

Object segmentation of solid waste material for further semantic segmentation

Waseem Shaikjee & Dustin Terence van der Haar

Academy of Computer Science and Software Engineering, University of Johannesburg, South Africa,
waseem.shaikjee@gmail.com, dvanderhaar@uj.ac.za

Abstract. Recycling is crucial in reducing solid waste and pollution. An optical waste sorter can reduce the amount of solid waste pollution. To locate and segment the waste objects from the captured image data, waste object segmentation is needed. Waste object segmentation presents a complex problem due to the physical nature of solid waste in the wild. Using the TACO Dataset, this paper assesses object segmentation concerning solid waste object segmentation in the wild with engineered features. We apply a watershed and an ensemble segmentation technique and review the technique against the TACO dataset.

1 Introduction

Solid waste mismanagement leads to the pollution of natural environments, and the ecosystems within, which has a drastically negative effect on the workings of the environment and the health of all biotic life found in the environment [1, 2]. Developing countries lack the resources to effectively deal with solid waste management, which leads to the mismanagement of solid waste regarding disposal, treatment, and storage of solid waste [3]. Recycling provides an environmentally friendly method of reducing solid waste, generating jobs, and helps in the management of solid waste [3].

Lower-income countries may not have the resources and infrastructure to effectively use recycling to reduce solid waste due to lower-income generated in those countries struggle to provide even the most primary of solid waste management services [3].

Technological advances can help in creating cheap, efficient, and impactful recycling processes that remove the responsibility of separating the waste from stakeholders but are still affordable and practical for countries with limited resources. An optical sorter using computer vision is also an ideal solution; however, of its drawbacks is on determining which category of solid waste an object belongs to waste objects from the image data [4]. Real-world waste objects are deformed, fragmented, and overlap with other solid waste objects [2]. Furthermore, the variation in the state of waste objects in real life creates a complicated segmentation problem of which the visual properties of an object cannot be learned and require salient segmentation techniques [5].

This paper compares and analyses how specific segmentation techniques function with object segmentation in the context of solid waste material recognition and separation in the wild. In addition, the paper tries to understand and identify the constraints and critical factors in the problem space of solid waste segmentation by focusing on low-level features. Low-level segmentation methods provide greater insight into why object segmentation of solid

waste objects in the wild is a challenging task and to create a better understanding of the problem space for future work. Knowing which significant features can further aid and guide the body of knowledge on optical solid waste material recognition and separation and propel future research. The paper is structured in the following order, similar work, experiment setup, results, and conclusion.

2 Similar Work

In a study by Umut Ozkaya and Levent Seyfi, a garbage detection system was developed which would detect the type of recyclable material using deep learning networks [6]. The TrashNet dataset, which contains the six most common types of recycling waste was used. TrashNet is a "lab" controlled dataset, where the background, scene, and lighting of an image are controlled [7]. The findings predicted up to 97.86% accuracy using GoogLeNet + SVM, with the worst-performing network SqueezeNet + Softmax achieving 83.43% on the TrashNet dataset [6]. Furthermore, transfer learning was applied to the deep learning networks, allowing for a pre-trained network on a larger dataset and adapting it to the waste classification problem [6]. However, this study used a dataset that does not reflect solid waste material typically found in the wild. Real-life waste is disfigured, broken down, and generally found in the same vicinity as other waste products.

Object detection relies on matching sections of an image to a specific pattern of data or set of features [8]. However, solid waste objects were rarely found to be identical. Garbage is usually dirty, crumpled, misshapen, or torn apart [2]. A technique is needed to find the solid waste object(s) with incoherent features in the image. The object(s) can be extracted from the image for further material recognition. With the use of low-level features such as regions and boundary lines, the solid waste object(s) in an image can be segmented from the rest of the image data [5].

Image segmentation is categorized into three main types namely; semantic segmentation, object detection, and Instance segmentation [9, 8]. Semantic segmentation is a pixel-based segmentation technique that labels each pixel with a corresponding class label [10]. Object detection can be used as a segmentation technique by identifying the objects in an image and using it as region proposals for further segmentation [11]. Instance Segmentation is a combination of semantic segmentation and object detection techniques [12]. Instance segmentation consists of object localization, object detection, and object segmentation, with the main objective being to detect the objects in the image and correctly identify separate instances of the same type of object [10]. Object localization finds spatial object information in relation to the image. Object detection classifies the object based on trained categories that match specific patterns. Object segmentation is responsible for segmenting the objects from the rest of the image while also segmenting objects that belong to the same category into separate instances.

Zhan and Hu presented a novel Salient Object contour detection algorithm that focuses on using edges to extract the contour of the objects in the image [5]. The algorithm extracts Canny edges and uses dual image tracing to strengthen the edges and reduce the noise in the image. Canny edge detection leaves discontinuities in the edge map, leaving breaks in the boundary of an object [5]. A morphology technique is used to complete the edges and provide a complete object contour using boundary similar region calculation [5]. The salient object contour detection method for segmentation presented by Zhan and Hu showed a high accuracy, and it is efficient as well as straightforward [5]. However, the paper presented does not indicate how the algorithm will work in an image containing cluttered and overlapping objects and cases with more than one object of interest.

Semantic segmentation uses a per-pixel classification system where each pixel is given a specific class label [13]. Semantic segmentation can be applied to the different levels of

features available, either low, mid, or high-level features. Low-level features typically represent a feature space of a single dimension, such as colour, edges, and corners while mid-level features present a relationship between two or more low-level features [13]. In the scope of semantic segmentation, mid-level features are sets of interest points that can be matched to an existing set of interest points used to predict the scene based on a template [13]. This study will focus on low-level features for semantic segmentation.

Solid waste recognition is a very niche topic in the computer vision domain, and not much research is available. A greater understanding of the problem domain is needed, especially concerning waste object segmentation, due to the state waste objects found in disposal sites. Low-level features will make it easier to understand the limiting factors and variables concerning solid waste segmentation in the wild.

This paper compares and analyses how specific segmentation techniques function with low-level feature object segmentation in the context of solid waste material recognition. The main techniques covered in this paper are watershed segmentation and an ensemble segmentation technique using an ensemble of low-level features. Watershed is a region-based segmentation technique that calculates local minima and maxima. It iteratively increases the local minima until a single pixel of the local maxima remains, creating the segmentation boundary [14]. The ensemble segmentation technique searches for boundaries in the image using an ensemble of low-level features. The low-level features include Canny edges, Sobel, super-pixels, K-means clustering of the colour space, and adaptive thresholds. The low-level features are ensembled using bitwise operations on the boundary line calculations, creating strong boundary lines separating regions in the image. Contour extraction is used to extract and filter out weak region boundaries in the image.

3 Experiment setup

The salient object segmentation techniques' performance on solid waste is assessed using the TACO dataset. The TACO dataset is a subset of the COCO image dataset specifically annotated for litter detection in the wild [2]. The TACO dataset is the most suitable solid waste image dataset for solving real-world waste image segmentation. It contains 60 categories and 28 super categories. In contrast, the TrashNet data set, which is the following competing trash data set, contains only six categories of waste [2, 7]. TrashNet uses a consistent background and does not truly represent the actual state of waste found after waste collection by waste management services. This paper focuses on the TACO dataset due to its superior category split over other Trash data sets, including a wider variety of waste object types and much more differentiation in the image scene. The TACO dataset is a subset of the COCO dataset and uses the exact mechanisms and protocols as the COCO dataset. COCO is a world-renowned dataset that places focus on object detection and object segmentation and has complex scenes which allow for a more accurate representation of solid waste segmentation in the wild.

The TACO dataset used in this study is grouped into 28 super categories. Each category contains the original image and a binary mask for each waste object found in the image. Since a single image may contain several solid waste objects of varying types, the original image may be reused; however, the binary mask for the specific object focused upon will be different. Each object is then cropped from the original images and grouped by the super categories and supplemented to the original dataset. The cropping is in accordance with the work of Proenca et al., wherein their study was done to ensure that there is always a visible litter object in the dataset [2].

The watershed and an ensemble of low-level feature segmentation techniques will be applied to the TACO dataset. This will compare against a Mask R-CNN implementation by the authors and founder of the TACO dataset [2]. The primary metric used to determine the

accuracy and effectiveness of the segmentation techniques is the Intersection of Union (IoU) calculation, the ratio of overlap and union of the predicted segmentation mask, and the ground truth label.

3.1 Watershed-based segmentation

The watershed technique used in this paper is based on a variant of watershed implementation using markers [20]. The original image is converted into a greyscale image whereby the OTSU threshold is applied to the image. Erosion and dilation methods can estimate the foreground by eroding the image, and the background can be estimated by dilation. A distance transform algorithm is calculated from the estimated foreground and estimated background to obtain the markers that identify the separate region or objects in the image. The markers are used to stop over-segmentation and keep separate instances of touching or overlapping objects. The watershed algorithm is calculated using the markers providing an output of pixels describing the boundary lines, as seen in figure 1. Once the edges are calculated using the watershed method, they are extracted using Canny edges. The shapes of the regions are extracted using contour extraction and filtered based on size to reduce the amount of noise in the image. The objects can be extracted from the image for further processing with the found contours.

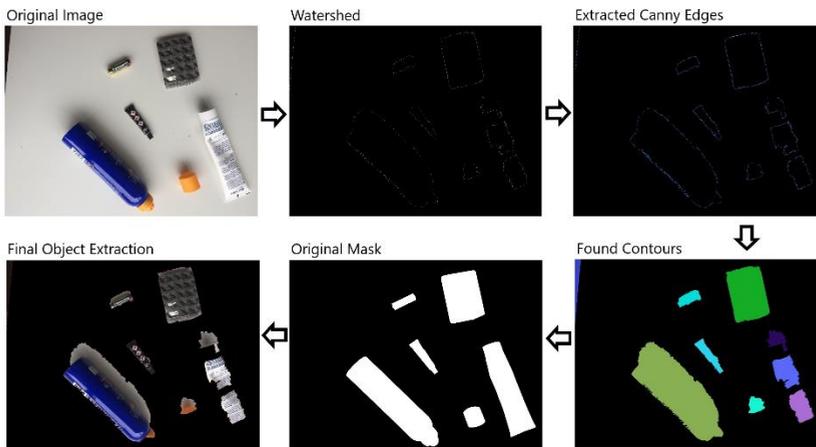


Fig. 1. Watershed Implementation

3.2 Mask R-CNN segmentation

The Mask RCNN implementation was done by Proença and Simões in their paper on the TACO dataset [2]. The implementation used Resnet50 with default weights from the COCO dataset. The network was trained for 100 epochs on the TACO dataset using SGD optimization with a batch size of 2, a learning rate of 0.001, and a learning momentum of 0.9 [2]. The input layer size covered 1024 by 1024 pixels. The paper used 4-fold cross-validation with a database split of 80% training split and 20% validation split.

3.3 Ensemble Segmentation

The ensemble segmentation algorithm uses an ensemble of algorithms to extract the most robust region and boundary lines to apply to any image variation. An ensemble of low-level

features and the process is illustrated in figures 2 - 4. The original image goes through a Min-Max normalization step to balance the pixel and colour distribution throughout the image. A greyscale transformation and an OTSU threshold (figure 2-step 2) are applied before going through the Canny edge detection step (figure 2 - step 3). This study does not apply blurring to the image before the Canny edge detection step; this is done to keep the boundary pixels calculated constant throughout the segmentation process. Bilateral filtering reduces the noise in the edge detection step (figure 2 - step 4). The Sobel edge function is performed on the image from the greyscale image (figure 2 - step 5). From the normalization step, an adaptive threshold is calculated on the image. The adaptive threshold calculates the ideal threshold value for each pixel. The adaptive threshold highlights regions around edges and helps trace edges around the regions in the image, ensuring that each region's boundary can close.

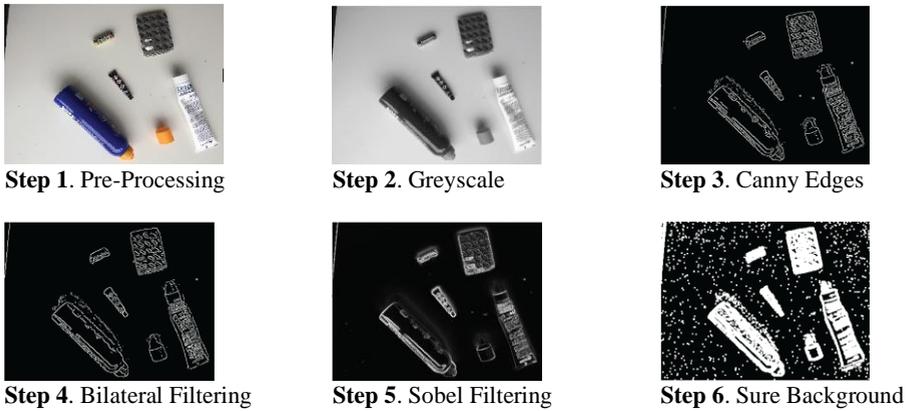


Fig 2. Ensemble segmentation steps 1 - 6

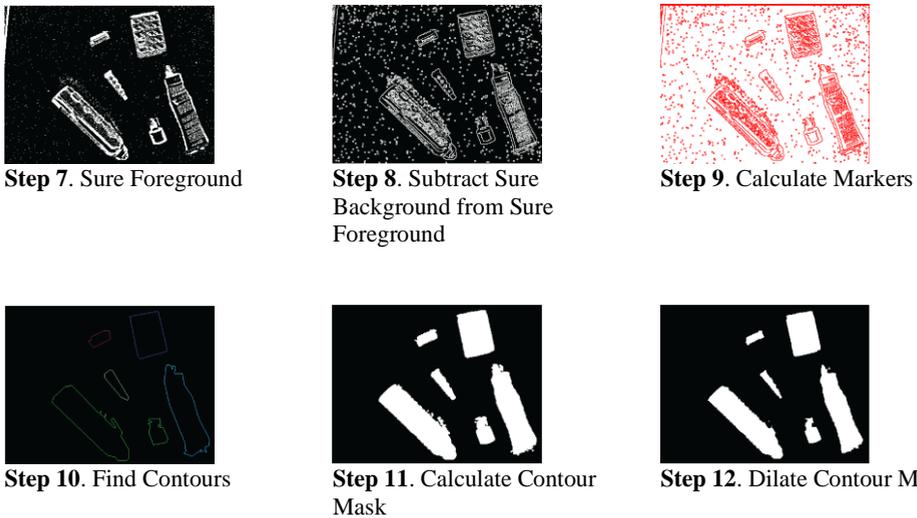


Fig 3. Ensemble segmentation steps 7 - 12

Steps 6 to 10, as depicted in figures 2 and 3, show the calculations of a boundary marker process, highlighting the area of potential boundaries between the background and foreground. Sure foreground is calculated by dilation of the output from the adaptive threshold function. The sure background is calculated by eroding the output of the adaptive

threshold calculation. The potential boundary markers are calculated by subtracting the sure foreground from the sure background. From the boundary markers, the contours are extracted and filtered.

The adaptive threshold output is through a bitwise OR operation in step 13 with the bilateral filtering output from step 4. This helps verify that the output from the bilateral filtering falls within the approximate boundaries of the objects. Step 14, illustrated in figure 4, uses the Sobel calculations from step 5 and the output from step 13 as the inputs for a bitwise AND operation. Step 14 selects only the boundary pixels that overlap the two calculations, ensuring that the strongest boundary pixels are propagated forward.



Step 13. Adaptive Threshold
| Step 4



Step 14. Step 5 & Step 13



Step 15. Dilation



Step 16. Accumulative
Thresholding



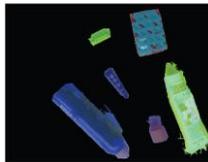
Step 17. Super Pixel | Step 15



Step 18. Step 15 & Step 17



Step 19. Final Edges



Step 20. Extracted Contours
and Objects

Fig 4. Ensemble segmentation steps 13 - 20

The super-pixels using Simple Linear Iterative Clustering (SLIC) are calculated from the original image. The super-pixels are then passed through a bitwise OR operation with the outputs from step 15. Using the super-pixels provides similar boundaries that could have been attained from clustering the image colour space and extracting the boundaries. Super-pixels allow for faster clustering based upon initialization and pre-setting the number of super-pixels. The outputs from step 15 and step 17 are joined using a bitwise AND operation that selects the strongest boundary pixels creating the final edges in step 19. Finally, the contours are extracted from the edges and allow the objects to be extracted.

4 Results

This paper conducted its experiments and analysis on the TACO dataset for a fair comparison. The results from the two proposed approaches will be compared against the Masked R-CNN detection implementation performed by Proenca et al., the creators of the TACO dataset. The results from the Mask R-CNN are tabulated in table 1. This paper applied

a watershed-based segmentation implementation and an ensemble segmentation technique using an ensemble of edge detection for strong boundary regions on the dataset. Table 2 illustrates the average IoU per superclass for each of the segmentation techniques, where the overall average IoU for the ensemble pipeline was 34%, and the Watershed pipeline was 26%.

Table 1. Average recall and precision for the Mask R-CNN TACO implementation.

Average Metric	IoU and Area	Results
Average Precision	(AP) @ [IoU=0.50:0.95]	0.294
Average Recall	(AR) @ [IoU=0.50:0.95]	0.39

Table 2. Average IoU for each class of solid recyclable waste

Average IoU - Ensemble technique	Average IoU - Watershed	Categories of waste
0.46	0.23	Aluminium foil
0.45	0.18	Battery
0.39	0.19	Blister pack
0.39	0.27	Bottle
0.29	0.22	Bottle cap
0.37	0.26	Broken glass
0.41	0.28	Can
0.43	0.32	Carton
0.25	0.31	Cigarette
0.39	0.35	Cup
0.28	0.31	Food waste
0.55	0.23	Glass jar
0.35	0.31	Lid
0.30	0.25	Other plastic
0.34	0.29	Paper
0.34	0.34	Paper bag
0.43	0.27	Plastic bag and wrapper
0.36	0.29	Plastic container
0.28	0.31	Plastic gloves
0.22	0.16	Plastic utensils
0.27	0.17	Pop tab
0.15	0.06	Rope and strings
0.22	0.06	Scrap metal
0.32	0.13	Shoe
0.37	0.18	Squeezable tube
0.16	0.10	Straw
0.34	0.29	Styrofoam piece
0.32	0.22	Unlabelled litter

Table 3. Average IoU across all classes of recyclable solid waste from the TACO dataset

Average IoU		No. of categories of waste tested.
Ensemble technique	Watershed	
0.337	0.235	28

4.1 Watershed

In table 3, it is noted that the watershed-based segmentation has the lowest accuracy when using small objects such as pop tabs, straws, and plastic utensils. The majority of the distribution of IoU lies within 50% grouping, showing that the watershed segmentation technique falls short of the majority of the dataset.

4.2 Ensemble-based segmentation

Table 3 also shows that the ensemble-based segmentation approach has the minimum accuracy when dealing with small objects such as pop tabs, straws, and plastic utensils. The outcome is similar to the watershed results and could indicate these complex segmentation problems when using low-level features that are boundary and edge-based. A large portion of the distribution of IoU lies in the 0 - 10% grouping. However, It is observed that a greater spread of the IoU moving in the upper percentile's groupings than in the watershed technique.

Compared to the Masked R-CNN detection by Proenca et al., the watershed segmentation technique has the worst-performing results. However, the ensemble segmentation approach is still competitive with the Masked R-CNN implementation found in the baseline paper [2].

5 Conclusion

Both the ensemble segmentation and watershed approach are good at determining boundary lines and edges. The difficulty lies in extracting the correct contours. Many solid waste objects have folds or creases; extracting the most external edges that encompass the entire object is challenging. The same holds when dealing with objects that overlap other objects. Another challenging task which was encountered was to differentiate between objects in the foreground and objects that are part of the background. The ensemble and the watershed approach struggle to select only the foreground objects, which reduces the IoU average. Images with less overlap with the objects and have easily distinguishable background scenes have a higher IoU percentage. The watershed approach is affected by contrast and brightness, favouring lighter regions. This does not affect the ensemble approach as it uses an ensemble of many different techniques to produce the most vital regions.

One of the biggest problems in the domain of optical sorters for waste management is the segmentation of waste material from the image data to calculate spatial relationships further and perform material recognition. Solid waste material is often broken, torn, crumpled, overlapping, and dirty. A segmentation technique capable of dealing with these

constraints is needed. Using low-level features for segmentation allows an isolation of the specific variables and factors that affect the segmentation of solid waste in the wild and provide a greater understanding of the existing body of knowledge.

This paper reviewed, compared, and analysed two low-level segmentation techniques on the TACO dataset. The watershed technique had an average IoU of 0.24%, with most of the dataset having an IoU of less than 50%. The watershed implementation struggled with smaller objects but performed better with larger objects. The watershed implementation is sensitive to contrast, brightness, and size changes. The ensemble segmentation technique had an average IoU of 34%, with a better distribution of IoU across the TACO dataset, with many falling between the range of 50 - 60%. The ensemble approach struggled with small objects but performed slightly better than the watershed approach.

There is much room for further research and expansion of the work of this paper. Many of the issues faced by the approaches discussed are the difficulty of differentiating between the potential waste objects and background. This study wishes to implement deep learning semantic segmentation approaches to help better analyse the solid waste segmentation problem and compare it against the low-level semantic approaches. A greater range of metrics can be collected from this work to help paint a better picture and help create a better understanding of the problem space.

References

- [1] F. P. Sankoh, X. Yan, Q. Tran, "Environmental and Health Impact of Solid Waste Disposal in Developing Cities: A Case Study of Granville Brook Dumpsite, Freetown, Sierra Leone," *Journal of Environmental Protection*, 4, 665–670 (2013).
- [2] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context for Litter Detection," 3 2020.
- [3] S. M. Kassim, "The Importance of Recycling in Solid Waste Management," *Macromolecular Symposia*, 320, 43–50, (2012).
- [4] L. Wang and L. Zhang, "Color Image Segmentation Algorithm Based on Affinity Propagation Clustering," in *Foundations of Intelligent Systems*, Berlin, 2012.
- [5] J. Zhan and B. Hu, "Salient Object Contour Detection Based on Boundary Similar Region," in *2012 Fourth International Conference on Digital Home*, 2012.
- [6] U. Ozkaya and L. Seyfi, "Fine-Tuning Models Comparisons on Garbage Classification for Recyclability," 8 2019.
- [7] M. Yang and G. Thung, "Classification of Trash for Recyclability Status," 2016.
- [8] J. F. Peters, *Foundations of Computer Vision*, Springer International Publishing, 2017.
- [9] C. Rasche, *Computer Vision, An Overview For Enthusiasts*, 2021.
- [10] R. Zhang, L. Jin, Y. Chen, Z. Guo, K. Zhao and Y. Zhao, "Ensembled Tricks for Instance Segmentation," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, 2020.
- [11] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 11 2013.
- [12] Y. Li and L. Ma, "A Super-pixel based Method for Instance Segmentation Post-processing," in *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2020.

- [13] A. Bassiouny and M. El-Saban, "Semantic segmentation as image representation for scene recognition," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014.
- [14] S. Beucher and C. Mathmatique, "The watershed transformation applied to image segmentation," in *10th Pfefferkorn Conf. on Signal and Image Processing in Microscopy and Microanalysis*, 1991.