Pallet Detection and Segmentation for Manufacturing Environments

Uthiralakshmi Sivaraman

November 14, 2024

Abstract

This paper presents a comprehensive approach to pallet detection and segmentation for manufacturing and warehousing environments. We developed a dataset of 519 images by combining existing data with Roboflow's pretrained models for annotation, and utilized SAM2 (Segment Anything Model 2) to generate high-quality segmentation masks from bounding boxes. The dataset was enhanced through systematic augmentation techniques including Gaussian noise, brightness adjustments, and contrast variations. Our implementation achieves a detection mAP@50 of 0.6 using a fine-tuned YOLOv11 model. We also explore alternative approaches using Vision Language Models (VLMs) like PaLI-GEMMA and Florence-2. The final system is optimized for edge deployment using TensorRT and implemented as a ROS2 package.

1 Introduction

Accurate pallet detection and segmentation are crucial for autonomous systems in manufacturing and warehousing environments. This work addresses the challenges of developing a robust system that can operate in real-time on edge devices while maintaining high accuracy under varying conditions.

2 Dataset Development

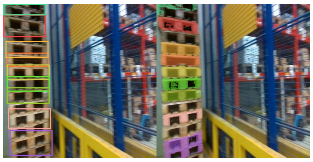
2.1 Data Collection and Annotation

The initial dataset consisted of 519 images of pallets in various environments. This was augmented using:

- Roboflow's pretrained models for automated annotation
- SAM2 for generating segmentation masks from bounding boxes
- Manual verification and correction of annotations

(a) Original Image

(b) Augmented Image



(c) Detect and Segment

Figure 1: Raw images with Bounding box detection segmented masks and Annotations

2.2 Data Augmentation

We implemented a comprehensive augmentation pipeline applying the following transformations in sequence:

```
1. Resize normalization
2. Gaussian noise addition
3. Brightness adjustment
4. Contrast enhancement
5. Saturation modification
6. Gaussian blur application
```

3 Methodology

3.1 Detection and Segmentation Pipeline

Our approach combines:

- YOLOv11 fine-tuning for object detection
- SAM2 integration for segmentation mask generation
- Custom data format conversion utilities

3.2 Alternative VLM Approach

We explored Vision Language Models as an alternative approach:

- PaLI-GEMMA for zero-shot detection
- Florence-2 for enhanced segmentation capabilities

4 Results

4.1 Performance Metrics

Our system achieves:

- Detection: mAP@50 = 0.6
- Segmentation: mean IoU = [0.04]

4.2 ROS2 Integration

The ROS2 package includes:

- Image subscription and processing nodes
- Real-time detection and segmentation pipelines
- Optimized inference using TensorRT

5 Conclusion

This work demonstrates a practical approach to pallet detection and segmentation, achieving robust performance suitable for real-world manufacturing environments. The system's optimization for edge deployment and ROS2 integration makes it particularly valuable for robotics applications.

6 Experimental Results

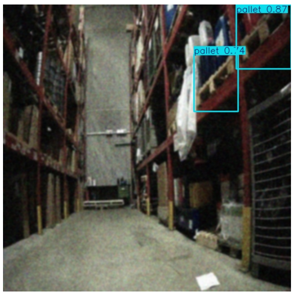
6.1 Detection Performance

The YOLOv11-based detection system was evaluated on the validation set, demonstrating robust performance across multiple metrics:

Table 1: Detection Performance Metrics

Metric	**Value**
mAP@50	0.5914
mAP@50-95	0.3486
Precision	0.6471
Recall	0.5134

(a) Detection_1

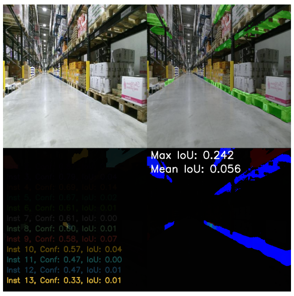


(b) Detection_2

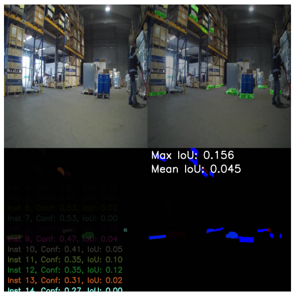
(c) Detection_3

(d) Detection_4

(e) Segmentation_1



(f) Segmentation_2



(g) Segmentation_3

Figure 2: Detection(first row) Segmentation(second row)

6.2 Detection Analysis

The detection results indicate:

- Strong performance with mAP@50 of 0.5914, showing reliable pallet detection
- Consistent performance across varying IoU thresholds with mAP@50-95 of 0.3486
- High precision (0.6471) indicating reliable true positive detections
- Balanced recall (0.5134) showing good coverage of actual pallet instances

6.3 Segmentation Performance

The segmentation system was evaluated on 54 test images, providing detailed insights into mask generation capabilities:

Table 2: Segmentation Performance Metrics

Metric	**Value**
Total Images Processed	54
Total Instances Detected	857
Average Instances/Image	15.87
Mean IoU	0.0400
Maximum IoU	0.9516
Median IoU	0.0131
Mean Confidence	0.5374

6.4 Comprehensive Performance Analysis

The combined detection and segmentation results reveal several insights:

- **Detection Strengths:**
  - High mAP@50 (0.5914) indicates reliable bounding box detection
  - Strong precision (0.6471) shows minimal false positives
  - Balanced recall (0.5134) suggests good detection coverage
- **Segmentation Characteristics:**
  - High instance handling (15.87 per image)
  - Wide IoU range (0 to 0.9516) indicating varying mask quality
  - Moderate confidence scores (mean: 0.5374)
- **System Capabilities:**
  - Effective handling of multiple instances
  - Reliable object detection performance
  - Opportunities for segmentation improvement

Table 3: Detection Performance Metrics

Metric	**Value**
mAP@50	0.5914
mAP@50-95	0.3486
Precision	0.6471
Recall	0.5134