Recyclable Waste Detection on ZeroWaste

Tao Zhang, Zeyu Gu, Zhengqi Dong, Yufan Lin
{mtao, zgu, dong760,eric1025}@bu.edu,

1. Task

Currently the issue of waste detection has drawn the public's attention. Due to a recent report by The World Bank, the waste production is predicted to be 3.4 billion tons by 2050 which would become a disaster [1]. How to deal with waste will be an inevitable problem. So, effectively and accurately classifying waste has become a solution. In this project, our group mainly focuses on the detection of waste. We will use the Zero Waste dataset, which is the largest public waste detection dataset, as a foundation to start our implementation. We aim to use computer vision and deep learning techniques to build a more efficient and accurate model for detection.

2. Related Work

The task of identifying different types of waste in images boils down to the task of object detection. This section presents some state of the art architecture used in object detection[2, 3] as well as some publicly available waste dataset.

2.1 SOTA architectures

There have been several successful architectures proposed during the last ten years that have achieved state of the art results in object detection. Here, we discuss two of the state-of-the-art architectures for object detection.

Scaled YOLOv4 Scaled YOLOv4 [3] is a redesign of the YOLOv4 network based on the CSP [5] approach and is able to be scaled both up and down to accommodate both small and large networks while maintaining optimal speed and accuracy. The CSP is a new way to architect the CNN that can save computations for various CNN networks. The authors of Scaled YOLOv4 propose a network scaling approach that can modify not only the depth, width, resolution, but also the structure of the network [3].

TridentNet Scale variance has been a major challenge in object detection. Although there are ways to solve scale variant problems, the increased inference time makes those methods less applicable to practical applications. Li et al. [2] proposed the trident blocks, which make use of dilated convolution to achieve different receptive fields. The resulting TridentNet is able to deal with the scale variation issue.

In [7], Bashkirova et al. showed that these two methods especially struggled with labeling small objects correctly, so this is one aspect we will pay attention to in our project.

2.2 Waste Detection Datasets

As automatic waste detection attracted more attention, several waste datasets [15, 16] have been created for researchers to develop better methodology to tackle the problem.

TACO TACO[15] is an open image dataset of waste in the wild. The TACO dataset currently contains 1500 annotated images with 60 classes[15]. The annotations are provided in the well-known COCO format. TACO is often used to evaluate the performance of object detection models. One downside of the TACO dataset is that it may contain some user induced errors and bias due to the crowd-sourcing nature of the dataset.

ReSort-IT ReSort-IT[16] is one of the more recent datasets created for the purpose of developing better object detection models based on deep learning. It contains 16000 synthetic images. The dataset is publicly available on Github. One downside for this dataset is its synthetic nature, which may differ a lot from real world waste site scenarios.

We discuss our choice of dataset in section 4.

3. Approach

Object detection is to detect the classes of an object and its location information in the given image. In the ZeroWaste paper, authors have applied three detection models, RetinaNet, MaskRCNN, and TridentNet, to show the feasibility of this approach. The best result was shown in the RetinaNet model with AP=24.2, AP50=36.3, AP75=26.6, AP_s=4.8, AP_m=10.7, and AP_l=26.1, which will be used as the baseline model for performance comparison in our project.

Our work will be extended based on the ZeroWaste project and push the limit to the next level. Our approaches will first mainly focus on Object Detection. After achieving a desirable accuracy and detecting speed on `ZeroWaste-f`, the fully-supervised dataset, we will then consider to other techniques or approaches to further improve the performance, which potentially can make the training to be more efficient and scalable to the real-world image dataset, such as exploring advanced model in image segmentation, solving long-tailed problem, low accuracy with small object, and laborious process of data collection and annotation.

In order to achieve this goal, there are several tasks we plan to implement:

1. Reimplement Grounded Language-Image Pre-training(GLIP) [11], which showed the best accuracy performance over most of evaluation metrics on the benchmark result trained on MSCOCO dataset [10].
2. Test and determine which is the best one-stage object detection model for our project, by leveraging the trade-off between accuracy and speed amount three models: Yolo-v4 [12], scaled Yolov4 [13], and YoloR [14].
3. Data collection: What is the accuracy we can get from those models with different throughput 15 fps, 30 fps, 45 fps, and 60 fps? Report the test/val datasize, AP, AP50, AP75, AP_S AP_M, and AP_L on both validation and testing dataset (Read here to learn more about the evaluation metrics, COCO Evaluation Metrics, A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit, Object Detection Metrics With Worked Example)
4. Compare the result we collected from a one-stage object detection model with the GLIP, and determine which one is more suitable for our project?

Extended works: After that, we will consider implementing at least one of the extended works that described below. The actual accomplishment will mainly depend on the amount of time we have left.

1. Long-tailed problems are commonly seen in object detection, which is related to the imbalanced distribution of data size on categories. Read more about this paper here, Equalization Loss v2: A New Gradient Balance Approach for Long-tailed Object Detection, to see if there is any possible solution to minimize it in our project.
2. As mentioned in the ZeroWaste paper[7], data collection and annotation is expensive and laborious. Read some relevant papers (e.g., Few-shot, or single shot object detection) to see if there is any good approach to reduce the amount of training data while maintaining a good performance in accuracy.
3. In the ZeroWaste paper [7], we noticed that small objects tend to have larger labeling errors than large objects (e.g., APs = 4.8, AP_m=10.7, AP_l=26.1). Therefore, we plan to read some relevant papers and see if there is any good approach to improve the accuracy on small objects. (e.g., Feature Pyramid Network, IPG-Net: Image Pyramid Guidance Network for Small Object Detection, or Awesome Tiny Object Detection)

4. Dataset and Metric

Our project, following Professor Kate's group's work on automated waste recycling, will be based on the industrial-grade waste detection and segmentation dataset named ZeroWaste, specifically the fully-labeled one `ZeroWaste-f`. The dataset was split into training, validation and test sets and stored in the widely used MS COCO format for object detection and segmentation using the open-source Voxel51 toolkit [7]. There are 3002 training images, 572 validation images, 929 test images in the dataset. Major classes include cardboard, soft plastic, rigid plastic and metal. It will be suitable for fully-supervised learning.

One metric to be defined as a measurement for evaluating success is the **detection precision**, which is expressed in percentage to measure the accuracy of classifying the object compared to the ground truth. Consequently, we expect to reach higher detection precision than the detection methods illustrated in [7].

Other metrics include **precision-call curve**, which is a plot of precision and recall at varying values of confidence, and **IoU**, which evaluates the degree of overlap between the ground (gt) truth and prediction (pd). Based on this metric, **Average Precision** (AP) is induced. AP@α is Area Under the Precision-Recall Curve(AUC-PR) evaluated at α IoU threshold [8] We expect our model will achieve a higher precision-call curve and average precision than the object detection models stated in [7].

5. Approximate Timeline

Make a plan with approximate deadlines, e.g.

Task	Deadline
Implemented Yolo-v4 and GLIP detector	03/11/2022
Implemented at least one items in extended works	03/18/2022
Collecting data and plot result	03/25/2022
Project Status Report Due	04/04/2022
Project Presentation and Draft Report	04/25/2022
Prepare final report	05/03/2022

References

[1] World Bank Group. "Global Waste to Grow by 70 Percent by 2050 Unless Urgent Action Is Taken: World Bank Report." World Bank, 24 Sept. 2018, www.worldbank.org/en/news/press-release/2018/09/20/global-waste-to-grow-by-70-percent-by-2050-unless-urgent-action-is-taken-world-bank-report.

[2] Y. Li, Y. Chen, N. Wang, and Z. Zhang, Scale-Aware Trident Networks for Object Detection. 2019.

[3] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, Scaled-YOLOv4: Scaling Cross Stage Partial Network. 2021.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, Mask R-CNN. 2018.

[5] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, CSPNet: A New Backbone that can Enhance Learning Capability of CNN. 2019.

[6] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. 2018.

[7] Dina Bashkirova and K. Saenko, "ZeroWaste: Towards Deformable Object Segmentation in Extreme Clutter," 2021.

[8] Koech, Kiprono Elijah. "On Object Detection Metrics with Worked Example." Medium, Towards Data Science, 18 Dec. 2021, https://towardsdatascience.com/on-object-detection-metrics-with-worked-example-216f173ed31e.

[9] Papers With Code, Semantic Segmentation Benchmarks, https://paperswithcode.com/task/semantic-segmentation

[10] Papers With Code, Object Detection Benchmarks, https://paperswithcode.com/task/object-detection

[11] L. H. Li *et al.*, "Grounded Language-Image Pre-training," *Arxiv*, 2021.

[12] C.-Y. Wang, A. Bochkovskiy, en H.-Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network", in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, bll 13029–13038.

[13] A. Bochkovskiy, C.-Y. Wang, en H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection", CoRR, vol abs/2004.10934, 2020.

[14] C.-Y. Wang, I.-H. Yeh, en H.-Y. M. Liao, "You Only Learn One Representation: Unified Network for Multiple Tasks", arXiv preprint arXiv:2105. 04206, 2021.

[15] Proença, P. F., & Simões, P. (2020). TACO: Trash Annotations in Context for Litter Detection.

[16] Koskinopoulou, M., Raptopoulos, F., Papadopoulos, G., Mavrakis, N., & Maniadakis, M. (2021). Robotic Waste Sorting Technology: Toward a Vision-Based Categorization System for the Industrial Robotic Separation of Recyclable Waste. IEEE Robotics Automation Magazine, 28(2), 50–60. https://doi.org/10.1109/MRA.2021.3066040