Improve the Defect Detection Classification Efficiency in Robotic Manipulation Activity

Adwait Kadam (220084) under Prof. Tushar Sandhan and Dr. Prem Raj

Indian Institute of Technology Kanpur SURGE 2024

Abstract

Modern warehouses and factories leverage robots for tasks such as shipping, sorting, picking, and packaging. These robots must manage a variety of objects in dynamic environments, often encountering errors and defects that affect efficiency and product quality. This project aims to enhance defect detection during robotic manipulation using the **ArmBench** Dataset, which includes **42,000+** images of robotic activities captured from various angles. Key defects in the dataset include **multi-pick** and **package defects**, each with specific annotations. The baseline model, **ResNet-50**, is compared against a proposed **Vision Transformer (ViT)** model. The ViT model leverages a self-attention mechanism, proving effective in image processing tasks. Also, a **custom image dataset** of images including two different objects was prepared to train a model for image segmentation task , which would aid the **robotic-arm** in **picking** ,**dropping** and **identification** tasks.

Introduction

Objective of the project is to improve the detection of defects occuring during robot picking and dropping activities using the **ArmBench** Dataset. The Dataset comprises of images of robotic manipulation activities captured from various view points. We aim to get better performance metrics on the same dataset than the baseline.

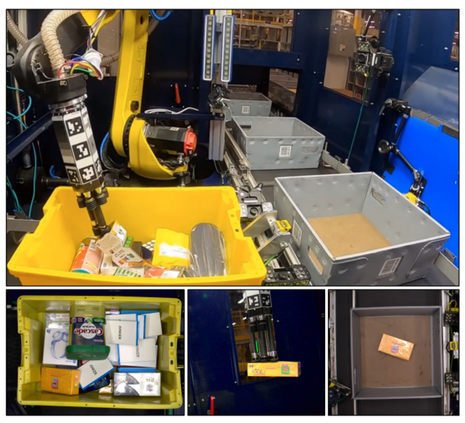

Figure 1. Robotic Activity in Amazon Warehouse

A large scale object dataset is collected using a robotic manipulation system operating in **Amazon warehouse**. The robotic arm pics one object at a time as shown in **Fig.1** from the yellow container and places it in grey tray(top). The dataset contains images for different phases of manipulation **i.e.,image of objects** in the yellow container before picking(bottom-left), during transfer (bottom-mid) and after placement(bottom-right). In addition to sensor data, the dataset also provides high quality annotations.

Research objectives

The present study investigates the following objectives:

- **Objective 1:** Improve the Defect Detection Classification Efficiency in Robotic Manipulation Activity.
- **Objective 2:** Image Segmentation Task on Custom Image Dataset for Improved Robot-Picking Efficiency in Robotic-Arm.

Dataset

The dataset comprises 42,000 images with specific defects. Two types of robot-induced defects are included in the dataset:

• Multi-pick is used to describe activities where multiple objects were picked and transferred from the source container to the destination container.

•Package defect is used to describe activities where the object packaging opened and/or the object separated into multiple parts deconstruction.

•Two subclasses, open and deconstruction, are defined for package defect.

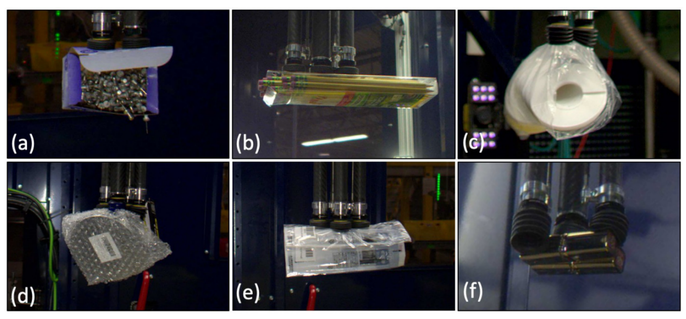

Figure 2. Various Defects Explained

Model Proposal and Training

- The baseline uses ResNet-50, This convolutional neural network is known for its effectiveness in image classification tasks due to its deep residual learning framework
- The model proposed is Vision Transformer or ViT model. Transformers, initially designed for natural language processing tasks, have shown remarkable performance in image processing tasks due to their self-attention mechanism which allows the model to focus on important parts of the image

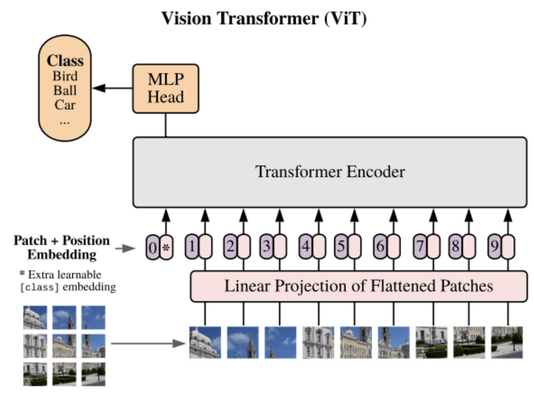

Figure 3. Cluster Visibility regions of COST2100 Model

Results and discussion

In this section, we evaluate the performance of the fine-tuned Vision Transformer (ViT) model on the test dataset. The evaluation metrics include class counts, recall, false positive rate (FPR), and the confusion matrix. We compare these results with a baseline model (ResNet-50).

Confusion Matrix

Actual\Predicted	multi-pick	nominal	package-defect
multi-pick	53.79	41.62	4.59
nominal	6.44	91.14	2.42
package-defect	7.13	19.13	73.74

Table 1. Confusion Matrix

Recall and FPR

Recall for **multi-pick 0.55**, **nominal 0.91**, **package-defect 0.73**, and **FPR** for **multi-pick 0.065**,**nominal 0.285**, **package-defect 0.027**

Model Comparison

Metric	Multi-pick	Package Defect
Recall	0.34	0.73
FPR	0.05	0.05

Table 2. Baseline Model (ResNet-50) Metrics

Metric	Multi-pick	Nominal	Package Defect
Recall	0.55	0.91	0.73
FPR	0.065	0.285	0.027

Table 3. ViT Model Metrics

Further Research on hierarchial classification is being currently implemented

Image Segmentation using YOLOv8

Figure 4. Sample Image from dataset of 100+ Annotated Images

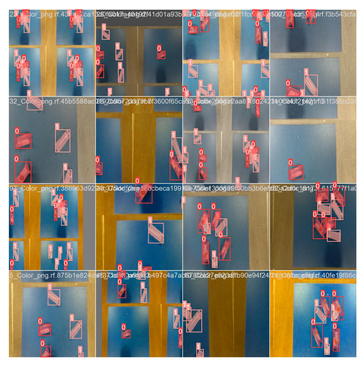

Figure 5. Train set

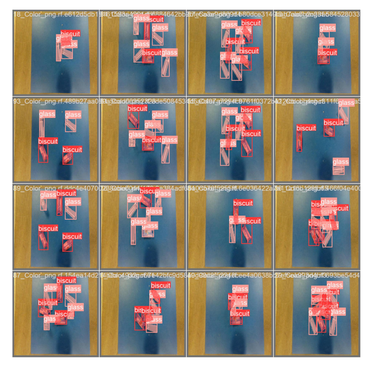

Figure 6. Validation set

Plots and Results

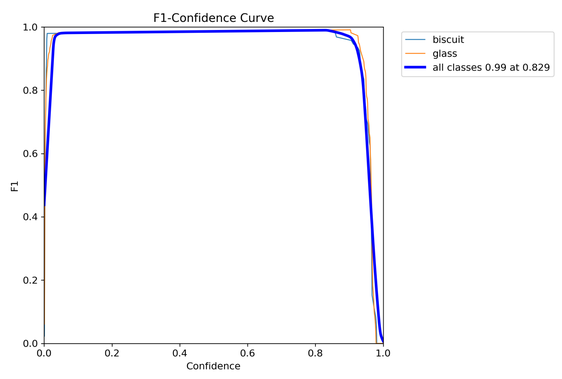

Figure 7. Box plot (segmentation via box)

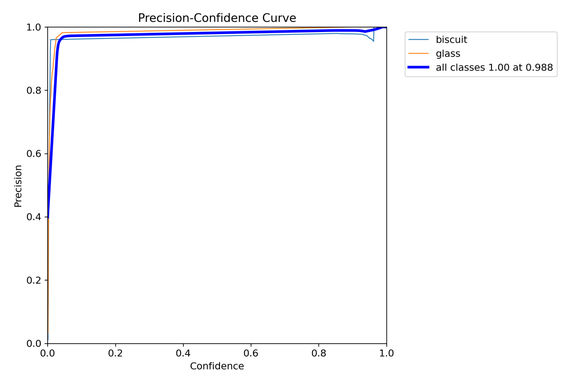

Figure 8. Mask plot (segmentation via mask)

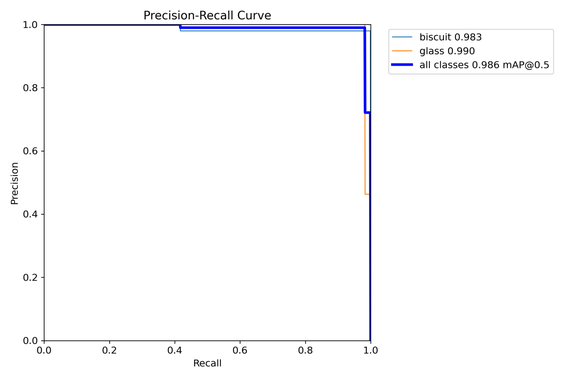

Figure 9. Box plot

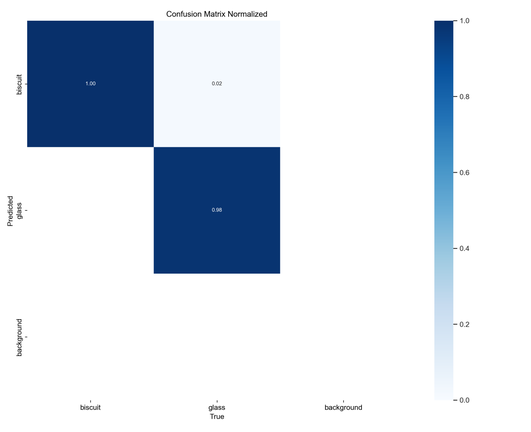

Figure 10. Mask plot

References

• A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, and others, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2021. https://arxiv.org/pdf/2010.11929v2

• Amazon Science, "Armbench: An Object-Centric Benchmark Dataset for Robotic Manipulation," http://armbench.s3-website-us-east-1.amazonaws. com/index.html, Accessed: 2024-06-18.

• R. Kondor, J. Truszkowski, Q. Li, P. Moon, J. Sharp, and P. Stone, "Armbench: An object-centric benchmark dataset for robotic manipulation," arXiv preprint arXiv:2010.11929, 2020. https: //assets.amazon.science/16/f3/35fa6104443b86fd52d82640ec94/ armbench-an-object-centric-benchmark-dataset-for-robotic-manipulation. pdf