

February 7, 2022

# **Fish Family Classification using Deep Learning**

Data analysis and preparation

Sam Tihen

CMPSCI5390

University of Missouri - St. Louis

<https://www.overleaf.com/read/sqkwpscdbrsk>

# 1 Introduction

Deep Learning can be used to classify what type of fish is pictured in a photograph. This project intends to explore that topic.

To create the dataset, online image searches were used to find and download more than 1000 images, split across 5 different Linnaean families of fish.

The images were then imported into Apple Photos and exported as JPEG High Quality with a maximum height and width of 500px to reduce the overall dataset file size and limit any file format issues.

## 1.1 Project Motivation

I created a website called MyScubaDives.com in 2011, and the primary purpose of the website was to create a database of all dive sites and dive operators around the world, and allow people to log their dives and share images of what they saw. Around this time, image "tagging" and facial recognition was being popularized by Facebook, and I thought it would be interesting to attempt to use AI to help people identify what fish were in their photos. The site ultimately failed to gain popularity, and I lost interest in the project as usage and revenue never really grew. This class project seems like a good opportunity to explore what I failed to do a decade ago.

## 2 Dataset

The dataset contains approximately 200 images of each of the following categories:

1. Clownfish — Item Count: 201
2. Lionfish — Item Count: 215
3. Parrotfish — Item Count: 210
4. Triggerfish — Item Count: 215
5. Butterflyfish — Item Count: 210

The photographs are of multiple species within each family, and are taken from various angles and with various lighting conditions.

Each image is limited to one family of fish, but multiple individuals within the same family may be present.

There are 1051 images total.

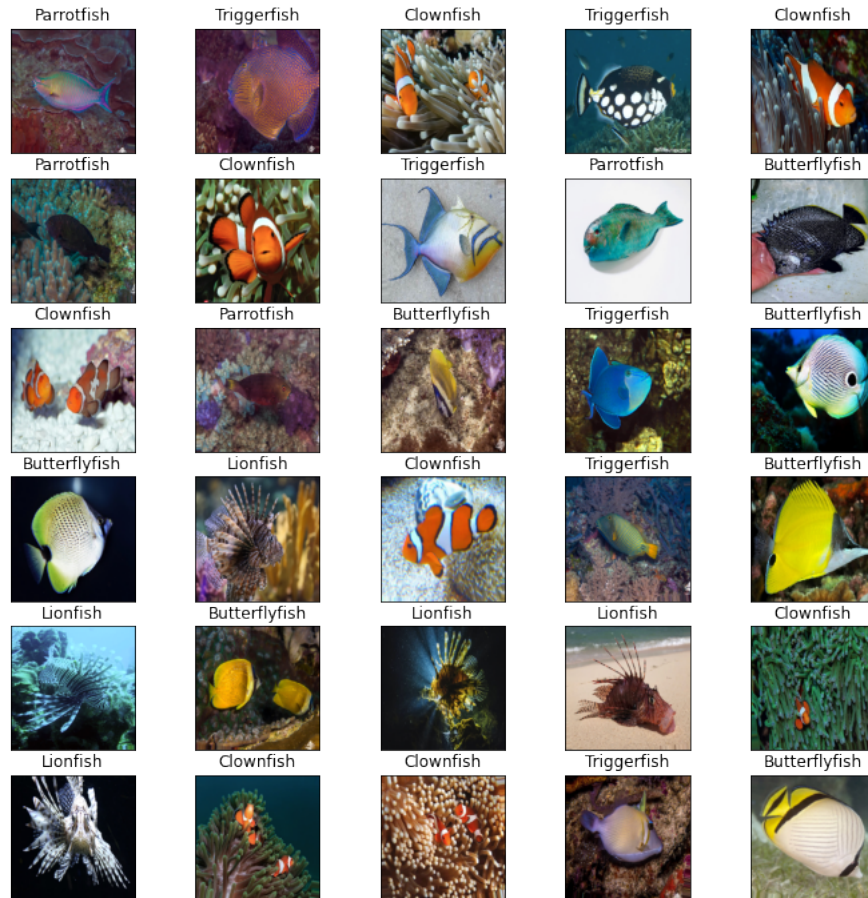


Figure 1: Dataset example

## 2.1 Project Data Processing Code

The data for this project was manipulated using Google Colab and Python.

URL: <https://colab.research.google.com/drive/12tNNGjWssaP43wdXAOhgY41qJ8AIDNhn>

## **2.2 Data Normalization**

The data in the initial dataset will be normalized by rescaling by dividing by the maximum.

## **3 Conclusion**

## **References**