Progress Report

Ziyu(Caroline) Xu

March 7, 2025

Introduction

The purpose of this project is to develop a robust fish detection model using the Caltech Fish Counting Dataset (CFC-DAOD). The dataset has challenges in domain adaptation because it requires the model to generalize from the source domain (Kenai Left Bank) to the target domain (Kenai Channel). Based on the exploration of the dataset and the challenges, YOLO was selected as the primary detection model due to its efficiency. This report presents the exploratory data analysis and model selection process that informs the development of an optimized fish detection model.

Data Analysis

Dataset Overview

The dataset follows the COCO annotaton format which includes images, annotatons, and categories. The images have metadata such as file name, image ID, image height and image width. The annotatons have annotaton about bounding box coordinates (x, y, width, height) for detected fish. The categories for both train and test contain only one fish species which the category id is 1 and the name is 'fish'. Table 1 shows the training set consists of 76619 images with 132010 annotatons, while the test set includes 13090 images with 41761 annotatons.

Dataset	**Images**	**Annotations**	**Categories**
Source (Train)	76619	132010	1
Target (Test)	13090	41761	1

Table 1: Summary of the dataset composition, including the number of images, annotations, and catogories for the training (Source) and test (Target) sets.

Target Size Distribution

Table 2 provides a summary of the bounding box dimensions for fish instances in the dataset. The width of the bounding boxes ranges from 2 to 214 pixels, with an average of 68.55 pixels. The height ranges from 1 to 164 pixels, with an average of 22.21 pixels. Bounding box areas range from 6 pixels^2 to 20,424 pixels^2, with an average of 1,671.21 pixels^2. This variation in size highlights the presence of both very small and relatively larger fish in the dataset, requiring a model capable of detecting small objects effectively.

Metric	**Minimum**	**Maximum**	**Mean**
Width (pixels)	2.0	214.0	68.55
Height (pixels)	1.0	164.0	22.21
Area (pixels^2)	6.0	20424.0	1671.21

Table 2: Summary of Bounding Box Dimensions in the Dataset

Figure 1 provides further insight into the size distribution of bounding boxes by illustrating their area ratios relative to the entire image. The histogram shows that the majority of bounding boxes occupy a small fraction of the image, with most area ratios clustering below 0.01. The red dashed line at 5% area ratio highlights that almost all fish are small object.

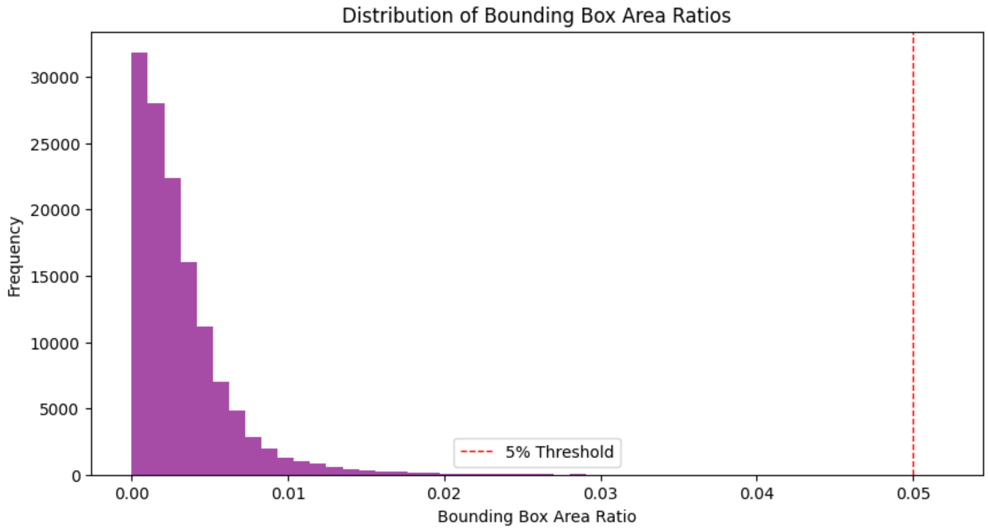

Figure 1: Distribution of Bounding Box Area Ratios in the Dataset

To complement this analysis, Figure 2 presents a scatter plot of bounding box width versus height, with color representing the bounding box area. This visualization reinforces the observation that most fish have narrow, elongated bounding boxes (low height-to-width ratios), and a significant number of fish have very small bounding boxes in both dimensions. The color gradient further highlights the concentration of small objects in the dataset.

Given these findings, it is clear that YOLO's anchor sizes may need to be changed. The presence of a significant number of small bounding boxes suggests that the detection performance could benefit from adjusting anchor sizes to match the distribution of object dimensions. Additionally, increasing the input resolution (e.g., from 640×640 to 1024×1024) may improve the detection of smaller fish by maintaining more details.

Background Complexity Analysis

The background complexity of the dataset was analyzed using color complexity scores. It measures the variation in hue and saturation across images. The average color complexity score is 34,077.68, with a standard deviation of 23,909.92. This indicates a large difference in background characteristics. This high variability suggests that the dataset contains a mix of images with simple and highly complex backgrounds. Figure 3 presents the distribution of color complexity scores. The histogram shows a multi-modal distribution, with distinct peaks at different complexity levels. The first peak, located at a lower complexity range, likely corresponds to images with relatively uniform backgrounds, such as clear water conditions. The second and third peaks, appearing at higher complexity scores, indicate the presence of images with more complicated backgrounds. It might includes underwater vegetation, varying lighting conditions, or reflections.

The dataset's diverse background complexity shows challenges for object detection because models trained on simpler backgrounds may struggle to generalize to more complex environments. High background variation can also increase false positives due to lighting changes and texture noise. To address this, preprocessing techniques such as color normalization and adaptive brightness adjustment can help standardize image appearance.

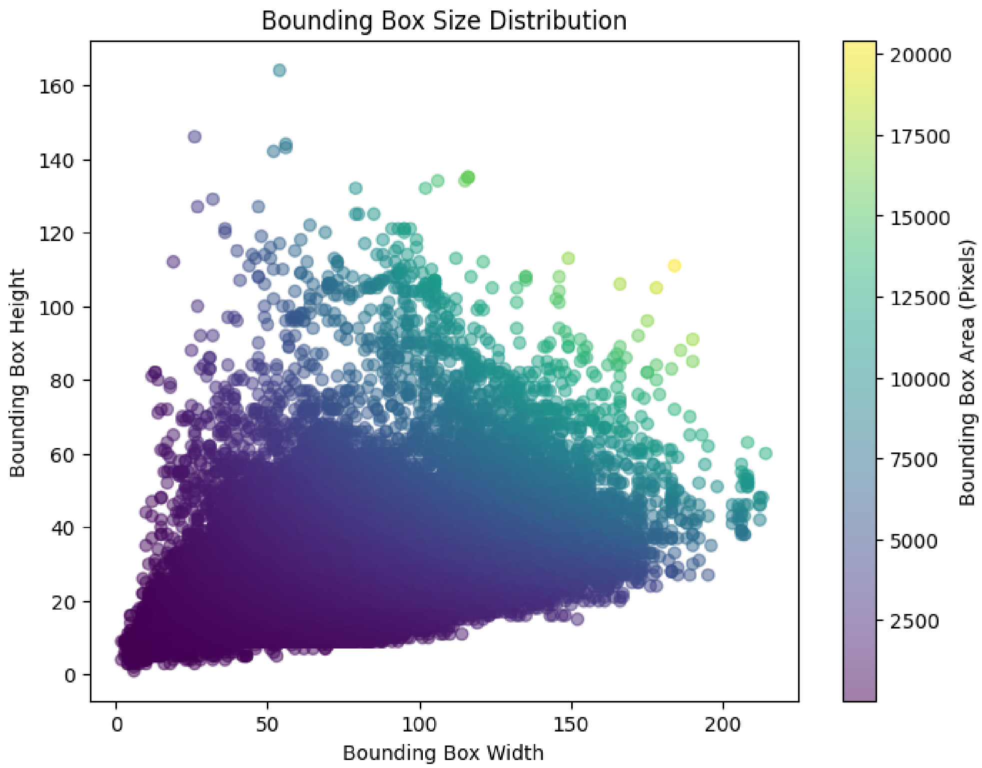

Figure 2: Bounding Box Size Distribution in the Dataset

Texture Complexity Analysis

The texture complexity of the dataset was assessed using edge density, which quantifies the proportion of edge pixels in each image. The analysis demonstrates an average edge density of 0.6756, with a standard deviation of 0.5590. It indicates significant difference in texture across images. This variation suggests that there are some images feature relatively smooth backgrounds, but others contain a high degree of texture. It might due to factors such as underwater vegetation, water surface disturbances, or image noise. Figure 4 illustrates the distribution of edge density scores. The histogram shows a right-skewed distribution, where most images have low edge density. A smaller subset exhibits significantly higher values.

High texture complexity can lead to false positives because strong background edges may be misinterpreted as fish. However, low texture complexity may make it harder to distinguish fish from their surroundings. To improve robustness, Gaussian blur can help reduce background noise in high-texture images. CutMix augmentation can enhance generalization. Additionally, training with multi-scale features will improve detection across varying texture levels.

Target Density Analysis

The distribution of fish across images was analyzed to assess the likelihood of object overlap and potential occlusion challenges for the detection model. The dataset contains a total of 132,010 fish annotations across 76,619 images. It results in an average fish density of 1.72 fish per image. This relatively low density suggests that most images contain only one or two fish. It reduces the probability of overlap. Additionally, the average distance between fish instances is 246.84 pixels. It could further indicate that fish are generally well-separated within images. Given these findings, the dataset does not present significant challenges related to occlusion. Since fish are not frequently overlapping, non-maximum suppression (NMS) in YOLO can remain at its default setting.

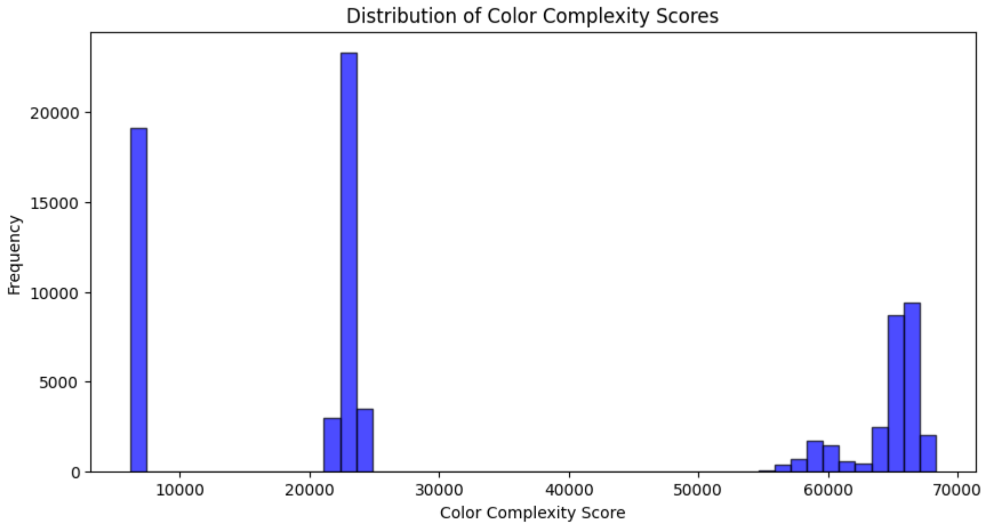

Figure 3: Distribution of Color Complexity Scores in the Dataset

Method

The methodology for this study focuses on developing an optimized fish detection model using YOLO. The primary challenges include small object detection, high background complexity, and variable texture conditions. This section details the model selection reasons, anchor tuning strategy, data augmentation techniques, and preprocessing methods aimed at improving detection performance.

Model Selection: Why YOLO?

YOLO was chosen as the primary detection model because of its efficiency in real-time object detection and its ability to handle small objects through optimized anchor boxes. Unlike two-stage detectors such as Faster R-CNN, which rely on a region proposal network. YOLO processes images in a single pass which makes it computationally efficient. Given that the dataset includes small fish, YOLO's anchor-based detection approach provides a suitable framework.

Optimizing YOLO for Small Objects

To enhance YOLO's performance for small fish, the following modifications would be implemented:

- **Adjusting Anchor Sizes:** The default anchor box dimensions will be updated based on the bounding box size distribution in the dataset. This ensures that the predefined anchors align closely with the actual fish dimensions.
- **Increasing Input Resolution:** Training YOLO with *1024×1024 resolution* instead of the standard *640×640* will keep more details which enhance the model's ability to detect small fish.
- **Using Mosaic Augmentation:** This augmentation technique combines multiple images into one which allows the model to learn from fish of different sizes and positions in varied contexts. This method might be particularly effective in small-object detection, as it exposes the model to diverse fish scales.

Preprocessing

The dataset's background and texture variability require targeted preprocessing to enhance model robustness. **Color normalization** and **adaptive brightness adjustments** will standardize image conditions, while **background suppression** methods may be applied in high-texture environments to reduce noise. To further

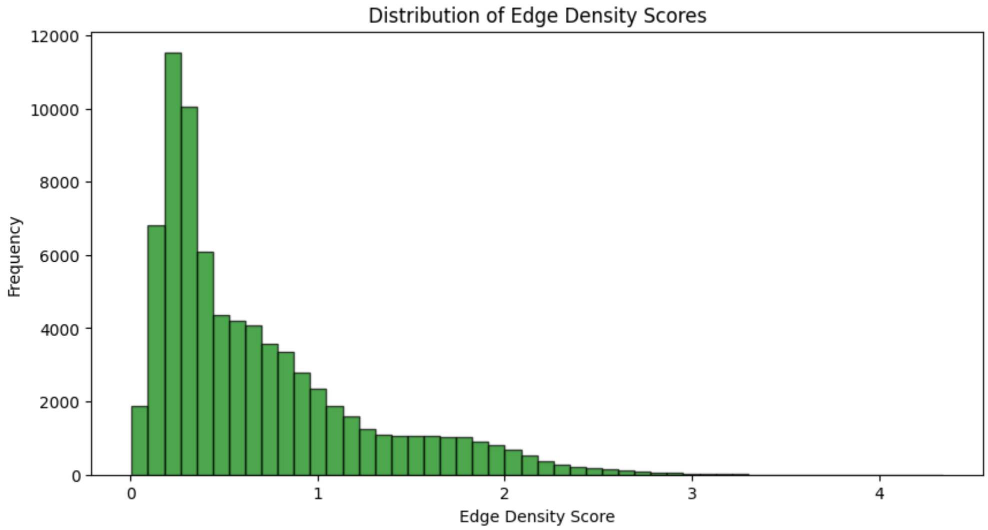

Figure 4: Distribution of Edge Density Scores in the Dataset

improve generalization, **Gaussian blur** will help mitigate edge interference, and **Mixup and CutMix augmentation** will expose the model to diverse textures.