Gugan Kathiresan

gugans.kathiresan2018@vitstudent.ac.in

7338702777

26th Dec 2020

ACTIVE SHELF MONITORING

RESULTS DOCUMENTATION

Dataset used: Mvtec d2s: Densely segmented supermarket dataset [1]

Annotation format: Bounding box object detection, COCO json

Architecture used: Faster RCNN – ResNeXt Backbone depth=101 groups=32 [2]

Environment: Python 3, Pytorch, Detectron2 toolkit [3]

Train test validation split: 1) **(Primary)** 4380:1800:1800 (Training no occlusion single object images, Validation and test had few cases of occlusion and multiple objects)

2) 70:10:10 (Training test and validation all contain occlusion, rotated, and multiple objects)

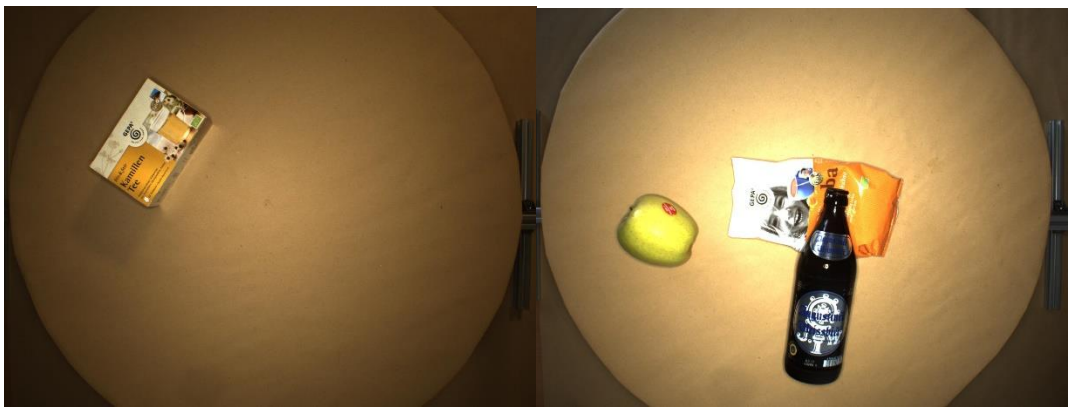
Dataset Pre-Processing

The following experiments were conducted on the Mvtec d2s dataset. Due to the large size of images present in the original dataset, which is mostly because of the various augmentations and settings of the objects, we have utilized only subsets in our experiments.

In an attempt to focus on bounding box detection of the objects only, the d2s segmentation annotations were converted to bounding box annotations using Roboflow [4]. In order to simulate different scenarios, we have utilized two different subsets of the dataset.

Our primary subset included 4380 training images, 1800 validation, and 1800 test images. In this dataset the training images did not include occluded or multiple objects. This was done as an effort to obtain and find the most precise object detection model. At the same time validation and test sets contain multi-object and occluded images as well.

This subset will from this point onwards will be referred as the "1st subset", and Figure 1 displays an example of a training and testing image sample.



a) b)

Fig 1. a) Training set image b) Test set image of 1st subset

The secondary dataset, which onwards will be referred to as the "2nd subset" employed a 70:10:10 train-validation-test split for a total of 12535 images. This set included occluded and multi-object images in the training as well as the validation and test sets.

Training Results

The training set results for varied warmup and iterations number can be seen in table 1.

Table 1. a) For 1st subset

Warmup iterations	Training iterations	Evaluation Results for BBOX
2000	5000	| AP | AP50 | AP75 | APs | APm | APl | |:---:|:---:|:---:|:---:|:---:|:---:| | 59.854 | 75.262 | 69.215 | 13.867 | 46.164 | 57.534 |

Table 1. b) For 2nd subset

Warmup iterations	Training iterations	Evaluation Results for BBOX
2000	5000	| AP | AP50 | AP75 | APs | APm | APl | |:---:|:---:|:---:|:---:|:---:|:---:| | 75.521 | 88.150 | 83.551 | 9.463 | 56.397 | 75.386 |

Testing Results

The test set results for varied warmup and iterations number can be seen in table 2.

Table 2. a) For 1st subset

Warmup iterations	Training iterations	Evaluation Results for BBOX
2000	5000	| AP | AP50 | AP75 | APs | APm | APl | |:---:|:---:|:---:|:---:|:---:|:---:| | 54.714 | 67.558 | 61.023 | 10.093 | 44.277 | 51.166 |

Table 2. b) For 2nd subset

Warmup iterations	Training iterations	Evaluation Results for BBOX
1000	2500	| AP | AP50 | AP75 | APs | APm | APl | |:---:|:---:|:---:|:---:|:---:|:---:| | 63.531 | 77.622 | 71.821 | 11.995 | 48.517 | 61.484 |
1000	5000	| AP | AP50 | AP75 | APs | APm | APl | |:---:|:---:|:---:|:---:|:---:|:---:| | 68.525 | 82.316 | 76.463 | 14.856 | 53.211 | 68.170 |
2000	5000	| AP | AP50 | AP75 | APs | APm | APl | |:---:|:---:|:---:|:---:|:---:|:---:| | 73.877 | 86.772 | 81.525 | 14.218 | 58.577 | 74.374 |

Fig 2. Evaluation results of 1st subset

Fig 3. Evaluation results of 2nd subset

---

mAP Results

The mAP results for the best performing model configurations can be seen in table 3.

Table 3.

Max mAP for Dataset	Evaluation Results for BBOX
1st subset	54.71373911715362 %
2nd subset	73.87689221213813 %

Inference

From the obtained results it can be seen that Faster-RCNN provides only moderately precise with a 55% mAP across all IOUs. But when considering only 50% of IoU the model performs significantly better with a 68% mAP.

While this study did employ 2 subsets, due to the fact that the 2nd subset included the occluded and multi-object images, the results of the 2nd subset is termed inconclusive. Only the results of the 1st subset outline the detection and feature extraction of the implemented model and thus, its results being accurate.

One of the few challenges faced during training was the need for large number of iterations in order to receive better results. The implemented model suffered from the classic issue of missed detection due to occlusion. This could be improved with the use of external augmentation.

References

[1] Follmann, P., Bottger, T., Hartinger, P., Konig, R., & Ulrich, M. (2018). Mvtec d2s: Densely segmented supermarket dataset. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 569-585).

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1137-1149.

[3] Wu, Y., Kirillov, A., Massa, F., Lo, W. Y., & Girshick, R. (2019). Detectron2.

[4] https://roboflow.com/