Product Recognition Report

A. What are the challenges and your proposed solution?

Primary challenge of the task is the lack of data. Because of this we can't just simply train a new model to automatically segment our image. The solution is to use one-shot image segmentation. By using the Siamese Mask R-CNN we can detect the target object given a reference image and perform instance segmentation on it. The primary issue is that the existing pretrained models were trained on the COCO dataset and perform poorly in this case. Solution for this is to use the One-Stage One-Shot Object Detection model to detect the objects in our scene and then run the siames model on the detected region. The detector model performs well and allows us to reduce the search area for the siamese model, which increases our performance.

If we trained the siamese model on grocery dataset, we would not need the object detector, as the model should perform well on its own.

B. What's the performance of your algorithm? How do you measure it?

The performance is very poor as the pre-trained models were trained on the COCO dataset and not a grocery dataset. The one-shot detection model performs rather well as it is able to detect most of the object. The major drawback is doing batch search. Searching only for one object class at a time yields better performance than doing batch detection.

The primary issue in the solution is the siamese mask model, looks like it is heavily coupled to the COCO dataset and has trouble generalizing for new data classes.

Generally to test the performance we would have a test dataset for which we know the correct mask. Then we just check the intersection over union metric to get the performance of the model.

C. Which annotation tool did you use? Explain why.

None. The models require a reference image of the target object for the detection. These images should have only the desired object, so the surrounding area needs to be removed. This was done manually with photoshop.

D. If you choose to use open-source vision models, what are they? Explain why.

Two models were used for this. The primary model is the Siamese Mask R-CNN for instance segmentation. It is built on top of Mask RCNN and allows it to detect and mask a target in the image by providing a reference image of the target. This model is not very performant on it's own for this use case as it tends to miss classify objects.

Another one is the One-Stage One-Shot Object Detection model which is used to detect the object and it's class by just providing one image. This model is used to reduce the search space for the mask siamese model. As this model is faster and has better accuracy than the siamese model, it allows us to prune unnecessary parts of the image and focus on the part which really contains the object and perform segmentation on it.

E. Can your solution generalize?

Yes, the solution is not strictly tied to this specific use case. The model can detect any type of object, as long as the object was in the training dataset or has similar features to a previously seen object. For example, detecting birds in the scene. Although every bird has different characteristics, the core features are the same.

F. Can your solution handle partial/full occlusion? How will you test it?

The solution can handle partial occlusion, but not total nor when two different parts of the same object are visible. The more the object is occluded the lower confidence the model has of the class of the object and the detection. This could have side effects like having the model misclassify the object.

The object detection model could learn to detect split objects, but this would require to retrain the models to handle this situation.

Generally to test for occlusion we would have a dataset in which we have a partially visible object and know the correct position and mask, this way we could compare our solutions with the groundtruth. Additionally, we could manually generate test cases where we impose an object into a scenario and remove or occlude part of the desired target object.

G. Can your solution recognize the same set of products in a different store environment? What can be improved to increase your solution robustness?

Yes, the solution is not tied to this specific store. It can easily be used to detect other objects in any other scenario, as long as there are reference images of the target object.
There are many things that can be done to include the robustness and the performance, but the most essential ones are the following:

- Improve the dataset. Adding more high quality images of the target object with different views.
- Retrain the siamese model on a grocery dataset. The model performs well on it's own, but this would drastically improve the performance. If this is done, the one-shot object detection model could be removed from the pipeline.

Repository with code:
https://github.com/Mika412/masked_test/tree/main

Siamese Mask R-CNN:
https://github.com/bethgelab/siamese-mask-rcnn

OS2D: One-Stage One-Shot Object Detection:
https://github.com/aosokin/os2d