

# ParticleSAM: Small Particle Segmentation for Material Quality Monitoring in Recycling Processes

Yu Zhou\*, Pelle Thielmann\*, Ayush Chamoli\*, Bruno Mirbach\*, Didier Stricker\*<sup>†</sup>, Jason Rambach\*

\*German Research Center for Artificial Intelligence (DFKI), <sup>†</sup>RPTU Kaiserslautern  
{yu.zhou, pelle.thielmann, bruno.mirbach, didier.stricker, jason.rambach}@dfki.de, chamoli@rptu.de

**Abstract**—The construction industry represents a major sector in terms of resource consumption. Recycled construction material has high reuse potential, but quality monitoring of the aggregates is typically still performed with manual methods. Vision-based machine learning methods could offer a faster and more efficient solution to this problem, but existing segmentation methods are by design not directly applicable to images with hundreds of small particles. In this paper, we propose ParticleSAM, an adaptation of the segmentation foundation model to images with small and dense objects such as the ones often encountered in construction material particles. Moreover, we create a new dense multi-particle dataset simulated from isolated particle images with the assistance of an automated data generation and labeling pipeline. This dataset serves as a benchmark for visual material quality control automation while our segmentation approach has the potential to be valuable in application areas beyond construction where small-particle segmentation is needed. Our experimental results validate the advantages of our method by comparing to the original SAM method both in quantitative and qualitative experiments.

## I. INTRODUCTION

The construction industry is known for its high demand for raw materials, particularly aggregates such as sand, gravel, and crushed stone. According to the European Aggregates Association (UEPG) [1], the European aggregates demand is 3 billion tonnes annually. This has led to resource shortages and environmental impacts that urgently need to be addressed and resolved.

Increasing the use of recycled (RC) aggregates as a substitute for primary aggregates, particularly in the construction industry, is a cornerstone tactic for minimizing the environmental impact and achieving sustainability goals. Traditional characterization methods of recycled Construction and Demolition Waste (CDW), which often rely on manual sorting and expert visual inspection, yield delayed results and are prone to human error. The integration of image-based monitoring technologies into the production and quality inspection process of CDW has the potential to replace the labor-intensive and time-consuming manual analysis methods. This can be applied in tasks such as the estimation of the Particle Size Distribution (PSD) to reduce the cycle time of qualified RC aggregates and enhance their overall quality.

Currently, the development and application of deep learning models in the field of CDW analysis face challenges due

to the lack of publicly available open-source RC aggregates dataset. Such datasets are crucial for the training, validation, and evaluation of deep learning models. Moreover, while state-of-the-art deep learning models have demonstrated remarkable performance in standard instance segmentation tasks, their effectiveness diminishes when applied to industrial images, which are typically high resolution and often contain multi-layers and densely occluded small particles.

To address these challenges, we build upon the Segment Anything Model (SAM) [2] and propose necessary modifications to segment large numbers of small-sized particles from industrial material flow images with improved accuracy and efficiency. To encourage research in the topic, we build a benchmark high-resolution multi-layer crushed CDW particle dataset simulated from isolated particle images. Our proposed data generation engine automatically segments isolated particle instances from industrial single-layer images, refines them, and simulates augmented multi-particle images.

To summarize our main contributions:

- We generate a benchmark image dataset of multi-layer CDW particles at varying overlap levels and particle size distributions.
- We propose ParticleSAM, an adaptation of the Segment Anything Model (SAM) for segmenting images of densely placed small particles, tested for CDW, but applicable to any image dataset of similar characteristics.

## II. RELATED WORK

### A. Sensor-based Characterization of Construction Waste

Recently, sensor-based approaches have gained traction in CDW recycling, utilizing imaging, spectroscopy, and machine learning to improve the analysis of material characterization.

Near-Infrared (NIR) camera-based approach can achieve a real-time monitoring of CDW materials [3], excelling at material differentiation and performing well in low-light environments. However, NIR provides very low resolution, making it incapable of small particle size prediction. 3D laser triangulation (3DLT) sensor [4] and fusion of RGB and depth cameras [5] are integrated to acquire additional depth information except 2D data. They are combined with CNN-based classification and segmentation models, respectively. By fusing the depth information, the detection of particle size can be optimized during monitoring; however, even in a stable laboratory environment, these systems require highly precise and complex installation and calibration processes.

This work was partially funded by the German Ministry of Education and Research (BMBF) under Grant Agreements 02WDG1693D (KIMBA) and 033R390(ReVise\_UP).

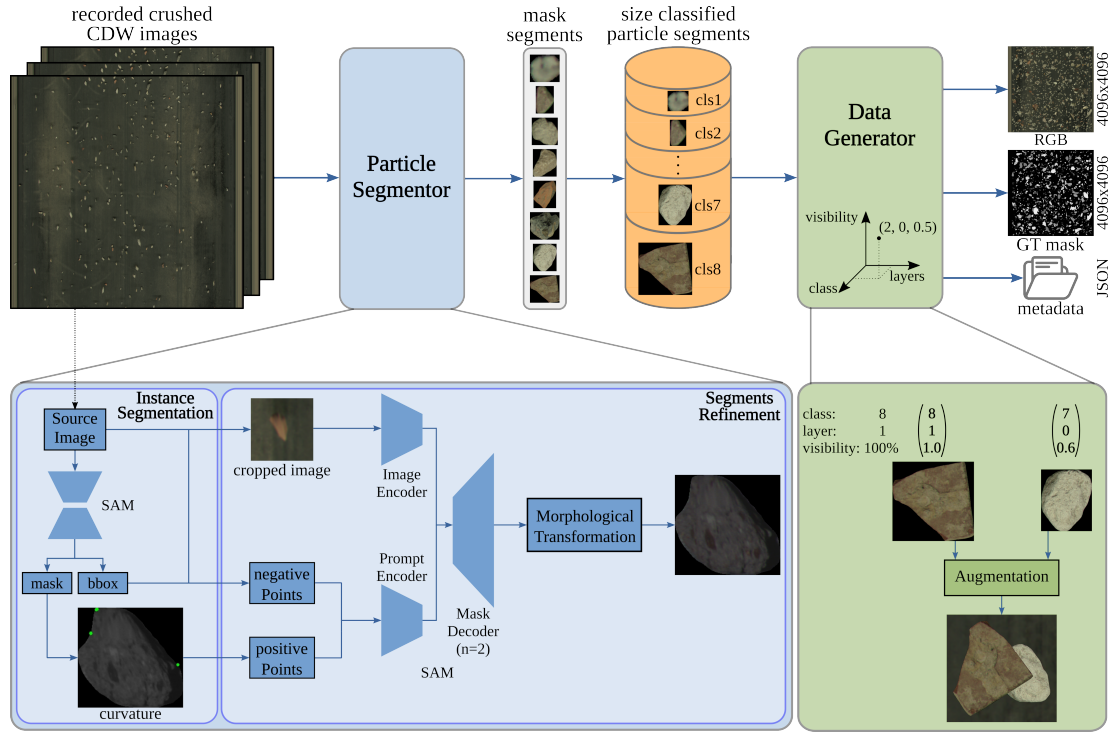


Fig. 1. Data Simulation Pipeline: The Particle Segmentor extracts isolated particles from conveyor belt recordings using SAM twice. Segmented particles are classified by size and stored for the Data Generator. The Data Generator augments particles based on class, layer, and visibility, producing high-resolution RGB images, ground truth masks, and metadata. The recorded crushed CDW image provided by Lieve Göbbels from the department of Anthropogenic Material Cycles (ANTS), RWTH Aachen University.

Methods focusing solely on RGB inputs have demonstrated that competitive performance can also be achieved by combining deep learning technologies. As in [6], they contributed an open access CDW dataset as the basis for training classification models. In [7] and [8], segmentation models are modified to provide detailed boundary and shape information, enabling more precise analysis than object detection. Nevertheless, both studies focus on initial solid construction waste, with limited contribution to the recyclability and reusability of aggregates derived from crushed CDW. In [9], the authors first introduce a synthetic datasets composed of real brick and sand-lime samples, and then re-train CNN detection models on this synthetic data to enhance their accuracy in domain-specific object detection tasks. Even so, the very low resolution (300 pixels) of their data renders it completely unsuitable for small particle segmentation.

### B. Segmentation Foundation Model

Foundation models have advanced the field of instance segmentation by training on massive large-scale datasets and enabling them adaptable to various downstream tasks with minimal fine-tuning. SAM [2] and Mask2Former [10] both employ the state-of-the-art vision transformer-based encoder-decoder architecture with attention mechanisms to achieve flexible and robust instance segmentation. Moreover, SAM is pre-trained on a vastly larger dataset compared to Mask2Former. Recent works have explored integrating SAM with downstream

segmentation tasks. For example, researchers have fine-tuned SAM with weak annotations [11], using its output to generate high-quality pseudo-labels for training lightweight instance segmentation models. Med-SA [12] is a specialized adaptation of SAM, fine-tuned on medical imaging datasets, it adapts 2D SAM to 3D medical images to achieve higher precision in medical image segmentation.

SAM is the first general foundation model for instance segmentation, capable of zero-shot transfer to a wide range of downstream tasks. Our research focuses on adapting the SAM model to enhance its segmentation capability for CDW images, addressing a critical challenge in sensor-based analysis of recycled aggregates.

## III. METHODOLOGY

### A. Data Simulation Pipeline

Up to now, research conducted on CDW datasets [13] has mainly focused on raw CDW, which serves as the input material for processing plants during recycling. In contrast, our study shifts the focus to processed CDW aggregates, which undergo particle size sorting before being distributed to construction industries as RC materials. As mentioned above, there is currently no such open source dataset. Kronenwett et al. [9] represents a first step in this direction, but their dataset is limited in particle size and type diversity and contains low-resolution synthetic images.

Our idea has been to generate augmented multi-particle images with varying overlap and size distribution. Fig. 1 shows the full data simulation pipeline. We first extract isolated particles from single-layer image recordings with a segmentor, and then use a generator to automatically output augmented images after a size classification.

**Particle Segmentor:** Inspired by SAM’s data engine, we design a Particle Segmentor (see Fig. 1), which reconfigures the two encoders and decoders of SAM to achieve automatic instance segmentation and segments optimization. Initially, we segment particles from images of crushed CDW, which were captured while the aggregates were being transported on a conveyor belt. In industry environments where the lighting conditions are often suboptimal, segmentation faces significant challenges due to shadows and motion blur. To address these issues, we add a refinement module: corner points from bounding box and curvature points from the mask output are extracted as negative and positive points, respectively, serving as input of the prompt encoder. Subsequently, morphological transformations are applied to optimize segmentation masks by removing noise, filling gaps, smoothing edges, and connecting broken regions.

**Particle size Classification:** According to the manual sieve analysis defined by DIN 66165-1 suggested by [14], we classify the obtained particles into 8 classes from 4 mm to 63 mm based on their farthest pair distance (see Table I).

**Data Generator:** To facilitate the efficient generation of large-scale, high-resolution annotated RGB images, we developed an automatic Data Generator. This approach eliminates the need for time-consuming manual annotation and ensuring higher ground truth accuracy. The data generator performs random augmentations to particles, like flipping, rotation and colorization, while avoiding scaling to preserve the authenticity of size and shape. The augmented particles are then randomly placed onto a 4096x4096 pixels conveyor belt background, simulating various occlusion conditions to enhance both robustness and realism. To overcome the storage bottleneck associated with classical binary masks as ground truth, we consolidate all mask annotations in a single portable grayscale (PGM) format image.

**Multi-stage dataset:** We construct our 4096x4096 pixels image data in 3 stages. The L1 images consist of instances from a single class, where instances are dropped to cover the entire conveyor belt background as much as possible without any overlapping. The L2 images build upon L1 by introducing occlusion, while ensuring that the visibility of each individual object remains within the range of [60%, 100%], effectively simulating real industrial scenarios. The L3 images further advance to mixed-class, where each image contains instances from different classes. In CDW crushing processes, smaller particles tend to settle to the bottom due to conveyor belt vibrations, while bigger particles remain on the top. To replicate these real-world conditions, we introduce a multi-layer structure to L3 (see Table I), while keeping particles in each layer to maintain a visibility range of [60%, 100%]. Since PSD (Particle Size Distribution) is a key criterion for

TABLE I  
CATEGORIZED SEGMENTED INSTANCES AND LAYERS

| Class   | Size Range [mm] | # Particles | Layer   |
|---------|-----------------|-------------|---------|
| Class 1 | 4.0–5.6         | 486         | Layer_0 |
| Class 2 | 5.6–8.0         | 637         |         |
| Class 3 | 8.0–11.2        | 403         |         |
| Class 4 | 11.2–16.0       | 348         | Layer_1 |
| Class 5 | 16.0–22.4       | 333         |         |
| Class 6 | 22.4–35.0       | 185         | Layer_2 |
| Class 7 | 35.0–45.0       | 106         | Layer_3 |
| Class 8 | 45.0–63.0       | 8           | Layer_4 |

analyzing RC materials, we additionally annotate the PSD for all stage images to facilitate more precise analysis. To enhance data diversity and better align with real-world conditions, we incorporate multiple PSD distributions, including uniform distribution, Gaussian distribution, and random distribution. Table II provides an overview of all stages.

### B. Adapting SAM for Small Particle Segmentation

Despite SAM’s strong performance in general instance segmentation, it struggles with complex scenes, especially for small or overlapping objects. Its reliance on prompt inputs limits full automation, posing challenges for our application. By refining its encoder-decoder architecture with multi-stage simulated data, we enhance its ability to segment dense, small particles in high-resolution images.

We adopt SAM’s image encoder and mask decoder to develop a real-time automated segmentation model for aggregates names **ParticleSAM**. First, unlike the minimum-mask area merging approach in SAM, we introduce a maximum-area filtering to remove the impact from the conveyor belt background and improve inference speed, particularly for segmenting fine-crushed particles. Next, we focus on two aspects of adaptation:

**Feature Extraction Enhancement:** In SAM’s everything demo, features are extracted using a 32x32 grid of sampling points. To enhance feature representation of our 4096x4096 pixel images, we increase this to a 64x64 grid. Additionally, we use the image splitting functionality to split input image into four crops, allowing the model to extract local features before final feature fusion. While this approach enriches features encoding, it also increases inference time by over 5

TABLE II  
DETAILS OF DIFFERENT EVALUATION DATASETS

| Data | # Images         | Visibility of Particles | # Particles per Image | Class  | Layers | Occlusion |
|------|------------------|-------------------------|-----------------------|--------|--------|-----------|
| L1   | 115 <sup>a</sup> | 100%                    | [93, 3525]            | single | single | none      |
| L2-1 | 159 <sup>a</sup> | 60%–100%                | [195, 500]            | single | single | low       |
| L2-h | 121 <sup>a</sup> | 60%–100%                | [195, 4831]           | single | single | heavy     |
| L3-0 | 79 <sup>a</sup>  | 0%–100%                 | [162, 1983]           | mixed  | single | low       |
| L3-m | 239 <sup>a</sup> | 0%–100%                 | [698, 3125]           | mixed  | multi  | medium    |
| L3-h | 640 <sup>a</sup> | 0%–100%                 | [698, 6251]           | mixed  | multi  | heavy     |

<sup>a</sup>Resolution of 4096x4096

times. To achieve real-time monitoring in the application, we choose the smallest backbone of SAM (ViT-B) and no sub-crops to form our ParticleSAM.

**Postprocessing Optimization:** We tune the post-processing parameters, including Non-Maximum Suppression (NMS), Intersection over Union (IoU) thresholds, and stability constraints, to efficiently remove padding and filter duplicate masks, thereby enhancing the decoder’s efficiency. This optimization improves segmentation precision by reducing false positives and ensuring greater consistency, making the results better aligned with our industrial application.

Last, to systematically optimize the model’s performance, we use a progressive hyperparameter tuning strategy on our simulated data. An adaptation subset of images is randomly selected from both the L1 and L2 data. The tuning process begins with the subset of no-occlusion data (L1), and parameters are optimized in descending order of particle size. Similarly to the L1 phase, adaptation starts with the largest class in L2 and gradually moves to smaller classes. Fig. 2 depicts the different adapted results.

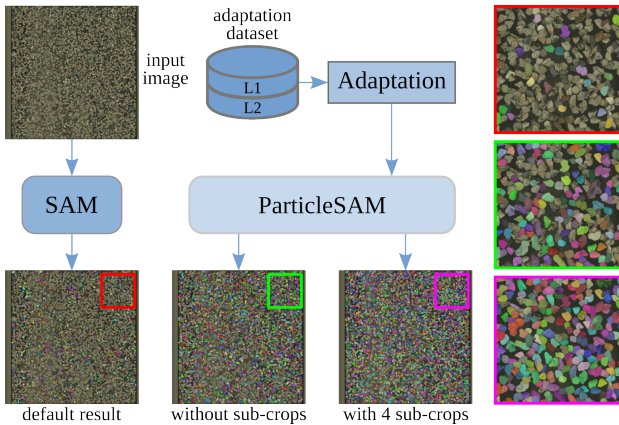


Fig. 2. Overlay of input images with colored segmentation masks showing the improvement of particle segmentation of ParticleSAM compared to SAM with 4096x4096 image data. Images on the right are zoomed-in crops. The source conveyor belt background provided by the department of Anthropogenic Material Cycles (ANTS), RWTH Aachen University.

#### IV. EXPERIMENTS

**Metrics:** We report the mean intersection-over-union (mIoU) metric to quantify the accuracy of particle segmentation. We also use the mean average precision (mAP) metric with thresholds to evaluate the model’s ability to recognize particles, e.g.  $mAP_{50}$  considers a segmentation is correct when the IoU between the predicted and the groundtruth masks is larger than 50%.

**Baseline:** Since we are not aware of other particles segmentation research for crushed waste, we leverage SAM’s Automatic Mask Generator (AMG) as a baseline to compare our ParticleSAM’s performance. Our ParticleSAM is adapted from the lightweight backbone of SAM (ViT-B), and the AMG retains the same ViT-B backbone.

**Evaluation Data:** Excluding the subset of L1 and L2 images used for adaptation, the remaining images are utilized

TABLE III  
PARTICLE SEGMENTATION COMPARISONS BETWEEN SAM AND OUR PARTICLESAM ON DIFFERENT DATASETS

| Data | Model | mIoU         | $mAP_{50}^a$ | $mAP_{60}$   | $mAP_{70}$   | $mAP_{80}$   | $mAP_{90}$   |
|------|-------|--------------|--------------|--------------|--------------|--------------|--------------|
|      |       |              | (%)          | (%)          | (%)          | (%)          | (%)          |
| L1   | SAM   | 55.80        | 53.60        | 46.53        | 43.27        | 42.18        | 41.23        |
|      | Ours  | <b>68.17</b> | <b>61.12</b> | <b>54.26</b> | <b>50.83</b> | <b>49.51</b> | <b>48.51</b> |
| L2-l | SAM   | 61.09        | 73.16        | 72.50        | 71.87        | 70.62        | 65.53        |
|      | Ours  | <b>85.53</b> | <b>89.64</b> | <b>89.32</b> | <b>88.86</b> | <b>87.98</b> | <b>85.72</b> |
| L2-h | SAM   | 27.35        | 38.82        | 33.39        | 29.59        | 25.70        | 21.22        |
|      | Ours  | <b>45.15</b> | <b>48.10</b> | <b>42.65</b> | <b>38.33</b> | <b>33.97</b> | <b>29.11</b> |
| L3-0 | SAM   | 18.26        | 34.23        | 32.48        | 31.30        | 30.33        | 27.65        |
|      | Ours  | <b>44.13</b> | <b>47.65</b> | <b>45.90</b> | <b>44.70</b> | <b>43.69</b> | <b>41.98</b> |
| L3-m | SAM   | 18.95        | 55.23        | 53.95        | 52.77        | 51.31        | 44.67        |
|      | Ours  | <b>69.25</b> | <b>82.78</b> | <b>82.03</b> | <b>81.26</b> | <b>80.27</b> | <b>76.88</b> |
| L3-h | SAM   | 8.70         | 30.06        | 28.25        | 26.80        | 25.32        | 21.87        |
|      | Ours  | <b>34.79</b> | <b>43.24</b> | <b>41.19</b> | <b>39.46</b> | <b>37.79</b> | <b>34.98</b> |

<sup>a</sup> $mAP_{50}$  represent IoU threshold at 50%

to evaluate the model’s performance. Table II provides an overview of all the test datasets. Additionally, to assess the model’s improvement in segmenting extremely small objects, we generate extra L3-0 data, which contains only layer<sub>0</sub> and layer<sub>1</sub> particles with size ranging from 4 to 11.2 mm. Moreover, to compare performance under occlusion, we partition L2 and L3 data into less occluded and heavily occluded. Specifically, due to the annotations of PSD per image, the less occluded image keeps the same PSD as the heavily occluded images, but the number of stones per class is reduced by 50%.

**Discussion:** Table III shows segmentation performance of SAM and our ParticleSAM with various metrics from different evaluation data. ParticleSAM consistently outperforms SAM across all datasets. Notably, on the occluded mixed-class L3 data, SAM fails to deliver meaningful results, whereas ParticleSAM demonstrates a remarkable improvement of 26.09% and 50.30% on mIoU. Moreover, on the L3-0 dataset, which consists exclusively of particles smaller than 22 mm, by comparing mAP under different thresholds, we can observe that our model not only improves accuracy but also exhibits more stable segmentation performance with less drop from  $mAP_{50}$  to  $mAP_{90}$ .

Although our model achieves a significant improvement in particle segmentation, its performance declines under occlusion. Comparison in L2-l and L2-h, L3-m and L3-h, heavily occlusion leads to a noticeable performance drop, from 85.53% to 45.15% and 69.25% to 34.79%, respectively. In the most heavily occluded dataset L3-h, ParticleSAM’s detection capability is limited as the number of particles increases, failing to recognize more particles (See Fig. 3). As an alternative, sub-cropping, discussed in Section III-B, can mitigate this limitation and detect more particles under heavy occlusion; however, this approach increases inference time more than 5 times, making real-time monitoring impractical.



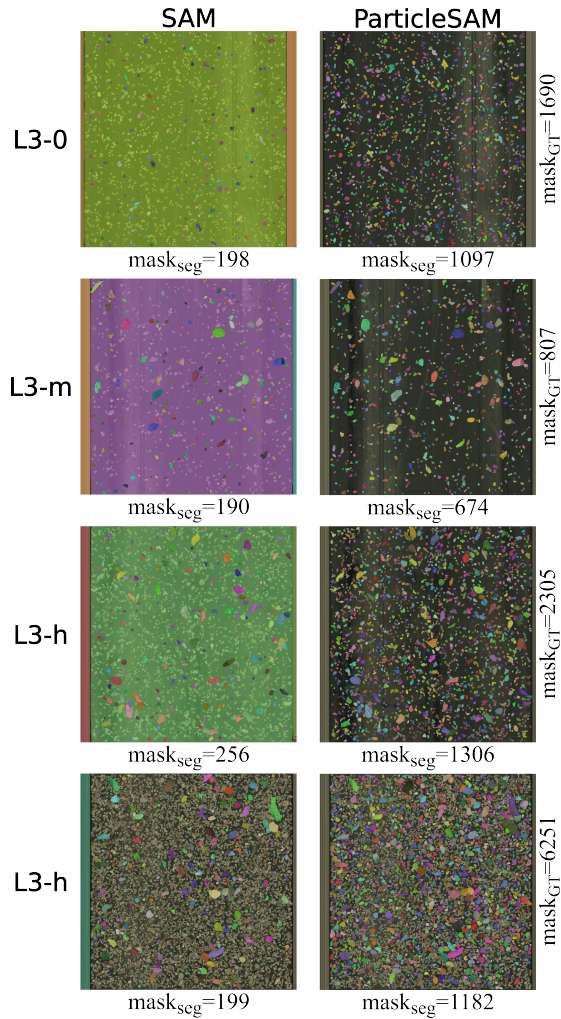


Fig. 3. Comparison of SAM and ParticleSAM segmentation across L3 dataset occlusion levels, with the number of segmented vs. total GT masks shown. The source conveyor belt background provided by the department of Anthropogenic Material Cycles (ANTS), RWTH Aachen University.

## V. CONCLUSION

We propose ParticleSAM, an adaptation of SAM for particle segmentation in high-resolution images with numerous small and dense objects. The challenge is addressed by parameter adaptation, the encoder of SAM to handle larger images and extract features on a smaller scale and the decoder to optimize the segmentation precision. Moreover, to the best of our knowledge, we are the first to introduce a large-scale simulated dataset in recycled CDW, specifically designed for crushed aggregates. This dataset provides a novel benchmark for fine-tuning and evaluating segmentation models in challenging scenarios as high-density occlusions or small particle material analysis, bridging a critical gap in existing datasets. The results on our dataset demonstrate a significant enhancement in particle detection and segmentation performance compared to SAM, especially on images with overlapping particles. Although our work was intended for the visual quality inspection of CDW, our proposed method ParticleSAM can be applied to other domains, where visual monitoring requires

the segmentation of multiple, small or overlapping objects, in which case our work can serve as a guideline. Similarly, our data generation pipeline for multi-layer, overlapping particle images can be easily adapted to any type of object, e.g. in medical and agriculture applications.

## ACKNOWLEDGMENT

We would like to thank the Department of Anthropogenic Material Cycles, RWTH Aachen University, for providing the raw images and backgrounds for the synthetic images, recorded at their facility.

## REFERENCES

- [1] Aggregates Europe, "Facts and figures," 2025, accessed: 2025-03-13. [Online]. Available: <https://www.aggregates-europe.eu/facts-figures/facts/>
- [2] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4015–4026.
- [3] J. Hernández Parrodi, N. Kroell, X. Chen, T. Dietl, E. Pfund, B. Küppers, and C. Nordmann, "Near-infrared-based material flow monitoring of construction and demolition waste nahinfrarot-basierte stromüberwachung von bau-und abbruchabfällen," *Mineralische Nebenprodukte Und Abfälle. Berlin: Thomé-Kozmiensky Verlag GmbH*, pp. 92–111, 2021.
- [4] N. Kroell, E. Thor, L. Göbbels, P. Schönfelder, and X. Chen, "Deep learning-based prediction of particle size distributions in construction and demolition waste recycling using convolutional neural networks on 3d laser triangulation data," *Construction and Building Materials*, vol. 466, p. 140214, 2025.
- [5] J. Li, H. Fang, L. Fan, J. Yang, T. Ji, and Q. Chen, "Rgb-d fusion models for construction and demolition waste detection," *Waste Management*, vol. 139, pp. 96–104, 2022.
- [6] T. Ji, J. Li, H. Fang, R. Zhang, J. Yang, and L. Fan, "Rapid dataset generation methods for stacked construction solid waste based on machine vision and deep learning," *Plos one*, vol. 19, no. 1, p. e0296666, 2024.
- [7] S. K. Dodampegama, L. Hou, E. Asadi, K. Zhang, and S. Setunge, "Adversarial domain adaptation with a modified yolov8x-seg model for construction and demolition waste segmentation," 2024, preprint available at SSRN 5106445.
- [8] S. Heo and S. Na, "Developing wastesam: A novel approach for accurate construction waste image segmentation to facilitate efficient recycling," *Waste Management & Research*, p. 0734242X241290743, 2024.
- [9] F. Kronenwett, G. Maier, N. Leiss, R. Gruna, V. Thome, and T. Längle, "Sensor-based characterization of construction and demolition waste at high occupancy densities using synthetic training data and deep learning," *Waste Management & Research*, vol. 42, no. 9, pp. 788–796, 2024.
- [10] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1290–1299.
- [11] A. M. Das, R. Chaudhry, K. Kundu, and D. Modolo, "Prompting foundational models for omni-supervised instance segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1583–1592.
- [12] J. Wu, W. Ji, Y. Liu, H. Fu, M. Xu, Y. Xu, and Y. Jin, "Medical sam adapter: Adapting segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.12620*, 2023.
- [13] A. Langley, M. Lonergan, T. Huang, and M. R. Azghadi, "Analyzing mixed construction and demolition waste in material recovery facilities: Evolution, challenges, and applications of computer vision and deep learning," *Resources, Conservation and Recycling*, vol. 217, p. 108218, 2025.
- [14] X. Wu, N. Kroell, and K. Greiff, "Deep learning-based instance segmentation on 3d laser triangulation data for inline monitoring of particle size distributions in construction and demolition waste recycling," *Resources, Conservation and Recycling*, vol. 205, p. 107541, 2024.