Project Report: Object Detection on Aquarium Dataset using Faster R-CNN

---

1. Introduction

Object detection is a fundamental task in computer vision that involves identifying and localizing objects within images. This project aims to detect marine objects in an aquarium dataset using a Faster R-CNN model with a MobileNetV3 backbone. The primary goal is to train the model to detect and classify objects, such as fish and other aquatic life, in underwater scenes. The performance of the model is evaluated using the Mean Average Precision (mAP) metric, a standard benchmark in object detection tasks.

---

2. Dataset

The dataset used for this project consists of underwater images annotated in COCO format, which contains object categories and bounding box information. The COCO format includes annotations for each object in terms of its position, category, and the corresponding bounding box coordinates. The dataset is split into training and testing subsets.

Dataset Highlights:

- **Categories**: Various marine species and objects.
- **Annotation Format**: COCO-style with bounding box coordinates.
- **Data Augmentation**: Applied to improve model robustness (e.g., flipping, brightness adjustments).
- **Train/Test Split**: The dataset is divided into a training set for model learning and a test set for evaluation.

---

3. Preprocessing and Data Augmentation

To enhance the generalization of the model, data augmentation techniques were applied during the training process. These techniques help the model learn from various image orientations and color variations, making it more robust to unseen data.

Data Augmentation Techniques:

- **Resizing**: All images were resized to 600x600 pixels for consistency.
- **Random Horizontal & Vertical Flips**: To simulate different perspectives.
- **Brightness & Contrast Adjustments**: Randomly applied to handle varying lighting conditions in underwater environments.
- **Color Jitter**: Introduced to simulate changes in hue, saturation, and contrast.

In the test set, only resizing was applied to maintain image integrity for evaluation.

---

4. Model Architecture

The object detection model used in this project is Faster R-CNN, which is a two-stage detection framework:

1. **Region Proposal Network (RPN)**: This network generates candidate regions where objects might be present.
2. **ROI (Region of Interest) Head**: Once regions are proposed, the ROI head refines the bounding box predictions and classifies the objects.

For feature extraction, a **MobileNetV3 backbone** was used, which is lightweight and efficient, making it suitable for this task.

Key Modifications:

- The pre-trained Faster R-CNN model was fine-tuned on the aquarium dataset by adjusting the classification head to match the number of object categories in the dataset.

---

5. Training

The model was trained using a custom data loader that incorporated the augmented training dataset. The learning process aimed to minimize the loss function, which includes classification, bounding box regression, and region proposal losses.

Training Setup:

- **Batch Size**: 4 images per batch.
- **Optimizer**: Stochastic Gradient Descent (SGD) with momentum and weight decay.
- **Learning Rate**: Set to 0.01, with gradual adjustments throughout the training process.
- **Epochs**: The model was trained for 10 epochs, with losses monitored to ensure optimal learning.

The model was trained using a GPU, which significantly accelerated the training process and enabled efficient computation of gradients during backpropagation.

---

6. Evaluation Metrics

To evaluate the model's performance, the **Mean Average Precision (mAP)** metric was used. mAP is a widely recognized evaluation metric in object detection tasks. It measures how well the model's predicted bounding boxes match the ground truth boxes at various Intersection over Union (IoU) thresholds.

mAP Evaluation Process:

- Predictions were generated for the test dataset.
- The predicted bounding boxes were compared against the ground truth boxes at different IoU thresholds (e.g., 0.50, 0.75, 0.95).
- The precision and recall were calculated for each category, and the average precision was computed.

Key Evaluation Metrics:

- **mAP@[0.50:0.95]**: This is the primary evaluation metric, which averages the precision across different IoU thresholds.
- **mAP@0.50**: Evaluates the model at a single IoU threshold of 50%, where predictions are considered correct if they overlap significantly with the ground truth.
- **mAP for Small, Medium, and Large Objects**: The model's ability to detect objects of varying sizes was assessed using different object scales.

---

7. Results

The Faster R-CNN model with a MobileNetV3 backbone performed reasonably well on the aquarium dataset. The model demonstrated an ability to accurately detect and classify marine objects in challenging underwater scenes. The mAP results provided a clear indication of the model's strengths and areas for improvement.

Performance Summary:

- **mAP@[IoU=0.50:0.95]**: The overall average precision across multiple IoU thresholds.
- **mAP@IoU=0.50**: A more lenient threshold, showing higher precision in detecting objects.
- **Object Size Performance**: The model handled medium and large objects effectively but struggled with smaller objects, which is typical in object detection tasks.

---

8. Challenges and Limitations

While the model performed well, there were a few challenges and limitations:

- **Small Object Detection**: The model found it difficult to detect smaller objects due to the inherent challenges of underwater imagery, such as low contrast and occlusions.
- **Data Imbalance**: Some object categories were underrepresented in the dataset, which may have affected the model's ability to generalize to those categories.
- **Limited Dataset Size**: A larger dataset with more diverse examples could improve the model's performance, especially for rare or small objects.

---

9. Conclusion

In this project, we successfully trained and evaluated an object detection model using Faster R-CNN with a MobileNetV3 backbone on an aquarium dataset. The model achieved reasonable performance as measured by mAP, demonstrating its ability to detect and classify marine objects in underwater environments. The results show that the Faster R-CNN model, when fine-tuned, can be effective in underwater object detection tasks.

---

10. Future Work

To further improve the model's performance, the following steps are recommended:

- **Hyperparameter Tuning**: Experimenting with different learning rates, batch sizes, and regularization techniques could lead to better results.
- **Model Ensembling**: Combining predictions from multiple models (ensembling) might improve accuracy, especially for difficult-to-detect objects.
- **Advanced Augmentation Techniques**: Applying more sophisticated data augmentation strategies, such as geometric transformations and simulated underwater noise, could make the model more robust.
- **Larger and More Diverse Dataset**: Collecting a more extensive and diverse dataset could help the model generalize better to unseen underwater scenes.