systems to target specific species while avoiding overfishing or bycatch. In regulatory contexts, it could play a critical role in monitoring illegal fishing practices, providing real-time evidence that protected species are caught in regulated waters. Furthermore, such technology offers potential benefits in recreational fishing, helping hobbyists identify catches quickly and accurately. With the rapid advancements of automated systems, such possibilities could soon become a part of our daily life.

1.2. Problem definition & challenges

There are several challenges that can make it difficult for general image classification models to correctly identify fish species. First, the visual similarity between species can be high, especially between closely related species of fish. Secondly, there are several environmental conditions that can lead to misclassification, such as lighting, visibility of the water, and the life stage of the fish. Lastly, the limited availability of labeled data on lesser known types of fish can pose a challenge to general models that have shallow data on the subject.

1.3. Concise description of contribution

This project explores the use of a fine-tuned CLIP model for fish species classification, achieving a test accuracy of 71% across 31 species. By evaluating the performance and confusion matrix for each class, this work provides insight into the potential and challenges of automated fish classification for real-world applications.

2. Methods

2.1. Significance & Novelty

While CLIP is well known for its zero-shot capabilities, its application to domain-specific tasks such as fish species identification remains limited. For instance, the pre-trained CLIP model was only able to achieve a 24% zero-shot accuracy on the fish dataset used in this project.

However, by fine-tuning CLIP with a classification head, the model was able to achieve a test accuracy of 71%, showcasing its adaptability to domain-specific tasks and highlighting CLIP's potential for species identification tasks.

2.2. Algorithm & Reproducibility

At its core, CLIP consists of two encoders: an image encoder and a text encoder, both of which map images and text into a shared embedding space. During class inference, CLIP compares the embeddings of the input image to the embeddings of text prompts it has already learned and computes the similarity scores between them. Then, the text prompt with the highest similarity score is selected as the class.

The pre-trained CLIP model already comes with a wide variety of text embeddings allowing for zero-shot classification. However, this text embedding knowledge does not generalize well in certain specialized tasks, which is why fine-tuning the model is necessary to achieve better performance.

In the case of this project, a lightweight classification head (linear layer) was added on top of the image encoder. This classification head maps the image embeddings produced by CLIP to the 31 classes of fish species. During training, the parameters of this head were updated using backpropagation with the cross-entropy loss between the predicted class label and the ground truth label.

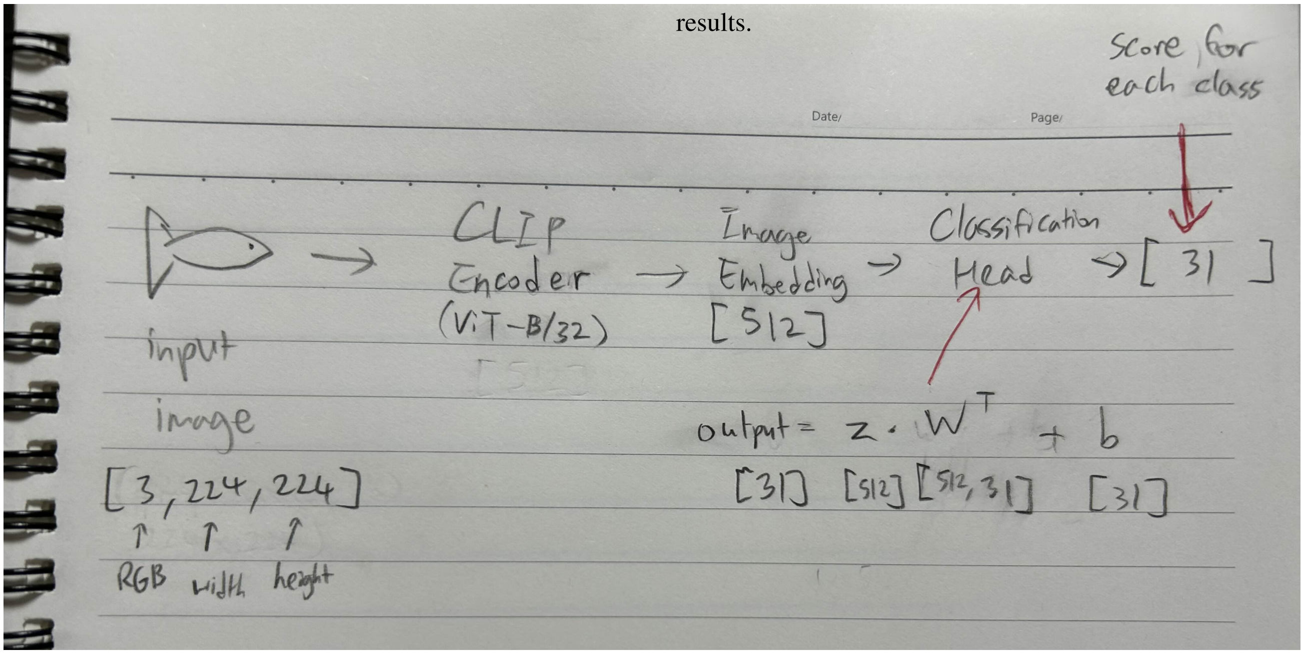


Figure 1. Workflow of the CLIP fine-tuning process.

3. Experiments

3.1. Dataset

The dataset being used is "Fish Dataset" on kaggle, a dataset of 13,304 images that contains 31 different fish species found at the Marinig Fishing Port in Cabuyao City (Lampa et al., 2022). This dataset was used to fine-tune a pre-trained CLIP model in hopes of achieving a higher classification accuracy than zero-shot classification.

3.2. Computer Resource & Experimental Design

The local computer environment handled the entire training process, using an Intel-i7 CPU, NVIDIA GTX 1660 Ti GPU, Windows 10 OS, and pytorch installed in a Python 3.10 environment.

First of all, the class names were extracted from the dataset and were used to produce generic text prompts (a photo of X). Then, using these text prompts, the zero-shot accuracy of the CLIP model was computed.

Next, the linear function of the classification head was fine-tuned by training with the dataset and backpropagating weights through a cross-entropy loss function. Due to the simplistic nature of the linear classification task, there was no need for a validation process with hyperparameter tuning.

Finally, the trained classification head was evaluated with a new test dataset. The bar plot for per-class accuracy and confusion matrix were printed for better visualization of the results.

3.3. Results

The pre-trained CLIP model achieved a zero-shot accuracy of 0.24 (2115/8811) on the training dataset.

After 7 epochs, the model achieved a training accuracy of 0.6770 with a loss of 1.5141.

The fine-tuned model achieved a test accuracy of 0.71 (1243/1761) on the test dataset.

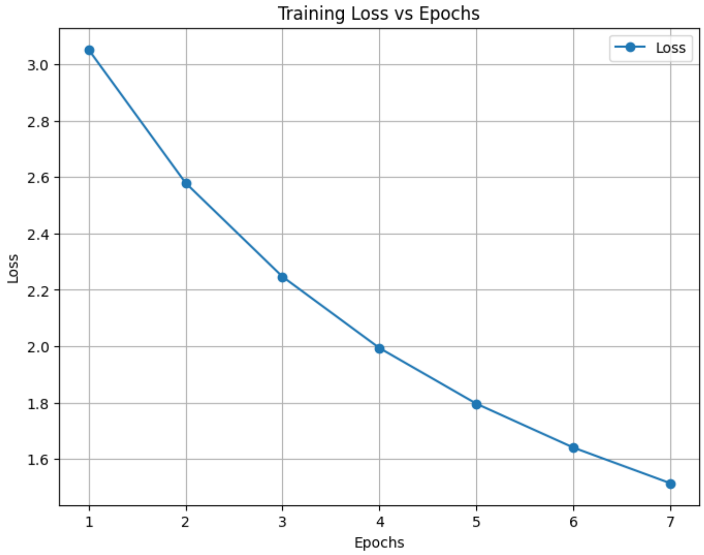


Figure 2. Visualization of training loss over epochs.

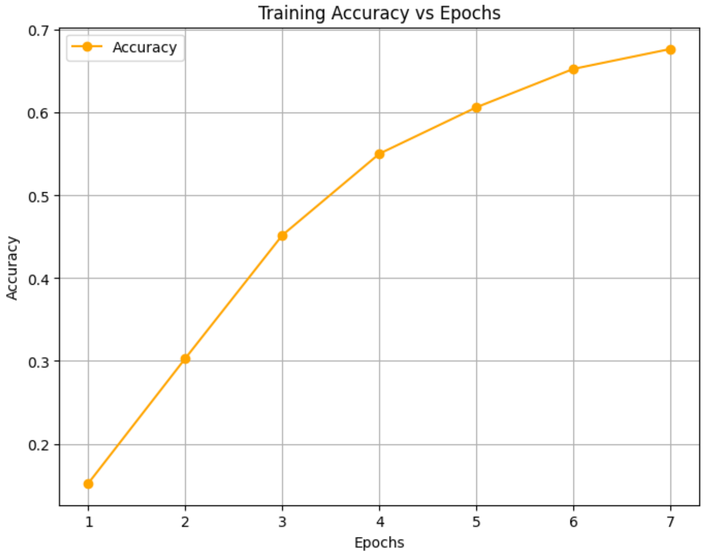


Figure 3. Visualization of training accuracy over epochs.

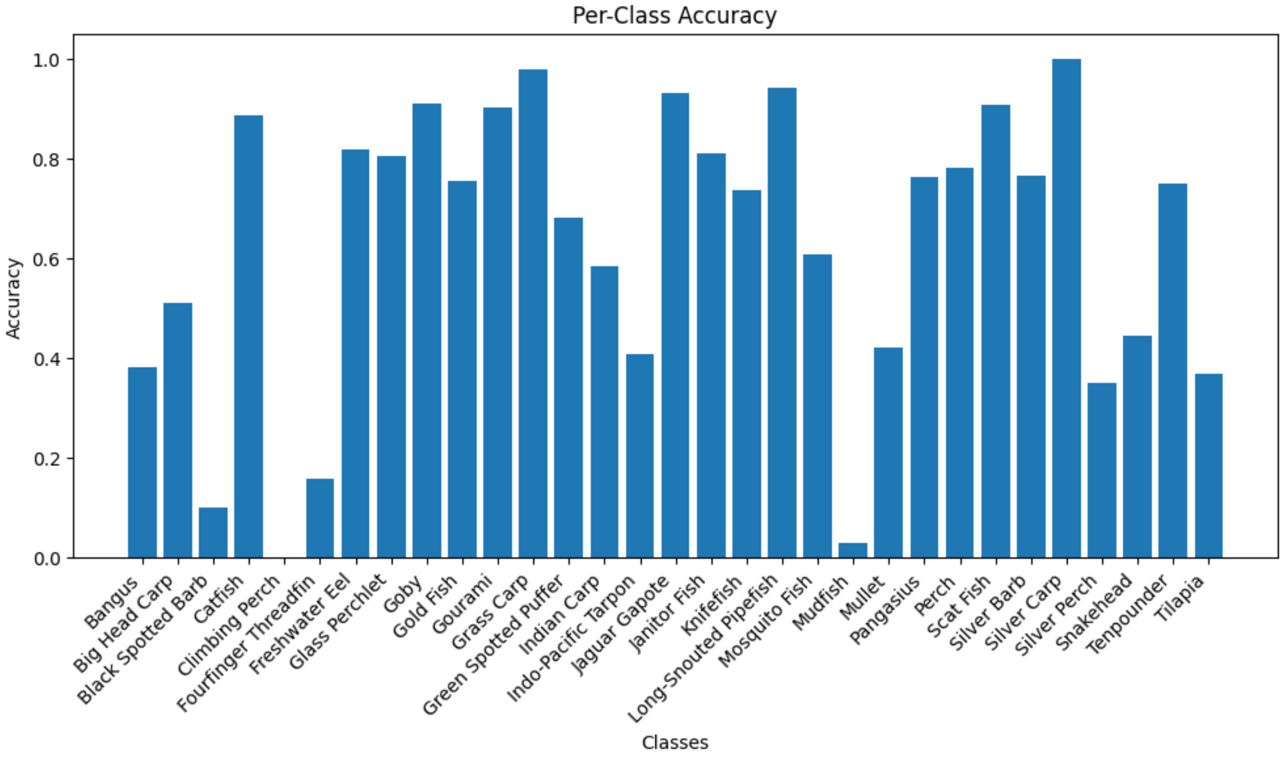
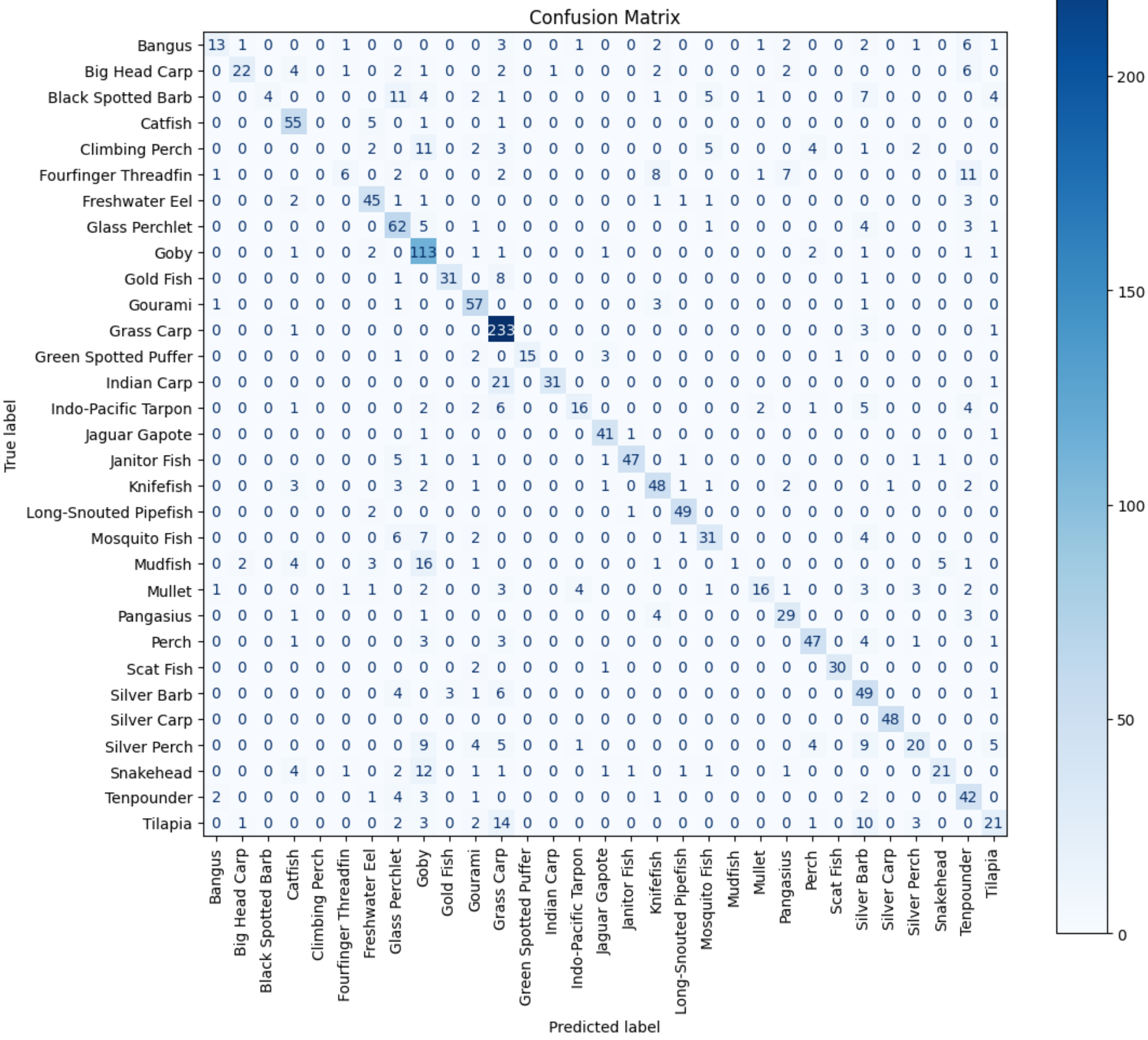


Figure 4. Bar plot depicting per-class test accuracy for each class.



	Bangus	Big Head Carp	Black Spotted Barb	Catfish	Climbing Perch	Fourfinger Threadfin	Freshwater Eel	Glass Perchlet	Goby	Gold Fish	Gourami	Grass Carp	Green Spotted Puffer	Indian Carp	Indo-Pacific Tarpon	Jaguar Gapote	Janitor Fish	Knifefish	Long-Snouted Pipefish	Mosquito Fish	Mudfish	Mullet	Pangasius	Perch	Scat Fish	Silver Barb	Silver Carp	Silver Perch	Snakehead	Tenpounder	Tilapia
Bangus	13	1	0	0	0	1	0	0	0	0	0	3	0	0	1	0	0	2	0	0	0	1	2	0	0	2	0	1	0	6	1
Big Head Carp	0	22	0	4	0	1	0	2	1	0	0	2	0	1	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	6	0
Black Spotted Barb	0	0	4	0	0	0	0	11	4	0	2	1	0	0	0	0	0	1	0	5	0	1	0	0	0	7	0	0	0	0	4
Catfish	0	0	0	55	0	0	5	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Climbing Perch	0	0	0	0	0	0	2	0	11	0	2	3	0	0	0	0	0	0	0	5	0	0	0	4	0	1	0	2	0	0	0
Fourfinger Threadfin	1	0	0	0	0	6	0	2	0	0	0	2	0	0	0	0	0	8	0	0	0	1	7	0	0	0	0	0	0	11	0
Freshwater Eel	0	0	0	2	0	0	45	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	3	0
Glass Perchlet	0	0	0	0	0	0	0	62	5	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	3	1
Goby	0	0	0	1	0	0	2	0	113	0	1	1	0	0	0	1	0	0	0	0	0	0	0	2	0	1	0	0	0	1	1
Gold Fish	0	0	0	0	0	0	0	1	0	31	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Gourami	1	0	0	0	0	0	0	1	0	0	57	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	0	0	0
Grass Carp	0	0	0	1	0	0	0	0	0	0	0	233	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1
Green Spotted Puffer	0	0	0	0	0	0	0	1	0	0	2	0	15	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Indian Carp	0	0	0	0	0	0	0	0	0	0	0	21	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Indo-Pacific Tarpon	0	0	0	1	0	0	0	0	2	0	2	6	0	0	16	0	0	0	0	0	0	2	0	1	0	5	0	0	0	4	0
Jaguar Gapote	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	41	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Janitor Fish	0	0	0	0	0	0	0	5	1	0	1	0	0	0	0	1	47	0	1	0	0	0	0	0	0	0	0	1	1	0	0
Knifefish	0	0	0	3	0	0	0	3	2	0	1	0	0	0	0	1	0	48	1	1	0	0	2	0	0	0	1	0	0	2	0
Long-Snouted Pipefish	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	49	0	0	0	0	0	0	0	0	0	0	0	0
Mosquito Fish	0	0	0	0	0	0	0	6	7	0	2	0	0	0	0	0	0	0	1	31	0	0	0	0	0	4	0	0	0	0	0
Mudfish	0	2	0	4	0	0	3	0	16	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	5	1	0
Mullet	1	0	0	0	0	1	1	0	2	0	0	3	0	0	4	0	0	0	0	1	0	16	1	0	0	3	0	3	0	2	0
Pangasius	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	4	0	0	0	0	29	0	0	0	0	0	0	3	0
Perch	0	0	0	1	0	0	0	0	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	47	0	4	0	1	0	0	1
Scat Fish	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0
Silver Barb	0	0	0	0	0	0	0	4	0	3	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	1
Silver Carp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0
Silver Perch	0	0	0	0	0	0	0	0	9	0	4	5	0	0	1	0	0	0	0	0	0	0	0	4	0	9	0	20	0	0	5
Snakehead	0	0	0	4	0	1	0	2	12	0	1	1	0	0	0	1	1	0	1	1	0	0	1	0	0	0	0	0	21	0	0
Tenpounder	2	0	0	0	0	0	1	4	3	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	42	0
Tilapia	0	1	0	0	0	0	0	2	3	0	2	14	0	0	0	0	0	0	0	0	0	0	0	1	0	10	0	3	0	0	21

Figure 5. Confusion matrix depicting all combinations of predicted class with ground truth class.

3.4. Analysis

There was a significant improvement in classification accuracy from 24% to 71%, so it is clear that the proposed method was at least somewhat successful.

However, it is worth noting that the test data had a disproportionate amount of Grass Carp which the model excelled at classifying as shown in Figure 5, so the test accuracy would be lower if every fish class is weighted equally.

The steady rise of training accuracy, decline of training loss, and high test accuracy indicates that the model was fitted well, not being too overfitted or underfitted to the training data. However, when looking at the per-class accuracy as shown in Figure 4, it is clear that the model could not accurately identify certain species such as Black Spotted Barb (4/40), Climbing Perch (0/30), and Mudfish (1/34).

In the case of the Mudfish, the majority of instances were predicted as Goby, a fish with similar visual properties. Upon closer inspection of the training data, there were 600 instances of Goby, while there were 200 instances of Mudfish, which likely skewed the model into predicting Goby in case the class was unclear.

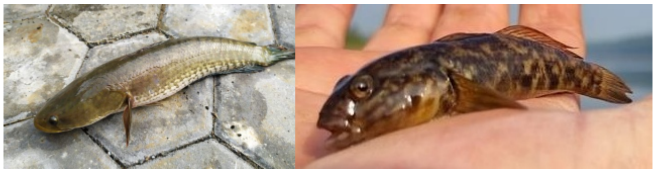


Figure 6. A side-by-side comparison of a Mudfish (left) and Goby (right), from kaggle dataset

4. Future Direction

To improve the accuracy of the worst performing fish species, there could be improved feature representation alongside CLIP image encoder to better distinguish them from more common fish classes. Also, a more balanced performance metric could be used instead of pure accuracy to improve performance on rarer fish classes.

All in all, this project shows that even a simple fine-tuned CLIP model can achieve respectable performance on fish classification. With a more diverse dataset and extensive training, the technology could indeed be used to more accurately identify a wider range of fish.

5. Github History

Github Link

6. Overleaf History

Overleaf Link

References

Lampa, M. D., Librojo, R. C., and Calamba, M. M. Fish dataset, 2022. URL https://www.kaggle.com/dsv/4323384.