Topic: Fish Detection and Counting in Aquaculture environment for automated monitoring and management

The farming of fish is an essential component of global food production,however efficient management of fish in ponds is crucial.one of the challenges of fish farming is manual monitoring and counting of fish.The aim is to train a model to detect fish and count all the fish in the pond using a counting algorithm. If we can have a real number of fish we can modify feeding schedules and minimize feed waste,we can track disease outbreak and lastly population count of fish can help the farmer to know good time to harvest fish, this can maximize yields while maintaining level of sustainable production.I have downloaded my dataset from kaggle, my dataset contain images of fish in water. I have annotated all fish images, capturing every fish in the dataset and surrounded it with a rectangle box named fish. I will install some cameras in water to detect and once it is detected that it is fish then the counting algorithm will count it. To prevent counting the fish more than one I will integrate the counting algorithm and tracking algorithm once the fish is detected by the trained algorithm it will be counted by counting algorithm and tracking algorithm will assign uniqueidentifier to the detected and counted fish, even if the fish will be moving around it will be counted only one time. The name of the dataset is Fishes in the wild, here are steps I have made my dataset ready to train a model

DATA CLEANING IN IN IMAGE DATASET

1. **Checking duplicate images and removing them**
   Duplicate images are images that are copies of other images in the same folder. Removing duplicate images has the advantages such as increased data quality and improving training of our model. I have checked and I created new folder that contains filtered images
2. **Resizing all images to have same sizes**
   I have resized all images in the filtered folder by 1024 by 768 pixels.resizing ensures that all images in the dataset have same dimension that simplify model training
3. **Removing corrupted images**
   These are images that can not opened or processed, these images might be incomplete or be in damaged states
4. **Converting different image format into one format**
   It involves changing file format of images into one format to ensures that all images are in a consistent format

5. **Removing background Noise**
   I have removed background noise using a code, it enhances image quality by eliminating unnecessary elements in the back of the images

DATA TRANSFORMATION

1. **Rotating other images**
   Means turning it around a center point, usually in increments of 90 degrees, it increases model generalization. I have only rotated 100 pictures
2. **Annotating the dataset**

Then after cleaning the dataset and rotating some pictures I annotated all the images using VGG image annotator(Via) I downloaded it from Github. I have used a bounding box of rectangles to annotate fish pictures. But some other pictures had 5 to 23 images on them.

After labeling I saved all the images in comma separated value(.csv)

EXPLORATORY DATA ANALYSIS IN ANNOTATED DATASET

Firstly I had to load again the dataset but in comma separated value format(csv) using pandas

1. **Showing few column of the annotated images**

   using head() function in pandas

2. **Showing total number of annotated fish images**

   I have shown total number of images using shape attribute in pandas and total images is 1,467

3. **Distribution of annotations per image**

   I have used histogram to plot distribution of annotation

   Distribution is showing that other images are annotated much more than others, it suggests that other images contain much more fish than other images.

4. **Distribution of Bounding box dimension**

   The skewness of the histogram is to the right(positive skewed) this suggests that most of bounding box are smaller and big bounding box are smaller in the annotated dataset

5. **File Size analysis**

   The histogram of file size is indicating that peak is in the middle meaning most of the annotated images have file size around the mean, suggesting a consistent resolution and quality for the majority of images.

6. **Scatterplot of Bounding box dimensions**

   The upward points suggest a positive correlation between the width and height of the bounding boxes. Theis mean that as the width increases the height also increase

Changing data format of annotated dataset csv to coco-json format

I have changed the format of the annotated dataset to coco-json format using code. I changed the format to COCO_JSON format because it provides a standardized and widely adopted way to present object detection,segmentation and keypoint annotation tasks.

FEW ANALYSIS OF THE ANNOTATED FISH IMAGE DATASET BUT IN COC0-JSON FORMAT

All the graphs plotted are just similar to the graph I have already explained. The difference is just the format. I have just added a scatterplot **of the bounding box location**: it is showing many points are clustered at the center, it suggests that annotations frequently occur in this region. This indicates common locations of objects in the images.

Challenges I have Faced when cleaning, transforming and analyzing the dataset

1. **Annotating a Large Number of Images**: Manually annotating thousands of images was time-consuming.
2. **Handling Diverse Image Quality**: Ensuring consistent quality across images required significant preprocessing efforts.
3. **Converting Data Formats**: Transitioning from CSV to COCO-JSON format involved complex transformations and validation.
4. **Ensuring Data Consistency**: Maintaining consistency in annotation and image processing was crucial for reliable model training.