Data Introduction

Data Source & Description

Origin: The dataset is sourced from Kaggle, titled "Underwater Plastic Pollution Detection."

Download Link: https://www.kaggle.com/datasets/arnavs19/underwater-plastic-pollution-detection/code

The dataset consists of underwater images polluted by debris and trash discarded in marine environments. The images have been preprocessed using the Dark Channel Prior method, as described in the research paper "Single Image Haze Removal Using Dark Channel Prior" by Kaiming He, Jian Sun, and Xiaoou Tang, to enhance contrast.

File Structure

Train Directory: Contains 3,628 images with corresponding labels.

Valid Directory: Contains 1,001 images with corresponding labels.

Test Directory: Contains 501 images with corresponding labels.

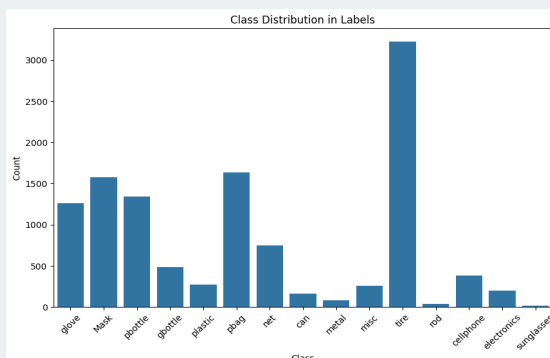## Image and Label Information

Bounding Box Coordinates: Each image includes bounding box coordinates (x-coordinate of the top-left corner, y-coordinate of the top-left corner, width, height) and class label values.

Number of Classes: The dataset includes 15 classes ['Mask', 'can', 'cellphone', 'electronics', 'gbottle', 'glove', 'metal', 'misc', 'net', 'pbag', 'pbottle', 'plastic', 'rod', 'sunglasses', 'tire'].

Image Size: All images are uniformly sized at 416x416 pixels.

Data Analysis

Label Distribution

Bounding Box Distribution

Mask	Can	Cellphone	Electronics	Gbottle	Glove	Metal	Misc

Net	Pbag	Pbottle	Plastic	Rod	Sunglasses	Tire