Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Please see Section 4
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] Please refer to the Section 5
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A] No theoretical results
   (b) Did you include complete proofs of all theoretical results? [N/A] No theoretical results
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Section 3 for our data URL and 4 containing the link to our Github project containing the code for all the experiments
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4 as well as the configuration logs for all of our experiments on our project page
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 4
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] see Sections 2 and 4
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] project URL in Section 4 and link to the dataset in Section 3
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] mentioned tMRF's consent in Section 3
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] mentioned that we removed the identifiable information in Section 3
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [Yes] provided the annotation guidelines in Section 3
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [Yes] We included this information to the Datasheet.

A Additional Information for the Reviewers

A.1 Author Responsibility Statement

The authors of this paper confirm that they bear all responsibility in case of violation of rights. The proposed ZeroWaste data will be released under the Creative Commons Attribution-NonCommercial 4.0 International License upon acceptance.

A.2 Data Hosting and Maintenance Plan

For this submission, reviewers can download our data on the FTP server hosted by Boston University (please see the URL provided with the submission).

Upon publication, our dataset will be stored in Zenodo data repository (`https://zenodo.org/`) and will be accessible upon request according to the dataset license. The `ZeroWaste` dataset page on Zenodo is: `https://doi.org/10.5281/zenodo.4899927`. All the consecutive versions of the dataset will also be published on Zenodo that provides a persistent versioned mirror of the uploaded data.

A.3 Links to the Data

The main project page can be found here: `http://ai.bu.edu/zerowaste/`. The `ZeroWaste-f` is stored in the MS COCO format and can be read using the standard tools, such as Pycocotools. Additionally, we host a web-server for the visualization of our dataset: `http://emmy.bu.edu:5000/`. Please see Figure 6 to learn how to use the visualization tool.

The code and configurations used in our experiments can be found here: `https://github.com/dbash/zerowaste`.

DOI: `10.5281/zenodo.4899927`

Link to metadata: `https://zenodo.org/record/4899927/export/schemaorg_jsonld#.YLpUHTdKhhE`

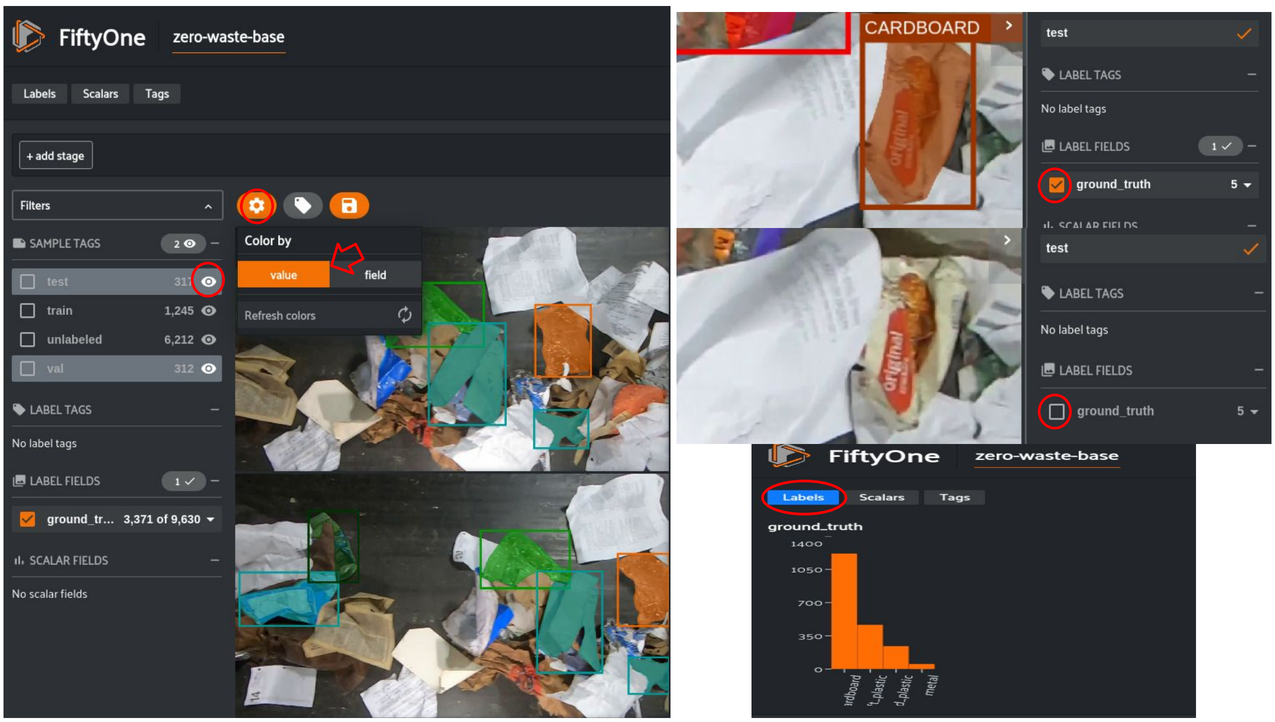


Figure 6: **Left:** To see examples from a particular split (*e.g.* test), please click the eye icon on the left sidebar of the corresponding split. To make the annotations appear in different colors depending on the class label, please select the "Color by value" option. **Right:** In order to show or hide the annotations, please select the "ground truth" option on the right sidebar after clicking on the particular image. Selecting "Labels" option in the top left corner of the page will show the class-wise statistics in the selected split. For more information, please refer to the official guide of Voxel FiftyOne.

B Appendix

B.1 Mask R-CNN experiments

Please see Table 6 for a detailed results of the experiments with Mask RCNN pretrained with COCO, and Table 5 for the TACO-pretraining results. The pretreining results on TACO dataset with `ZeroWaste` labels can be found in Table 6.

	AP *Cardboard*	AP *Soft Plastic*	AP *Rigid Plastic*	AP *Metal*	Total AP
Train	47.34	36.76	23.05	62.73	42.47
Validation	23.43	22.39	8.60	1.96	14.09
Test	28.33	13.17	17.45	0.01	14.74

Table 4: Class-wise average precision results of a COCO-pretrained Mask R-CNN on our `ZeroWaste-`*f* dataset.

	AP *Cardboard*	AP *Soft Plastic*	AP *Rigid Plastic*	AP *Metal*	Total AP
Train	41.60	34.00	20.21	60.14	39.12
Validation	27.28	23.37	6.37	6.43	15.86
Test	26.42	14.52	17.19	0.06	14.55

Table 5: Class-wise average precision results of a TACO-pretrained Mask R-CNN on our `ZeroWaste-`*f* dataset.

	AP *Cardboard*	AP *S. Plastic*	AP *R. Plastic*	AP *Metal*	AP *Other*	Total AP	AP50	AP75
Train	21.90	20.02	20.47	31.93	4.23	19.71	27.86	20.94
Test	5.12	14.96	11.90	5.79	2.91	8.14	13.25	8.55

Table 6: Class-wise average precision results of a COCO-pretrained Mask R-CNN on the TACO-zerowaste dataset.

B.2 Segmentation experiments

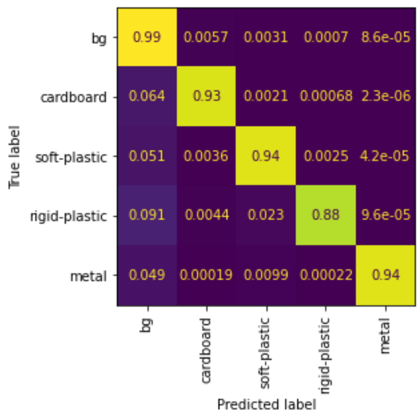
Table 7 shows the segmentation results of DeeplabV3+ for each class in `ZeroWaste` dataset. The confusion matrices of DeeplabV3+ for each split can be found on Figure 7. Please see the examples of the CCT pretdicitons in supervised and semi-supervised settings on Figure 12. Detailed DeeplabV3+ and CCT results can be found on Tables 7 and 8 respectively.

	Train			**Validation**			**Test**		
	IoU	Precision	Recall	IoU	Precision	Recall	IoU	Precision	Recall
Background	98.0	98.9	99.0	91.6	93.7	97.6	88.8	91.8	96.4
Cardboard	88.6	94.6	93.3	47.7	70.7	59.4	47.7	73.6	57.5
Soft plastic	88.6	93.6	94.2	53.8	75.1	65.5	47.3	63.1	65.3
Rigid plastic	80.3	89.9	88.2	7.8	50.7	8.4	11.4	46.4	13.1
Metal	87.1	92.2	94.0	0.0	0.0	0.0	0.1	4.7	0.2
mean	88.5	93.9	93.8	40.2	58.0	46.2	39.1	55.9	46.5

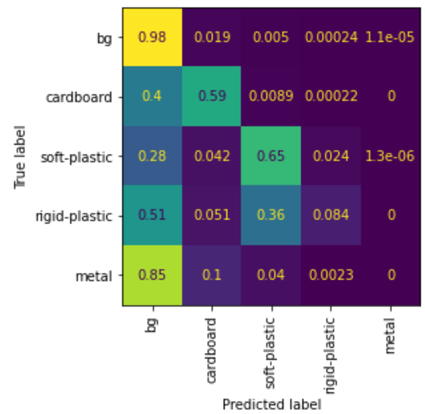
Table 7: Experimental results of DeepLabV3+ [53] on our `ZeroWaste-f` dataset.

	Background	*Cardboard*	*S. plastic*	*R. Plastic*	*Metal*	**mIoU**	**Pixel Acc.**
supervised	81.7	39.8	22.6	0.4	0.04	28.9	81.2
semi-supervised	83.3	40.3	23.2	0.0	0.0	29.4	83.6

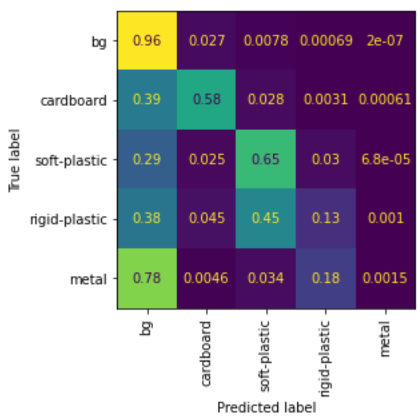
Table 8: Class-wise mIoU results of CCT on the test set of `ZeroWaste-f`. Results indicate that, while training with additional unlabeled examples slightly improves the segmentation accuracy of the most frequent classes (background and cardboard), it also results in the model misclassifying objects of the rare classes (metal).



(a) Train



(b) Val



(c) Test

Figure 7: Confiusion matrices of Deeplabv3+ on train, validation and test splits of `ZeroWaste-f`.

B.3 More data examples

The example of a frame from `ZeroWaste` before and after processing as described in Section 3 can be found on Figure 8. The examples of frames of `ZeroWaste-w` dataset of images *before* and *after* the collection of the foreground objects along with the RISE CAM results can be found on Figure 9.

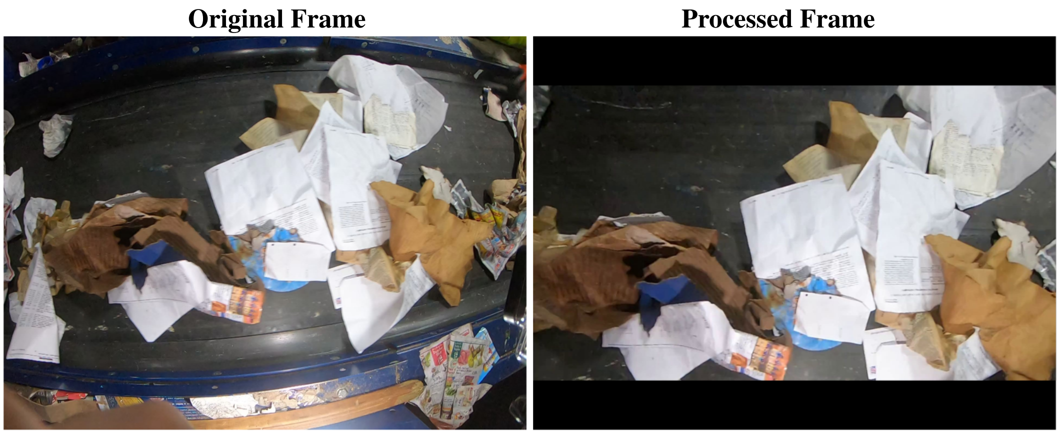


Figure 8: **Left:** sample video frame from `ZeroWaste` *before* collection of target objects to be removed from the conveyer belt. **Right:** the same video frame processed as described in section 3: fisheye effect removed, frame rotated to make the conveyor belt parallel to the image border, regions outside conveyor belt cropped out, motion blur removed.

(a)

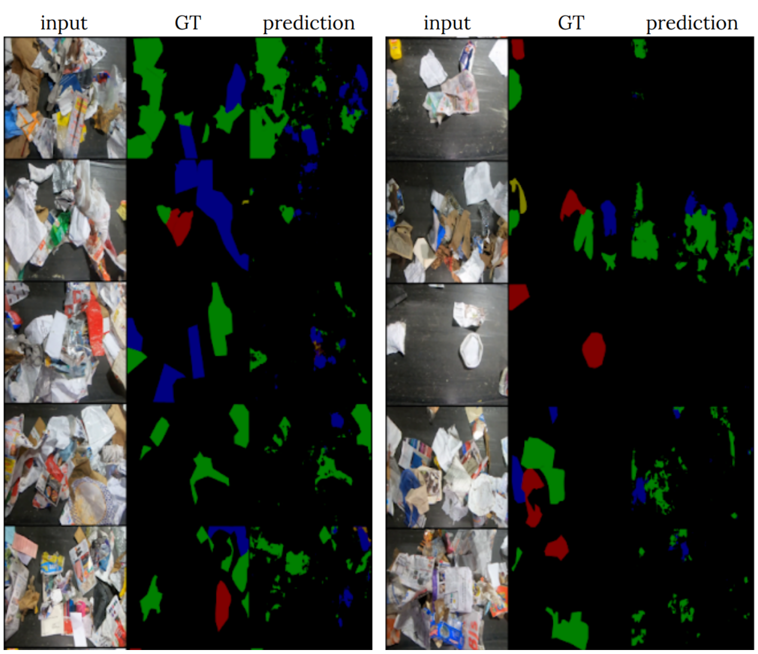
(b)

Figure 9: Examples of `ZeroWaste-w` data. **Left** examples of *After* collection frame. **Middle:** ground truth segmentation of the *Before* class example. **Right:** the corresponding CAM produced by RISE using a ResNet50 binary classifier trained on `ZeroWaste-w` data.

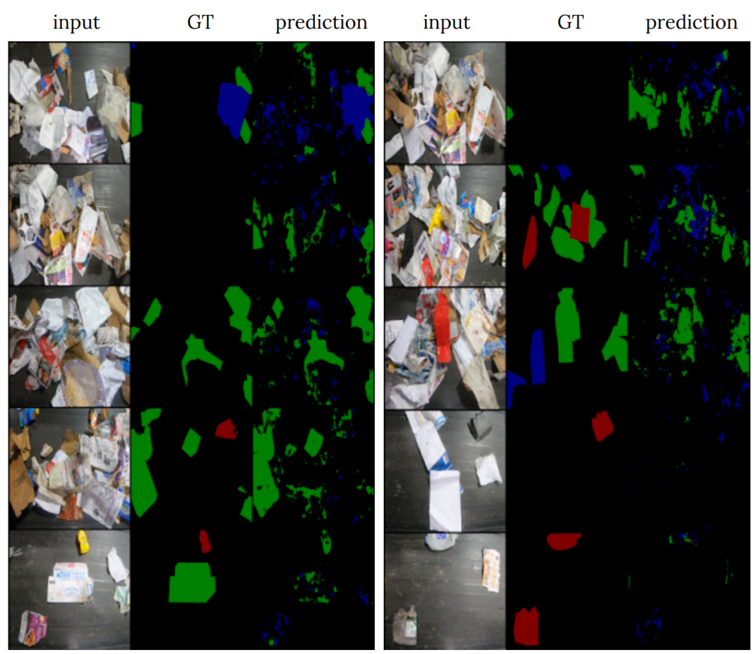
More examples of the fully annotated `ZeroWaste-f` dataset can be found on Figures 10 and 11. The illustration of the results of CCT in the semi-supervised and fully-supervised settings are shown on Figures 12a and 12b respectively.

Figure 10: Examples of images (**left**) and the corresponding polygon annotation (**right**) of the proposed `ZeroWaste` dataset.

Figure 11: Examples of images (**left**) and the corresponding polygon annotation (**right**) of the proposed `ZeroWaste` dataset.



(a) Semi-supervised results



(b) Supervised results

Figure 12: Examples of predictions of the semi-supervised (**a**) and supervised (**b**) versions of the CCT method on the images from the validation set.

C `ZeroWaste`Dataset Datasheet

C.1 Motivation

- **For what purpose was the dataset created?** The `ZeroWaste` dataset was created for the development and evaluation of detection algorithms for industrial waste sorting. While some efforts have been made by the computer vision community to create some data specifically targeted towards automated waste classification [10, 4] and detection [3, 2], the available data is very limited to the simplistic scenario of a constant background and little to no clutter. `ZeroWaste` fills this gap and introduces set of densely labeled frames of a conveyor belt from a real Materials Recovery Facility (MRF) that is specifically designed for the industrial waste sorting. We hope that our open-access dataset will push both the computer vision and the robotics communities towards more robust and data-efficient algorithms for object detection, robotic grasping and many other related problems.
- **Who created the dataset?** The dataset was collected by the Robotics Lab at Worcester Polytechnic Institute, namely James Akl and Fadi Alladkani under supervision of Professor Berk Calli, and processed and tested by the Image and Video Computing Lab at Boston University, namely: Dina Bashkirova, Ziliang Zhu and Ping Hu under supervision of Professor Kate Saenko, Professor Sarah Adel Bargal and Dr. Vitaly Ablavsky.
- **Who funded the creation of the dataset?** This paper is a part of an NSF-funded [67] collaboration project that investigates the implications of new AI and Robotics technology to MRFs and the recycling process.

C.2 Composition

- **What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** The dataset consists of a set of RGB frames from video data collected at the conveyor belt dedicated to sorting paper from objects of other material types in an MRF in Massachusetts. The polygon annotation of the objects that should be removed from the conveyor belt (objects of four classes: cardboard, soft plastic, rigid plastic and metal) are stored in JSON format according to the MS COCO [18] standard.
- **How many instances are there in total?** The fully annotated `ZeroWaste-`*f* dataset with polygon annotations contains 1874 images total: 1245 in the training split, 312 in the validation split and 317 images in the test split. The unlabeled `ZeroWaste-`*s* dataset contains 6212 images. The binary classification `ZeroWaste-`*w* dataset contains 1202 frames containing target objects (*before* class) and 1208 frames without the foreground objects (*after* class).
- **Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** The dataset contains a small fraction of the frames originally collected: we sampled every 10th frame from the beginning of the original 120 FPS videos for the `ZeroWaste-`*f* subset and every 30th frame for the `ZeroWaste-`*s* subset. The frames of *before* class are a subset of frames from `ZeroWaste-`*f* with at least one foreground object. Frames from *after* class are taken from the video after collection and are filtered so that they only contain frames with no foreground objects sampled 1 in 20.
- **What data does each instance consist of?** Each image from the `ZeroWaste-`*f* and `ZeroWaste-`*s* datasets is a frame from a video shot at the beginning of the conveyor belt (before collection). These frames contain one or multiple foreground objects (objects that must be removed from the belt, in this case anything but paper).
- **Is there a label or target associated with each instance?** Each frame from `ZeroWaste-`*f* is associated with a set of polygons (one for each foreground object) labeled with one of four class labels: cardboard, soft plastic, rigid plastic and metal. All polygons are stored in a single JSON file in the standard MS COCO format. Additionally, for each frame in `ZeroWaste-`*f* we include a semantic segmentation map with each pixel assigned the corresponding class label index as in MS COCO.
- **Is any information missing from individual instances?** There are no polygon annotations for the paper objects in this dataset since paper objects are considered background (they should stay on the belt).

- **Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?** Each frame is named in the following format: "X_frame_Y", where X is the number of the video (12 in total) the frame was taken from and Y is the frame number. Therefore, the dataset contains the information about the temporal relationship of between the frames.
- **Are there recommended data splits (e.g., training, development/validation, testing)** The `ZeroWaste-f` is split into training (1245 images), validation (312 images) and test (317) splits. We picked one sequence of 101 frames following some of the frames from the training set and two sequences of 101 images from the videos not present in the training set for the validation split, and added 10 more frames containing metal objects to compensate for the class imbalance. We chose three sequences of 101 frames that do not follow any of the frames from the training set as well as 10 frames containing metal objects to comprise the test set. We chose this setup so that all splits approximately follow the same class distribution, and to make the test split more challenging than the validation split by only containing the completely unseen data. We believe that the random sampling would not be the best choice for splitting of the frames since it would compromise the evaluation on the unseen data.
- **Are there any errors, sources of noise, or redundancies in the dataset?** The annotation was evaluated by a graduate student with common (non-expert) knowledge on the subject, and given the challenging setup it is possible that a few errors in the object labeling might occur, but the dataset has been thoroughly reviewed multiple times, so the label-level noise is possible but very rare.
- **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** Yes, the dataset is self-contained and does not rely on any external resources.
- **Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctorpatient confidentiality, data that includes the content of individuals' non-public communications)?** No, all information that might compromise the private information about the MRF employees or other individuals was excluded from this dataset.
- **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** No.
- **Does the dataset relate to people?** No.

C.3 Collection Process

- **How was the data associated with each instance acquired?** The each frame is associated with a frame from the raw RGB video collected at the MRF facility.
- **What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?** The recording apparatus is designed to fit the constraints of the facility: in order not to disrupt the MRF operation and be able to work in confined spaces available near the conveyor the recording platform needs to be compact, non-intrusive (to the workers), and portable (easy to move, battery-powered). Note that the cameras are not directly mounted on the conveyor but to a stand-alone platform, to reduce vibrations transmitted to the cameras. Additionally, (1) Damping pads are installed to counter the ground vibrations of the heavy machinery and reduce vibrations on the camera even further; (2) Weighted bases lower the center of mass to keep the apparatus stable. We used the GoPro Hero 7 for RGB footage, and we additionally collected the the near-infrared (NIR) footage simultaneously with the RGB footage using the MAPIR Survey3W NIR camera for the future work (specifically, it captures at a wavelength of 850 nm). The cameras in their encasings meet both the portability and ruggedness requirements. To maintain consistent lighting, two LitraTorch 2.0 portable lamps are installed with a light diffuser. This softens the light and spreads it more evenly in the scene. Both cameras were installed at around 100 cm above the conveyor, and the light sources at around 80 cm.
- **If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?** We deterministically

subsampled each tenth frame from the beginning of each of 12 small videos we collected for `ZeroWaste-f` dataset and every 30th frame starting after the `ZeroWaste-f` frames for the unlabeled `ZeroWaste-s` dataset. We sampled every 20th frame from the videos after *after* collection and left only the frames with no foreground objects.

- **Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?** The data collection was performed by the graduate students involved in the project. 75% of the annotations were acquired by a student researcher hired 20 hours/week for 20 USD per hour, and the rest of the data was annotated for free by the volunteering students at Boston University.
- **Over what timeframe was the data collected?** the raw video data was collected in October 2020 and processed and annotated during January 2020 - April 2021.
- **Were any ethical review processes conducted (e.g., by an institutional review board)?** No
- **Does the dataset relate to people?** No.

C.4 Preprocessing/cleaning/labeling

- **Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?**
  1. Rotation and cropping. The frames were rotated so that the conveyor belt is parallel to the frame borders and cropped to remove the regions outside the conveyor belt. We ensured that any personal information or identifiable footage of the workers at the conveyor belt was excluded from our data.
  2. Camera calibration. We removed the distortion [47] using the OpenCV [48] library to compensate for fish-eye effect caused by the proximity of the cameras to the conveyor belt.
  3. Deblurring. We used SRN-Deblur [49] method to remove motion blur resulting from a fast-moving conveyor belt. According to our visual inspection, SRN-Deblur achieves satisfactory deblurring and does not introduce the undesired artifacts that usually appear when classical deconvolution-based methods are used.
  4. Subsampling. We sampled every tenth frame from the video to avoid redundancy.
- **Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?** The raw data is stored internally on one of the machines at Boston University and can be sent upon request.
- **Is the software used to preprocess/clean/label the instances available?** we used the open-source OpenCV [48] tool for the data processing as well as the SRN [49] method (`https://github.com/jiangsutx/SRN-Deblur`) to remove motion blur.

C.5 Uses

- **Has the dataset been used for any tasks already?** We performed some initial experiments using `ZeroWaste` dataset on semantic segmentation and detection of the foreground objects. Namely, we ran a set of experiments with Mask RCNN [31] for supervised detection and transfer learning from the free-license TACO [3] dataset, DeeplabV3+ [53] for supervised semantic segmentation, and CCT [33] for semi-siupervised semantic segmentation. Additionally, we evaluated the accuracy of the class activation maps (CAMs) acquired by RISE [66] method and a ResNet50 [57] classifier trained on the `ZeroWaste-w` data for binary classification.
- **Is there a repository that links to any or all papers or systems that use the dataset?**
- **What (other) tasks could the dataset be used for?** This dataset can be used for training and evaluation of other detection and localization tasks, *e.g.* keypoint estimation, classification, as well as generative and domain adaptaion methods.
- **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** No

- **Are there tasks for which the dataset should not be used?** Although this dataset may be used for the development of a fully automated waste sorting system, authors emphasize that it was not intended for that purpose. Our goal was to improve the efficiency and safety of the waste sorting process and not to fully replace the human workers by the robots.

C.6 Distribution

- **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?** Yes, the dataset will be in the open access online for non-commercial use.
- **How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?** Upon publication, our dataset will be stored in Zenodo data repository (`https://zenodo.org/`) and will be accessible upon request according to the dataset license. The `ZeroWaste` dataset page on Zenodo is: `https://doi.org/10.5281/zenodo.4899927`.
- **When will the dataset be distributed?** the dataset will be opened upon acceptance by a peer-reviewed venue.
- **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?** The `ZeroWaste` dataset is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License [46] (please see the terms at `https://creativecommons.org/licenses/by-nc/4.0/`).
- **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** No

C.7 Maintenance

- **Who is supporting/hosting/maintaining the dataset?** The dataset is hosted and maintained by the members of the Image and Video Computing Lab at Boston University. Zenodo repository guarantees persistent accessibility to all versions of the dataset.
- **How can the owner/curator/manager of the dataset be contacted (e.g., email address)?** The primary curators of the `ZeroWaste` dataset are Dina Bashkirova (dbash@bu.edu) and Kate Saenko (saenko@bu.edu).
- **Is there an erratum?** No
- **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?** The dataset will be updated upon request of the users or once the new version of the data is available. The information about the updates and corrections will be included to the project page (`http://ai.bu.edu/zerowaste/`). All dataset versions will be available for download at `https://doi.org/10.5281/zenodo.4899927`.
- **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)?** The dataset is not related to people.
- **Will older versions of the dataset continue to be supported/hosted/maintained?** The older versions of the dataset will be available for download on the Zenodo page at any time.
- **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?** We welcome any initiatives form the research community. To propose an extension/correction of our dataset, please contact Dina Bashkirova (dbash@bu.edu) or Kate Saenko (saenko@bu.edu). All update requests will then undergo the appropriate review process prior to acceptance.