# Kaggle Fish

Univ.ai Project

ROHIT GOSWAMI

June 22, 2019

# Contents

# Problem Statement

Given unknown images and seven classes of fish, we need to isolate said fish and classify them.

# Data

## Training

- The training set has `json` bounding boxes and tags
- It is also tagged by folder

## Test

We will eventually have to handle the following use-cases before using the test set.

- ☐ Skewed
- ☐ No fish
- ☐ Wrong size
- ☐ Edges
- ☐ Backgrounds

# Methodology

## Class Approach

- Use transfer learning
- Pop the top and put two dense layers, one for classification and one for the box, that is 8 outputs and 4 outputs

## Our Approach

### TODO Fish/Not Fish

Here we want to first **augment** the existing data to allow us to make a binary fish or no fish model.

1. Augmentation

   **With existing data** Here we will be using random splits and information from the `json` bounds to work out our fish/no-fish data and hence train a simple logistic regressor.

   **With external data** Here we will ignore the fact that we have no bounding box information and instead pass in a **split-sized** image of a fish or no fish.

### TODO Bounding Box

Once we have the regions then we will use the bounding boxes to further clean the images

### TODO Classifier

Finally we are in a position to run a simple classifier on the not fish classes, that is a seven class classifier

# Data Exploration

## Structure

After obtaining the data and extracting said data, we note the following folder structure.

```
cd /Storage/DataSets/KaggleFish/train/train
ls
```

> ALB BET DOL LAG NoF OTHER SHARK YFT

Additionally, the bounding box data is in the form:

```
cd /Storage/DataSets/KaggleFish/bbox
head alb_labels.json
```

```
[
    {
^^I"annotations": [
^^I    {
^^I^^I"class": "rect",
^^I^^I"height": 151.06975503141317,
^^I^^I"width": 383.68430384213445,
^^I^^I"x": 547.1578480789353,
^^I^^I"y": 193.3597426816281
^^I    }
^^I],
^^I"class": "image",
^^I"filename": "img_07917.jpg"
    },
]
```


- We note that the `class` tag is not useful for us

## Quality

- Since we also have bounding box data for the no-fish cases, we decided to manually inspect some of the labels and the corresponding regions

### Fish

We tested the ALB tuna data-base and noted, that the head was chopped off by the bounding box