Supplementary Materials: WS^2 Dataset & Benchmark

1 Computing Binary Masks for Background Removal

We detail here the procedure used to compute the masks M_i^{bg} used for the Background Removal (BR) three-class training strategy described in Section 4 of our paper.

Let Λ be the set of classes $\Lambda = \{before, after\}$ and let $Tr^\lambda \in \{Tr^B, Tr^A\}$ be the subset of training images labeled as $\lambda \in \Lambda$. For each class $\lambda \in \Lambda$, the background estimator B^λ for λ is computed using the pixel-wise median across all images $I_i^\lambda \in Tr^\lambda$, converted in grayscale, $B^\lambda = \text{median}(Tr^\lambda)$. Then, for each image I_i^λ, M_i^{bg} is computed in the following way, $\forall p \in X_i^\lambda$:

$$M_i^{bg}(p) = \begin{cases} 0, & \text{if } \dfrac{I_i^\lambda(p) - B^\lambda(p)}{\text{MAD}(p)} < \tau_1, \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

where τ_1 is a threshold set to 0.125, p represent any pixel location in the image I_i^λ, and MAD is the Mean Absolute Deviation: $\text{MAD}(p) = \text{median}(\|I_i^\lambda(p) - B_\lambda(p)\|)$. Next, we refine M_i^{br} to retain pixels with high saturation, which are more likely to belong to objects rather than the background. Let $Z(p)$ be the saturation component of p in the HSV color space, we set $\tau_2 = 0.45$ and define:

$$\text{if } Z(I_i(p)) > \tau_2 \Rightarrow M_i^{bg}(p) = 1. \tag{2}$$

Finally, M_i^{bg} is multiplied pixel-wise by the RGB I_i^λ to set the background pixels to 0. As illustrated in Figure 1, once we obtain the binary masks to isolate the foreground for each image, we can invert these masks to create new images containing only the background. This process allows us to expand the dataset to a third set of images consisting solely of background. We experienced that this approach improves the final semantic segmentation models. To maintain class balance when training $\mathcal{K}_\theta$, we computed *background* images from half of the *before* images and half of the *after* images.

(a) (b) (c)

Figure 1: (a) Original image with the background included. (b) Image after the BR process. The BR reduces the bias introduced by the similarity of the background in the images of the same class. (c) Extracted background, used as a third independent class to help the classifier distinguish background from the objects.

This resulted in a total of 4 789 *background* images, instead of doubling the size of the original dataset. In the end, the new three-classes dataset is composed of 14 352 images, 4 712 *after* images, 4 851 *before* images, and 4 789 *background* images. To obtain purely background samples, we invert binary object masks rather than capturing images of an empty conveyor belt to avoid halting the line. By generating background-only images via mask inversion, our framework remains fully self-sufficient: it requires no manual annotators, no dedicated empty-belt recording sessions, and no modification of existing industrial processes to assemble the dataset or obtain the segmentation masks.

2 Dataset Documentation

2.1 Dataset Overview

The WS^2 dataset is a benchmark dataset specifically designed for weakly supervised waste sorting segmentation. Collected from a plastic waste sorting facility, it comprises high-resolution video sequences captured before and after a human operator manually removes unwanted items from a stream of mixed plastic waste on a conveyor belt, leaving only semi-transparent colored PET items.

The dataset is designed to facilitate the segmentation of removed waste items by leveraging the implicit supervision provided by the operator. Specifically, before images contain both items to keep and items to remove, while after images include only the items to keep. This dataset enables the evaluation of weakly supervised methods (such as CAM-based approaches) in complex industrial environments, eliminating the need for extensive pixel-level annotations required for training fully supervised segmentation networks.

The dataset consists of a training set with 9,563 unannotated before-and-after images, already organized into training and validation folders, and a test set containing 1,497 before-and-after images with corresponding pixel-level ground truth masks. The test set masks are binary annotated, distinguishing items to be removed from the rest. Images are grouped into folders containing video sequences, allowing models to leverage temporal information and enabling the evaluation of both frame-based and video-based segmentation methods. It includes both original images and background-removed versions, enabling direct comparisons between models trained with or without background information. The dataset follows a structured folder organization to support various training and evaluation setups.

2.2 Dataset Structure

The dataset is organized into the following folders:

```
train_val_dataset/
          training/
                before/
                      video_000/
                            frame_0001
                            frame_0002
                            ...
                      video_001/
                            frame_0001
                            frame_0002
                after/
                bg/ (optional)
          validation/
                before/
                after/
                bg/ (optional)

test_set/
          images/
                before/
```

```
        after/
masks/
        before/
        after/
```

2.3 Data Collection & Annotation

- **Source:** Images were captured from an industrial waste sorting plant (Seruso s.p.a.).
- **Cameras:** Two high-resolution Blackfly S BFS-PGE-200S6C cameras (5472 3648 pixels) were positioned before and after the human operator (HO) intervention.
- **Camera Setup:** Fixed on ceiling-mounted supports one meter above the conveyor belt and two meters apart from each other to ensure stability, avoid vibrations, and prevent interference with the HO's work.
- **Cropping & Image Processing:** Given the high resolution and large field of view, images were cropped to 1000 1000 pixels, ensuring they covered the full conveyor belt width while excluding the HOs workspace.
- **Belt Speed & Frame Rate:** The conveyor belt moved at approximately one meter per second, and the cameras were set to capture at 12 frames per second (fps) to maintain temporal consistency between before and after images.
- **Exposure & Gain Adjustments** Exposure time was reduced, and gain was increased to ensure clear and sharp images while minimizing motion blur.
- **File Format** Images were saved in JPEG format to optimize storage space while maintaining quality.
- **Annotations:**
  - The test set was labeled using **expert annotations**.
  - Experts manually highlighted bounding boxes around **illegal** objects.
  - These images were processed using **SAM (Segment Anything Model)** to generate precise segmentation maps.
  - The resulting masks were **manually refined at the pixel level** to create a high-quality, fully semantically segmented test set.
- **Optional Data:** Optical flow data must follow the same folder structure for POF-CAM applications.
- **Background-removed Datasets:** We provide both the original dataset and a background-removed version for training and validation.

2.4 Intended Use

The dataset is intended for:

- Weakly supervised segmentation of waste items.
- Benchmarking deep learning models for waste sorting.
- Investigating optical flow-based methods for segmentation.

Instructions for running the benchmark experiments can be found in the WS-WS-Dataset-main folder.

3 Benchmarking Code and Preprocessing

The `preprocessing_code/` folder contains scripts for preprocessing tasks:

3.1 Generating Optical Flows

- We use **SEA-RAFT** for optical flow computation. - To generate the optical flow, first download the official SEA-RAFT repository into the prepared folder `preprocessing_code/create_flows/SEA-RAFT`. - Run the following command:

```
python preprocessing_code/create_flows/create_flows.py --source_folder
   train_val_dataset/training --dest_folder optical_flows/training --
   cfg_file SEA-RAFT/config/eval/spring-M.json --cuda
```

- Modify `--source_folder` and `--dest_folder` to specify **training** or **validation** accordingly. - Optical flows are stored in the same structure as the dataset.

3.2 Removing Backgrounds

- Run the following command:

```
python preprocessing_code/remove_background/removed_bg.py --root_path /
   path/to/dataset/ --save_dir /path/to/output/
```

- This generates a background-removed dataset with the same structure as the original dataset. - To generate a dataset of **only backgrounds**, use:

```
python preprocessing_code/remove_background/removed_bg.py --root_path /
   path/to/dataset/ --save_dir /path/to/output/ --extract_bg True
```

- Be sure to rename the output folder to avoid overwriting data. - The **background class**, if used, must be organized similarly to `before` and `after` categories, inside the training or validation folders as described above.

4 Ethics Statement

4.1 Privacy & Consent

- No personally identifiable information (PII) is included in the dataset.
- The dataset exclusively consists of industrial waste images, ensuring no risk to individual privacy.

4.2 Responsible Use

- The dataset is intended for academic research and industrial waste sorting applications.
- It **must not** be used to train models for unrelated surveillance or monitoring tasks.
- Users must comply with all applicable regulations regarding environmental data.

4.3 Bias & Limitations

- The dataset focuses on plastic waste sorting and may not generalize to other waste streams.
- Variations in lighting and conveyor belt speed could impact model performance.

5 Licensing and Availability

5.1 Dataset License

The dataset is released under the **Creative Commons Attribution 4.0 International (CC BY 4.0)** license:

- **You are free to:** Share, adapt, and redistribute the dataset for any purpose, even commercially.
- **You must:** Give appropriate credit, provide a link to the license, and indicate if changes were made.
- **Restrictions:** You may not apply legal terms or technological measures that legally restrict others from using the dataset.

5.2 Long-Term Availability & Maintenance

The dataset is hosted on **Zenodo** for long-term preservation with DOI 10.5281/zenodo.14793517 at the following link: https://zenodo.org/records/14793518.

Any modifications or additions to the dataset will be documented in future releases.

6 Author Responsibility Statement

The authors bear full responsibility for this dataset in case of rights violations. The dataset has been released following ethical guidelines, and all licensing requirements are explicitly provided.