A Deep (Sea) Dive into Crown-of-Thorns Starfish Underwater Image Data

1 Introduction

This report presents findings from the exploratory data analysis (EDA) conducted on underwater image data for Crown-of-Thorns Starfish (COTS) detection. The guiding objective for this project is to develop an algorithm to detect COTS in underwater image data. Subsequently, the goal of this analysis is to better under the data structure and the characteristics of the underwater image data (e.g., patterns in scale and location of COTS) that can influence model design decisions to improve starfish detection.

The dataset is composed of `jpg` images and a `csv` file that details the video, sequence and bounding box dimensions per image. All images have width 1280 pixels and height 720 pixels.

2 Overview of Image Data

This section examines the distribution of images across videos and sequences at a high-level.

### 2.1 How have the original videos been decomposed into images?

There are 3 underwater videos divided into 8 sequences (4 for video 2), with unknown ordering, so it is unclear if all videos and sequences represent the same reef. The filming details (time, season, region) are also unknown. From Table 1, note that Video 0, with 8 sequences, contributes the least number of images, possibly due to filming smaller, more distant areas, resulting in more breaks and sequences. On the other hand, Video 0 contributes the most number of images with only 4 sequences of videos, possible due to filming larger areas at a time.

Table 1: Count of Sequences and Images per Video

Video ID	No. Sequences	No. Images
0	8	6708
1	8	8232
2	4	8561

Figure 1 illustrates the distribution of sequences across videos. Continuing from above, Video 1 has 8 sequences although more than half contain very few images, and with only two sequences contributing nearly 3000 images each. In contrast, Video 2 has only 4 sequences, but each contributes a substantial number of images, with at least 1,500 per sequence. Note the sequence labels are arbitrary so they have been excluded.

Figure 1: Images per Sequence in each Video
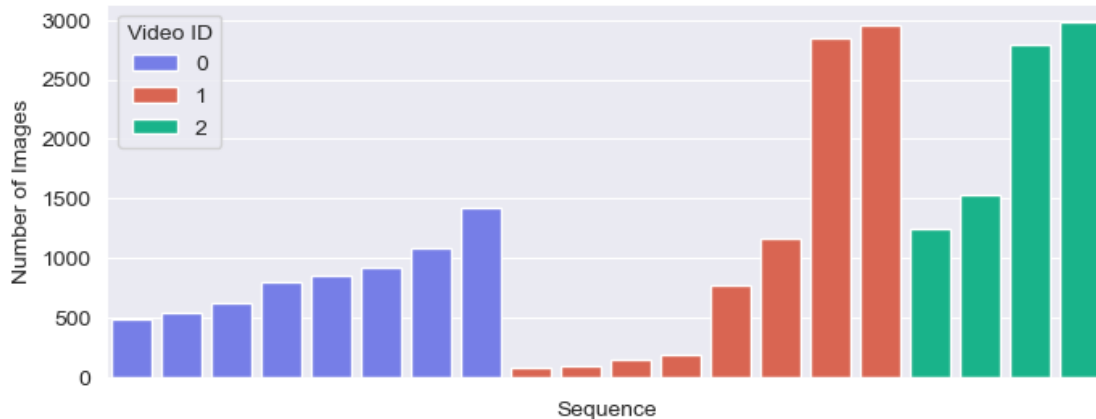
2.2 How are COTS distributed across videos and sequences?

Figure 2 shows that Video 0 has the highest proportion of COTS occurrences, with 31.9% of its 6,708 images containing COTS. In contrast, Video 2 has a larger dataset of 8,561 images but a significantly lower COTS presence at only 7.9%. This suggests that the videos may have surveyed different areas or at different times with varying COTS densities. When training the model, it may be beneficial to use weighted algorithms to give more emphasis to Video 0.
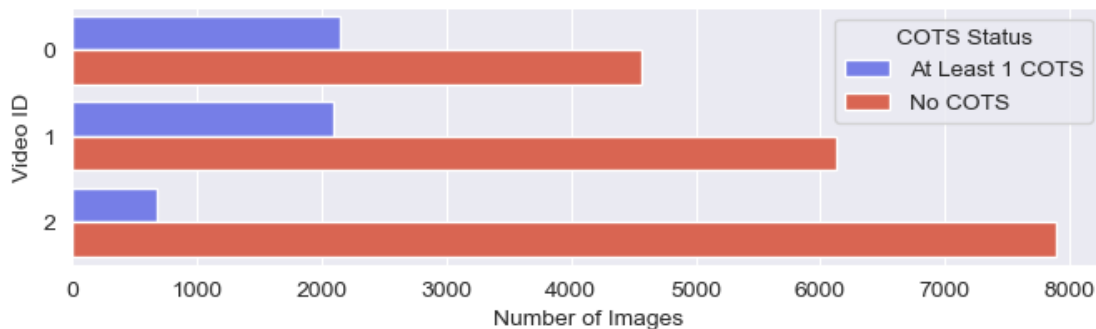
Figure 2: Images with and without COTS

Figure 3 orders images by their frame index, with gaps indicating breaks between sequences. Video 0 shows almost no COTS across all images, and those that do contain COTS have fewer than five each. Similarly, Video 1 has very few COTS overall, but when present, they often appear in high numbers (>10), which could help the model learn to detect clusters rather than individual COTS. Video 2 also has minimal COTS presence, except for what appears to be the third sequence, where a significant number of images contain more than 15 COTS, suggesting another example of high clustering of COTS.

Figure 3: Number of COTS in each Image per Video

```
<Figure size 400x300 with 0 Axes>
```
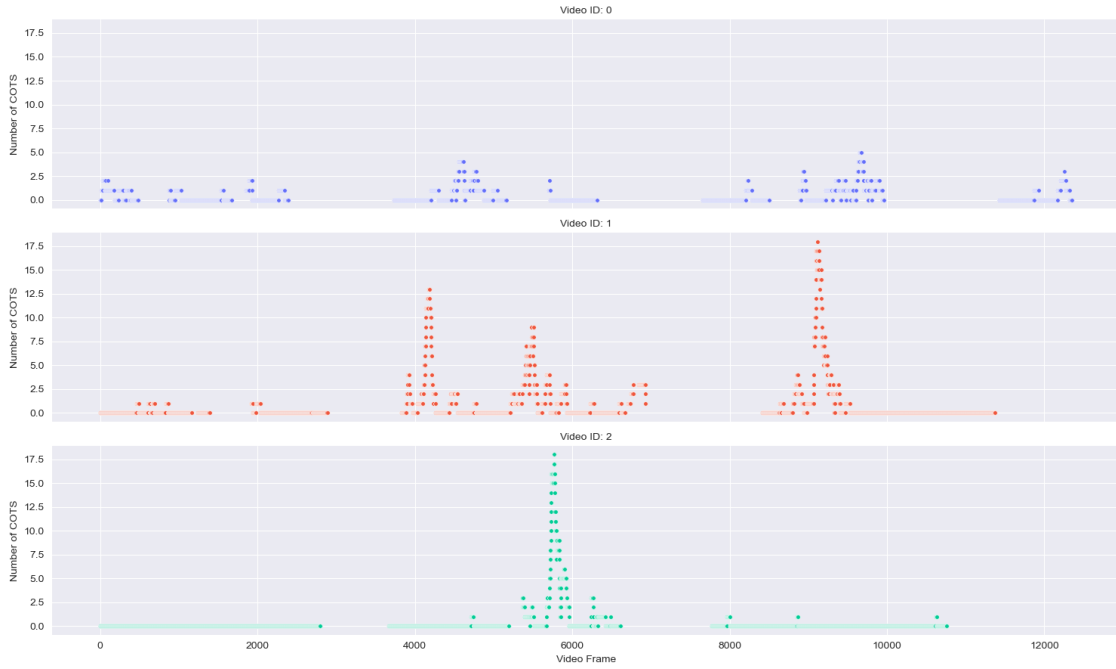

# 3 Examining Annotations & Distribution of COTS

This sections takes a closer look into bounding boxes and their distributions. Since bounding boxes are used to label COTS, they are geometrically correlated to the scale and location of the starfish and can be used to get a general idea of how the starfish present in the images.

3.1 How large are bounding boxes?

Figure 4 shows plots with outliers (bounding boxes > 10,000 pixels) excluded. In Video 0, bounding boxes typically average around 1,500 pixels, but there are more images with larger sizes compared to the other two videos. In Videos 1 and 2, bounding boxes generally range from the high 1,000s to 2,500 pixels, with very few exceeding 7,000 pixels. This aligns with the previous idea that Video 0 might represent more distinct areas and thus more variety in the sizes of the COTS captured.

Figure 4: Distribution of Bounding Box Sizes

```
<Figure size 2000x500 with 0 Axes>
```
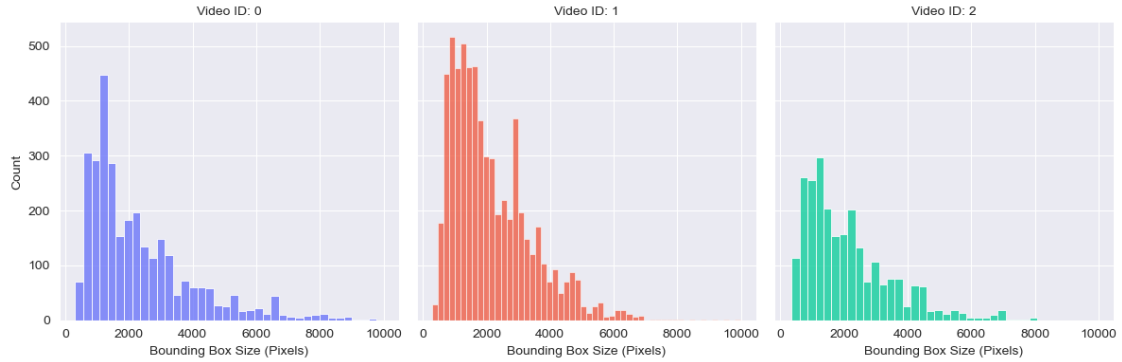

3.2 Where are bounding boxes?

Figure 5 presents a heatmap of bounding box coverage across images. In Video 0, there's a large concentration in the top-right corner, with some coverage in the center. Video 1 shows most bounding boxes concentrated around the (650, 400) pixel coordinate, with little coverage elsewhere. Video 2's boxes are more evenly spread out, but there's a slight increase in coverage in the bottom-right corner. The "string-like" structures in these plots suggest that the diver may have continuously filmed the same instance of COTS, so a series of images might represent the same COTS.

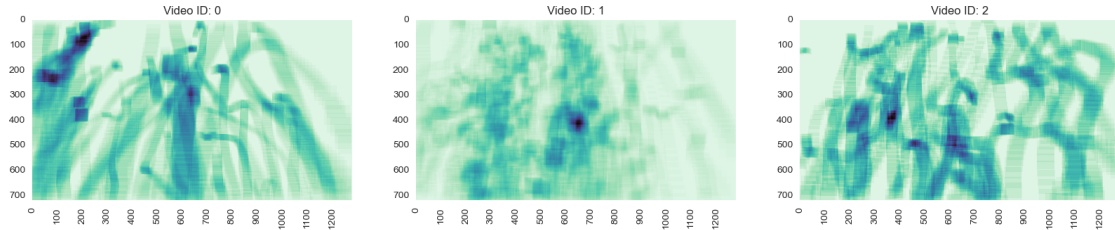Figure 5: Heatmap of Bounding Box Coverage

Figure 6 displays a scatterplot of bounding box centers, with colors distinguishing between different sequences. The "string-like" structures observed in the heatmaps from Figure 5 can be seen again, especially in Video 0 and 2. Video 1 shows many green and blue points, likely corresponding to the two sequences with over 3,000 images, as noted in Section 2.1. For Videos 0 and 2, the presence of the same sequence across multiple regions in each image suggests that a sequence may represent distinct COTS groupings or that the diver filmed the same group from different angles.

Figure 6: Scatterplots of Bounding Box Centers
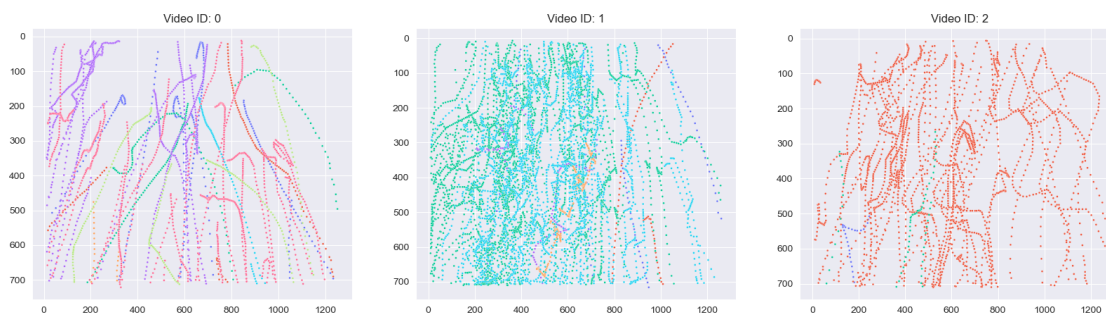

4 Exploring Image Characteristics

This section examines pixel data to determine any preliminary/distinguishing patterns.

4.1 Are there any differences in image brightness?

Figure 7 plots the average brightness of each image per video, facetted by the presence of COTS. While the brightness distributions vary significantly across videos, there is no clear indication of substantial differences in brightness based on COTS presence. However, Video 0 shows slightly lower brightness (around 120) compared to Video 1 and 2 (around 130). This could suggest that Video 0 was filmed on a darker day or that the images feature more grey reefs rather than bright blue water.

Figure 7: Distribution of Brightness Levels

```
<Figure size 500x200 with 0 Axes>
```
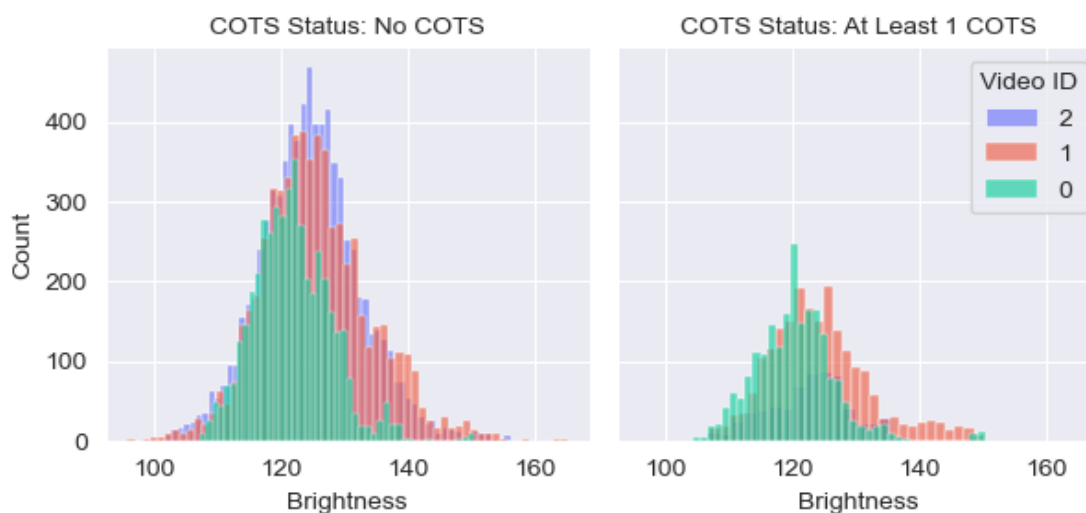
## 4.2 Are there any differences in colour intensities?

Figure 8 compares the intensity of the red, green, and blue (RGB) color channels across images with and without COTS. Ignoring the expected difference in scale due to the majority of images having no COTS, the following can be seen:

- For red, the average intensity is around 70 in images without COTS, compared to 50 in images with COTS.
- For green, the average intensity is around 150 for both types of images.
- For blue, the intensity in images without COTS is spread between 125-175, whereas images with COTS average around 175.

When training the model, it may be beneficial to emphasize red pixels since images without COTS show higher red intensity. Computing the variance of blue intensity in images could also be useful since images with COTS exhibit less variance in blue intensity.

Figure 8: Distribution of Colour Intensities

```
<Figure size 500x200 with 0 Axes>
```
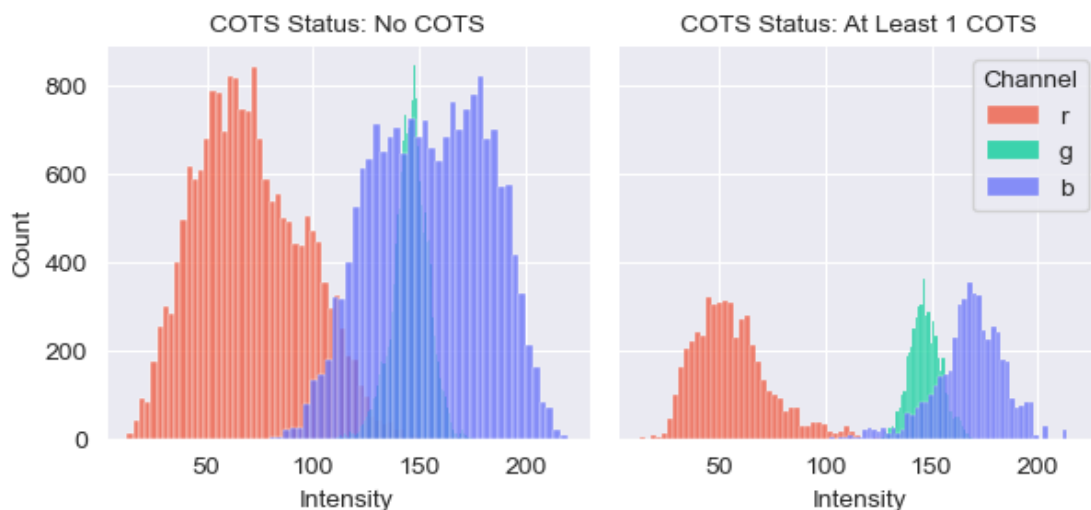

4.3 What is the average pixel value?

Figure 9 shows the "mean image," which averages each pixel across all images per video. This helps reveal the regions where COTS tend to appear, as reefs are usually grey in the dataset (and water is blue). In Video 0, the reef and COTS are concentrated in the middle-to-bottom-left of the frame. On the other hand, Videos 1 and 2 show that COTS tend to appear on the right side of both videos.

Figure 9: Images of Average Pixel Values

5 Conclusion

The above analysis of the underwater image data reveals several insights about COTS distribution and image characteristics. Video 0 has a higher proportion of larger bounding boxes but fewer images overall, with COTS clustered mainly in the left-most regions and in distinct "string-like" patterns. Video 1 and 2, while containing more images, show COTS concentrated in the central and right-most regions. The mean image analysis suggests that COTS in Video 0 appear in the middle-to-bottom-left, while in Videos 1 and 2, they are concentrated on the right side. Additionally, the RGB analysis suggests that red intensity is lower in images with COTS, while blue intensity is higher and less variable. These patterns can help to decide first steps for image processing during model development that can in turn improve COTS detection.