Project Report: Instance Segmentation for Campus Waste Detection Using Mask R-CNN

1. Introduction

Effective waste management is essential for sustainability on university campuses. Manual sorting and identification of recyclable materials can be inefficient and error-prone. This project leverages deep learning, specifically Mask R-CNN, to automate the segmentation and classification of campus waste into two primary categories: recyclable containers and food packaging. By utilizing instance segmentation techniques, we aim to enhance sorting efficiency, accuracy, and overall environmental impact.

2. Method

Architecture

The Mask R-CNN model, an extension of Faster R-CNN, is used for the task. It integrates Region Proposal Network (RPN) for object detection with a Fully Convolutional Network (FCN) for segmentation tasks. The backbone of the architecture is ResNet-50 combined with a Feature Pyramid Network (FPN), optimizing feature extraction at multiple scales.

Loss Function

Mask R-CNN employs a multi-task loss function defined as:

$L = L_{cls} + L_{box} + L_{mask}$

- **Classification Loss (L_{cls})**: Cross-entropy loss for object classification.
- **Bounding-box Loss (L_{box})**: Smooth L1 loss for bounding-box regression.
- **Mask Loss (L_{mask})**: Binary cross-entropy loss for pixel-wise mask prediction.

This combined loss function ensures simultaneous optimization of detection and segmentation tasks.

3. Dataset

Data Collection

The dataset comprises 200 images of waste items collected from the university campus. Each image is captured under varying lighting conditions, angles, and distances to ensure robust model training.

Dataset Link : https://drive.google.com/drive/folders/1cEJcowxr0p-MWaQatT97XMOKxX2V7ZCe?usp=sharing

Data Partitioning

The dataset was partitioned into three subsets:

- Training set (80%): 160 images
- Validation set (10%): 20 images
- Test set (10%): 20 images

Data Augmentation and Processing

Augmentation techniques include horizontal flipping, rotation, and scaling. Images were resized uniformly to 512x512 pixels, and annotations were formatted in COCO style, specifying bounding boxes and segmentation masks.

Normalization

Pixel values were normalized to the [0, 1] range for efficient training.

Sample Data Structure

Each training data sample includes:

```
{
  "file_name": "AACS7907.JPG",
  "image_id": 1,
  "height": 512,
```

```
  "width": 512,
  "annotations": [
    {
      "bbox": [x, y, width, height],
      "bbox_mode": "XYWH_ABS",
      "category_id": 0,
      "segmentation": [[...]]
    }
  ]
}
```


4. Experiments and Results

a) Hyperparameter Tuning

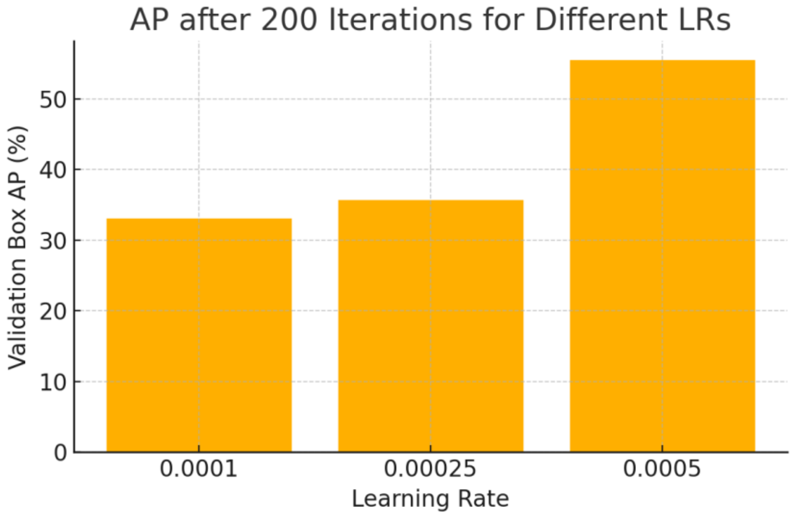
Hyperparameters tuned:

- Learning Rate: [1e-4, 2.5e-4, 5e-4]
- Iterations: 200 (initial tuning), 1000 (final training)

Best Parameters:

- Learning Rate: 5e-4
- Iterations: 1000

Performance plots show a significant improvement in validation AP at a learning rate of 5e-4.



b) Results Analysis

Pre-training Performance:

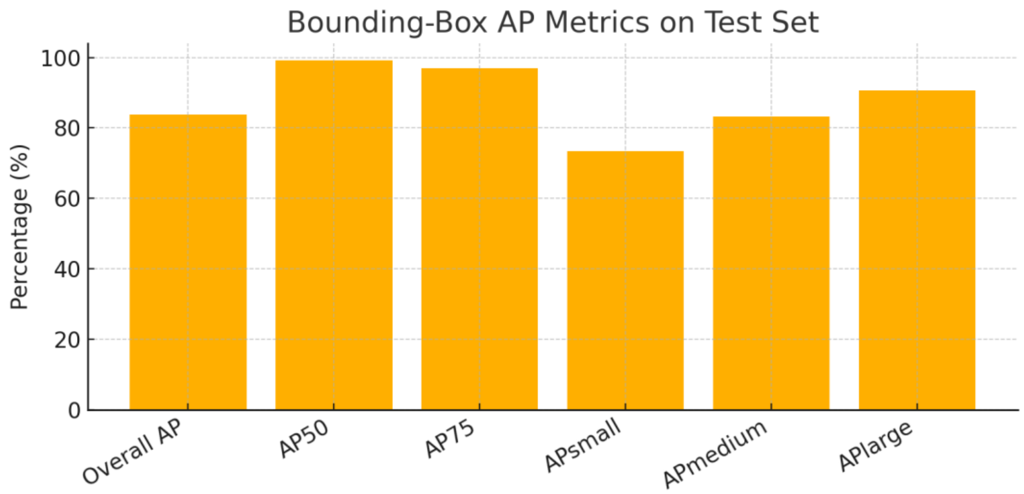
- Validation AP at 200 iterations ranged between 33.05% and 35.73% for lower learning rates.

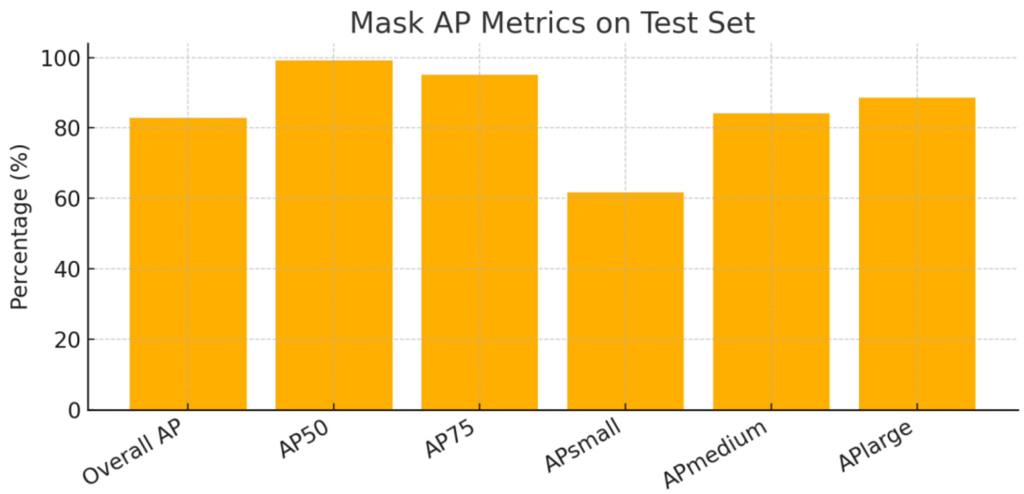
Post-training Performance: Final evaluation at optimal hyperparameters:

- **Bounding-box metrics:** Overall AP = 83.86%, AP50 = 99.13%, AP75 = 97.07%
- **Segmentation mask metrics:** Overall AP = 82.91%, AP50 = 99.13%, AP75 = 95.16%

Class-wise AP indicated balanced model performance:

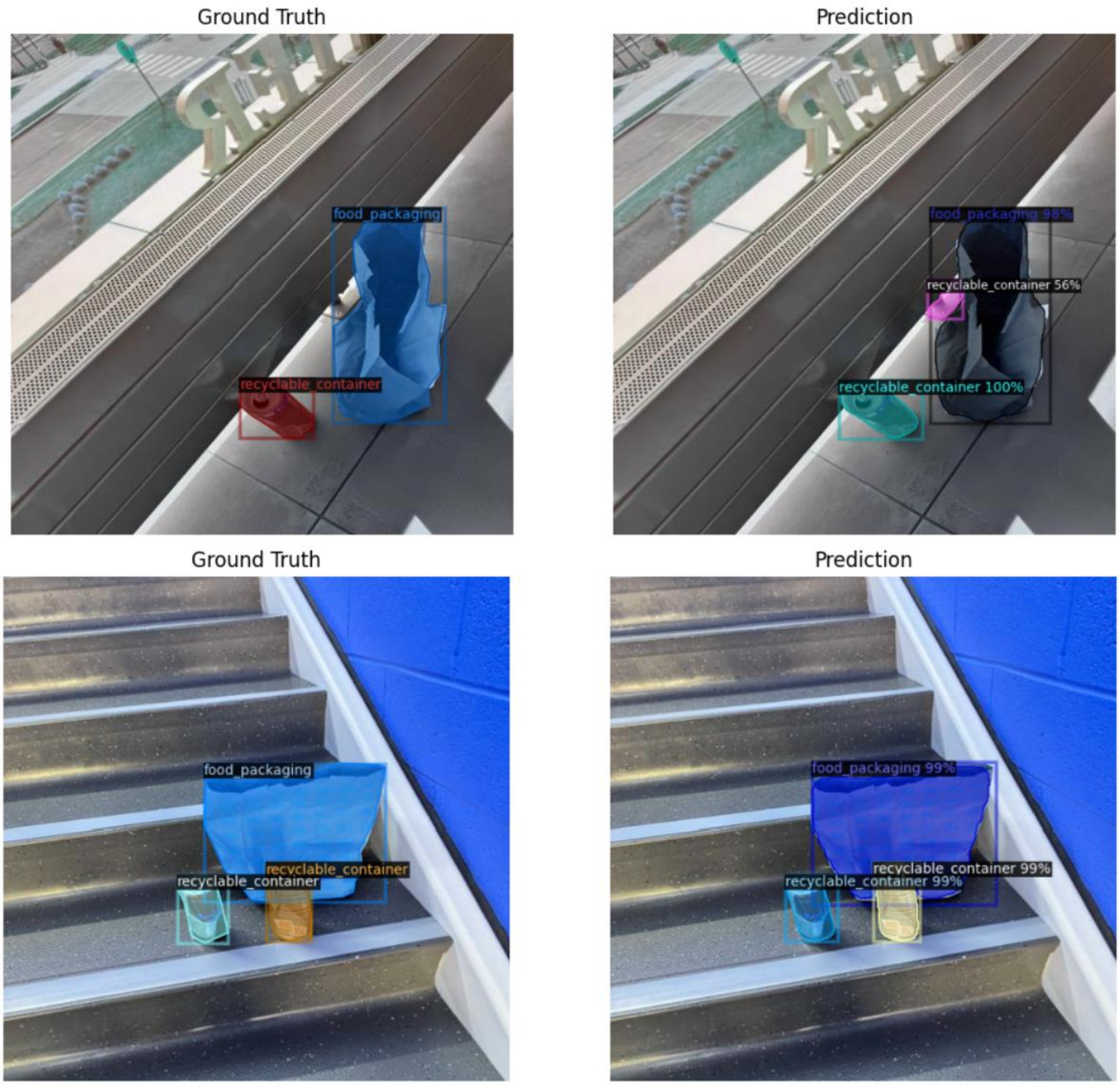
- **Recyclable container:** Box AP = 83.18%, Mask AP = 82.55%
- **Food packaging:** Box AP = 84.53%, Mask AP = 83.27%





Conclusion

The application of Mask R-CNN demonstrated strong performance for automated segmentation and classification of campus waste, significantly improving efficiency in waste management practices.


Ground Truth
food_packaging
recyclable_container
Prediction
food_packaging 98%
recyclable_container 56%
recyclable_container 100%
Ground Truth
food_packaging
recyclable_container
recyclable_container
Prediction
food_packaging 99%
recyclable_container 99%
recyclable_container 99%