Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

AUTOMATED DETECTION IN UNSTRUCTURED MATERIAL STREAMS: CHALLENGES AND SOLUTIONS

Alexander Schlosser[1], Takeru Nemoto[2], Jonas Walter[1], Sebastian Amon[1], Jörg Franke[1], and Sebastian Reitelshöfer[1]

[1]Institute for Factory Automation and Production Systems,
Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, GERMANY
[2]Siemens AG, GERMANY

ABSTRACT

Post-consumer packaging waste continues to rise, intensifying the need for automated sorting to increase recycling efficiency. Lightweight packaging (LWP), with its variable geometries, materials, and occlusions, remains especially difficult for conventional vision systems. Integrating *You Only Look Once* (YOLO) instance segmentation into robotic simulation platforms enables robust real-time detection. Experiments show that synthetic datasets yield consistently high segmentation accuracy, whereas real-world performance fluctuates. Crucially, increasing model size or resolution does not guarantee improvement; task-specific tuning and system-level integration are more effective. Simulation frameworks combining Unity, Robot Operating System 2 (ROS2), and MoveIt2 provide realistic evaluation and optimization. These findings demonstrate that AI-based segmentation and digital twins can deliver scalable, adaptive, and self-optimizing sorting systems, offering a practical pathway to sustainable material recovery and circular economy implementation.

1 INTRODUCTION

Recycling plays a pivotal role in the sustainable utilization of resources and is a critical strategy for mitigating environmental impact, particularly in the face of escalating waste volumes. The optimization of recycling processes assumes increasing importance to reduce reliance on landfills and promote a functional circular economy. A salient challenge in this context pertains to the heterogeneity of material flows, which complicates efficient separation and sorting.

Despite the advancements in sorting technology, which have led to recycling rates of approximately 53% for lightweight packaging (LWP), current technologies are approaching their operational limits (MEILO 3/18/2025). To enhance sorting performance, innovative approaches are necessary, enabling more precise separation processes and distinguishing between a greater number of fractions. Ensuring material purity is of paramount importance in this context, as impurities such as metals, wood, and organic substances can significantly compromise the quality of the recycled material (Wieser et al. 2022). While existing sensor-based sorting technologies offer an expanding array of detection options, the precise identification and selective separation of LWP within heterogeneous material streams remains a technological challenge (Robben and Wotruba 2019).

A critical factor in enhancing the sorting of LWP is the reliable identification of diverse materials, including plastics, metals, and composite materials, which can present challenges due to variability in shape, size, and degree of contamination. While imaging sensors are already widely employed, there are still challenges in the robust detection of transparent or heavily soiled plastics (Robben and Wotruba 2019).

The integration of advanced sensor systems with machine learning methodologies holds promise in enhancing the precision of material detection, thereby optimizing the efficiency of separation processes. The identification of challenges associated with detecting LWP within unstructured material streams, along with potential solutions, is a subject of ongoing investigation with the aim of further enhancing the efficacy and precision of recycling processes. (Kiyokawa et al. 2024)

2 STATUS AND CHALLENGES

To support the transition to a circular economy, the European Union has launched several initiatives, including the Circular Economy Action Plan, which defines specific recycling targets for the year 2030 (Europäische Kommission 2020). A key aspect for the implementation of these strategies is the efficiency of waste sorting, as precise separation of materials is essential for sustainable recycling (Fischer et al. 2023).

In highly developed countries, modern sorting facilities are increasingly relying on automated processes to increase the precision and efficiency of sorting. The effectiveness of automated sorting processes depends largely on modern perception technologies that enable the exact localization and classification of objects in heterogeneous waste streams.

This is usually done using cameras and, in some cases, additional sensors to identify the type of material (Mao et al. 2021; Mao et al. 2022; Raptopoulos et al. 2020). In camera images, the objects to be eliminated are usually recognized by machine learning algorithms (Mao et al. 2022; Zhang et al. 2022; Zhao et al. 2020). Precise segmentation plays a decisive role in differentiating between different material fractions. In addition, ensuring the highest possible throughput is crucial from an economic perspective (Kiyokawa et al. 2024; Lubongo and Alexandridis 2022).

Advances in deep learning and computer vision have significantly improved the recognition capabilities of industrial systems. Convolutional neural networks (CNNs) have proven to be particularly powerful for the precise detection, localization and classification of objects. Although these networks are becoming increasingly complex and require more computing power, optimizations in network architecture and hardware now enable real-time applications in sorting systems (Manakitsa et al. 2024; Marculescu et al. 2018). Studies show that Deep Learning (DL)-based sorting systems have high robustness and flexibility and are able to process heterogeneous waste streams (Mao et al. 2022; Raptopoulos et al. 2020). Despite these advances, challenges remain, particularly in the development of robust instance segmentation models for the accurate detection and separation of objects.

2.1 Instance Segmentation

Instance segmentation, a central task in the field of computer vision, deals with the precise identification and delimitation of individual objects within an image. Efficient and accurate segmentation models are particularly important in application areas such as autonomous driving and robotics. Various models are used for object segmentation. These models not only detect objects but also generate corresponding masks. A mask is a pixel-level binary representation that outlines the exact shape and location of an identified object within the image.

For reliable detection and handling, the objects must currently be highly separated on the conveyor belt (Mao et al. 2021). Contamination, dust, and suspended particles have been shown to impair the functionality of optical sensors, leading to inaccuracies in object identification (Kiyokawa et al. 2024; Lubongo and Alexandridis 2022).

Recent models, including YOLO and Real-Time Model for object Detection (RTMDet), demonstrate high accuracy and real-time capability (GitHub 3/20/2025; Lyu et al. 2022). Single-stage methods, such as You Only Look At Coefficient Ts (YOLACT++), enhance processing speed through parallel detection and segmentation; however, their effectiveness is constrained when dealing with overlapping objects (Bolya et al. 2018). It has been demonstrated that dual-stage models, such as Mask Region-based Convolutional Neural Network (Mask R-CNN), can achieve high accuracies by conducting two distinct processes: detection and segmentation (Ren et al. 2017). Multi-stage models, such as Cascade Mask R-CNN, enable iterative mask refinement (Cai and Vasconcelos 2017). Segmenting Objects by Locations (SOLOv2) and YOLOv9-Seg are segmentation models based on convolutional neural networks. SOLOv2 utilizes direct pixel-to-instance mapping, while YOLOv9-Seg integrates classical object recognition with segmenting masks, yielding faster but potentially less accurate segmentation (GitHub 3/20/2025). Additionally, Vision Transformer presents an alternative architecture based on self-attention mechanisms, enabling effective capture of global image contexts (Dosovitskiy et al. 2020).

Training datasets are imperative, for instance segmentation, as neural networks are only capable of precisely recognizing individual objects, segmenting their boundaries with pixel precision, and generalizing new, unknown images through annotated examples. The creation of extensive, diverse training datasets is a resource-intensive task; consequently, synthetic datasets are frequently developed (Kiyokawa et al. 2024; Koskinopoulou et al. 2021). A common method for this is the cut-and-paste method, in which real objects are extracted from an image and inserted into a new scene. This approach enables the creation of a variety of scenarios with different degrees of overlap and object positions, as well as automatic data annotation (Dirr et al. 2024).

2.2 Real-Time Optimization and System Integration for Robotics

The use of robots for material separation, including the removal of objects from conveyor belts, has already proven effective for applications with defined framework conditions (Lubongo and Alexandridis 2022; Raptopoulos et al. 2020; Zhang et al. 2022). In addition to object detection, it is imperative to swiftly and reliably estimate the objects' pose to ascertain an optimal grasping.

The integration of Deep Learning (DL) models into robotic systems necessitates advanced inference engines such as TensorRT (Tensor Runtime) to enhance efficiency. TensorRT is a high-performance inference engine designed specifically for real-time applications, offering optimizations such as precision matching, layer and tensor fusion, and kernel autotuning. Studies demonstrate that these optimizations significantly improve both inference speed and memory management efficiency. TensorRT has been identified as a significant technology for the real-time optimization of DL models (Zhou and Yang 2017).

Additionally, middleware frameworks such as ROS play a crucial role in facilitating the flexible development of scalable robot systems. Utilizing a modularized communication model, ROS enables the exchange of processes through topics, actions, and services. The ROS2 framework has introduced substantial advancements in real-time capability, security, and scalability. A notable innovation is the integration of the Data Distribution Service (DDS), which ensures robust communication for safety-critical applications such as autonomous vehicles or aerospace. Another advancement lies in the security architecture of ROS2, which includes mechanisms such as authentication, access control, and data encryption to ensure the integrity of robotic systems in safety-critical applications. ROS2 provides a future-proof middleware platform for robotic systems with improved real-time communication, enhanced security, and modular scalability. (Macenski et al. 2022)

The integration of these two technologies facilitates the development of advanced robotics applications with superior performance, reliability, and safety, catering to the mounting demands in industrial and research contexts (Macenski et al. 2022; Zhou and Yang 2017). The objective of this project is to ascertain the applicability of these models to recycling processes and to assess their contribution to automation and enhance efficiency.

2.3 Simulation in the Context of Object Recognition

In the context of robotic perception and industrial image processing, advanced object recognition and pose estimation methods are essential for reliably handling dynamic objects with unknown or highly variable geometries. These variations may arise due to non-rigid deformations, surface contamination (e.g., dirt, dust, or abrasion), or environmental factors such as moisture, glare, or translucency, all of which can significantly impair sensor data quality and model generalization. Novel or extended algorithms based on machine learning, some of which use synthetic training and input data, are intended to achieve sufficiently robust detection performance despite different variations, deformations, or soiling of the objects to be detected. For object recognition, a suitable approach for the recognition and 6D pose estimation of the objects to be grasped must be developed and validated. The challenges posed by the variability in type, material, and appearance of objects in the LWP material flow, including unknown objects, damage, dirt, and occlusions, as well as the continuous movement of the material flow, must be addressed. The detection of relevant objects in camera images involves the integration of conventional approaches with machine learning methods. (Bircanoglu et al. 2018; Koskinopoulou et al. 2021; Tremblay et al. 2018; Vo et al. 2019; Yang et al. 2021; Zitnick and Dollár 2014)

Subsequent to the initial detection of objects in camera images, the 6D pose estimation can be performed (Blank et al. 2019). In this domain, state-of-the-art approaches can be further developed for the use case (He et al. 2021; Labbé et al. 2020; Tremblay et al. 2018). In particular, continuous learning, defined as the continuous optimization of algorithms for object recognition and 6D pose estimation, holds considerable promise in achieving error-minimized object recognition and pose estimation, thereby facilitating an error-minimized robot-based material separation process (Graffieti et al. 2022). Recent research endeavors have focused on novel methodologies for gripping planning in piles, with the objective of augmenting the operational scope of robots to encompass the grasping of objects that are partially obscured within the pile. To this end, there is a necessity to formulate a trajectory and grasp planning approach that is contingent upon the dynamics of the conveyor belt, the interaction with other objects in the pile during the grasping process, and the identification of a suitable grasping point. This identification is predicated on the detection of the object and the estimation of its pose, particularly in cases where the object is partially occluded.

3 PICTURE IDENTIFICATION OF PURIFIED REVENUE STREAMS

Advancing automation in the recycling industry is imperative to meet the mounting demands for efficiency, accuracy, and throughput. The high productivity of the sorting plant hinges on rapid and uninterrupted material separation. The ability to swiftly detect objects and reliably estimate their positions is paramount for ensuring the efficacy of sorting.

Once an object of interest has been detected and its pose estimated, efficient and fast planning of the robot pose that must be adopted to grip the object is required. The development and validation of corresponding real-time capable approaches is imperative. The main objective of this approach is the development, implementation and validation of a segmentation pipeline for the detection of impurities in heterogeneous material flows. This is specially designed for use in LWP sorting systems. The pipeline's primary function is to precisely identify, segment, and classify objects. It processes visual data in real time to identify and selectively separate foreign and hazardous materials in the complex and dynamic environment of a sorting system. However, highly varying input data, as is the case with the LWP material flow (e.g., seasonal differences, damage, and dirt), remains a challenge. The aim of this research topic is to optimize the database of trained models by continuous learning during operation and thus to continuously improve the quality of the inference of neural networks for object recognition and object estimation. This, in turn, enables error-minimized object recognition and gripping process.

Distinct synthetic data sets are developed for the purpose of efficient modeling and evaluation. These data sets simulate realistic scenarios with different focal points. Given the limited availability of real training data, the cut-and-paste method is employed to create data sets. This approach facilitates the efficient generation of diverse scenarios with varying object positions, overlaps, and different lighting conditions (Dirr et al. 2024). It also enables automatic data annotation, thereby minimizing the time-consuming manual labeling process. This methodology establishes a refined and pragmatic training foundation, paving the way for the advancement of AI models in the domain of sensor-based sorting.

Three self-generated datasets are used for the following studies. The initially used ALPHA-Dataset is a synthetic data set. It is generated under controlled laboratory conditions and comprises simple objects with the purpose of evaluating the basic functionality of the pipeline and the performance of the grippers. The ALPHA-Dataset serves as a reference for fundamental segmentation tasks. The subsequent data set, BETA, which is also synthetic, depicts more complex scenarios and contains typical impurities and objects that occur in recycling plants. The third data set, GAMMA, consists of real data generated using the company's proprietary test setup. It accurately represents actual fault objects and practical fault scenarios.

YOLOv8-seg, YOLOv9-seg, RTMDet-Ins, and YOLOv11 were initially chosen for segmentation tasks. YOLOv11 represents a substantial advancement over previous versions of YOLO. The most significant innovation of YOLOv11 lies in its extended functionality and improved adaptability. While earlier models are primarily specialized in object recognition, YOLOv11 offers a wide range of computer vision functions. Beyond object recognition, it encompasses instance segmentation, image classification, pose estimation, and oriented object recognition (OBR). A notable enhancement is the capacity for multi-label detection within a single class, enabling the concurrent identification of multiple features within an object. The architecture of YOLOv11 is enhanced with an improved backbone and neck structure, resulting in optimized feature extraction capabilities.

This enhancement results in enhanced precision in object recognition. Concurrently, the model complexity has been mitigated, signifying that YOLOv11m possesses 22% fewer parameters than YOLOv8m without compromising recognition performance. A notable advantage of YOLOv11 is its enhanced flexibility regarding hardware specifications. The model's operational versatility is evidenced by its capacity to run on both cloud platforms and on NVIDIA GPU-powered systems, as well as on resource-constrained edge devices. This feature enables the model's broad applicability, particularly in mobile and industrial scenarios. Furthermore, YOLOv11 employs optimized algorithms for image processing, facilitating enhanced stability in recognition under diverse lighting conditions. (GitHub 2025b).

The performance of these models was evaluated on data sets representing typical scenarios from sorting plants. The evaluation was performed on a Nvidia GeForce RTX 3090 GPU, analyzing both segmentation accuracy and processing speed.

4 SIMULATION OF OBJECT RECOGNITION

Simulation of object recognition is a platform that enhances the efficiency and precision of object recognition and segmentation in sorting processes. It facilitates the comparison and testing of disparate robot-assisted sorting systems, thereby enabling the identification of optimal solutions for particular material flows. Beyond the technical parameters, economic factors such as budgetary constraints and spatial requirements are also considered. The platform provides a sound decision-making basis for physical implementations by enabling early error detection and optimization, thereby preventing inefficient or costly wrong decisions in the real application.

The simulation of a robotic sorting system must fulfill a variety of requirements. These include the reach of the robot arm, the payload capacity, and the cycle time, which determines how quickly a sortation process is completed. In addition, sorting accuracy, system footprint and ease of maintenance are critical factors in minimizing downtime. Standardized interfaces allow easy integration into existing industrial processes. The platform is currently being developed to capture additional key performance indicators, as only pick rate has been evaluated to date. Other criteria for evaluating system performance are planned and will be implemented in future versions.

To efficiently implement these requirements, a modular simulation environment was developed that combines several specialized software tools (Figure 1). Unity forms the core of the physical simulation and uses the PhysX engine to realistically represent the sorting systems. High-quality visualization allows detailed analysis of industrial scenarios and simulation of large sorting systems. ROS2 serves as the central platform for control and communication between the various simulation components and the real robot systems. MoveIt2 supports robot motion planning by integrating advanced algorithms for collision avoidance and safe motion execution. ROS# connects Unity to ROS2, enabling precise control of motion sequences so that simulation results can be transferred directly to simulated systems.

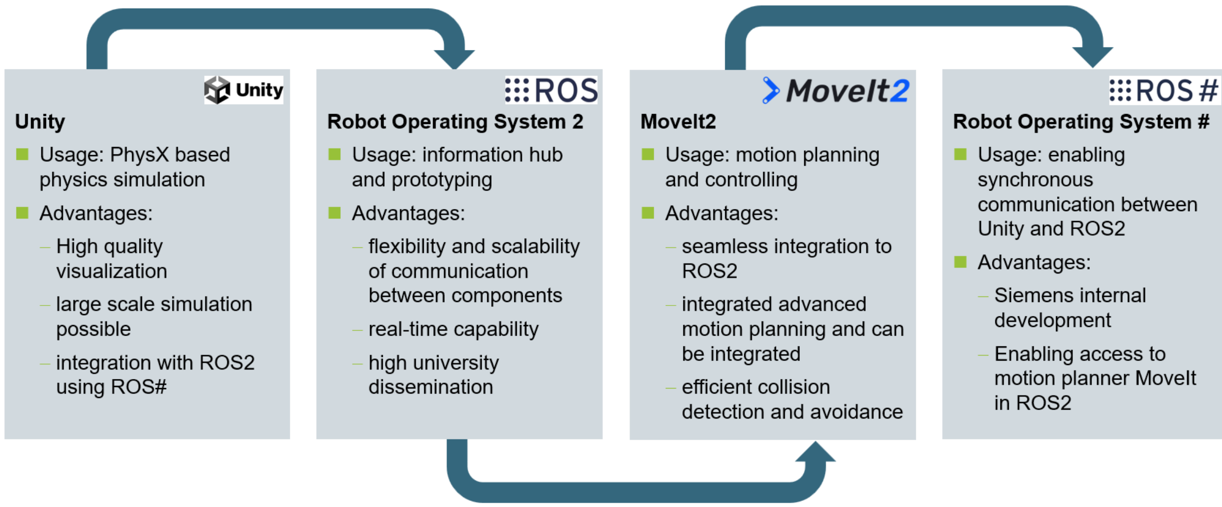

Figure 1: Modular simulation environment for robotic sorting systems. The framework integrates Unity (PhysX-based) for high-fidelity physical modeling and visualization of large-scale sorting processes. ROS2 enables centralized communication and control across simulation and real components. MoveIt2 supports motion planning with collision avoidance, while ROS# links Unity and ROS2, enabling precise transfer of simulated trajectories to physical or virtual robots.

A robotic sorting system consists of several key components that work together. The material flow ensures a continuous inflow of objects to be sorted, which are transported on a conveyor belt. An integrated image recognition system analyzes the position and material properties of the objects at a defined detection point. This data is sent in real time to the motion planner, which calculates the most efficient sequence of gripper movements to achieve the maximum number of picks per cycle. The robotic arm with the gripper then performs the sorting tasks by placing the objects into specific bins for plastic, metal, or paper based on pre-defined criteria such as material type or shape.

This platform represents a significant advancement over previous simulation systems. It enables realistic simulation and real-world testing of robotic sorting systems that can optimize efficiency and resource use in the recycling industry. Through the increased use of simulation platforms, innovative technologies and methods can be tested in the future without the need for large investments in physical prototypes. Future research should focus on the further development of simulation algorithms to further improve the generalizability of the systems. The use of reinforcement learning and adaptive AI-based motion planning methods could enable even more precise and efficient control of the sorting processes.

5 RESULTS

To test the performance of the detection environment with a live camera feed, various objects are to be detected, named, and their position determined. The detection is timestamped, and the accuracy is determined. The results of the investigations show the performance of modern real-time segmentation models in the context of robot-assisted sorting of recycling materials (Figure 2). A deep learning-based instance segmentation pipeline was created to detect and classify foreign materials in diverse lightweight packaging waste streams, supporting circular economy strategies. Examples of identifiable object classes include "Battery", "Shoe", "Cartridge", and "Phone" representing a practical categorization based on frequently observed contaminants.

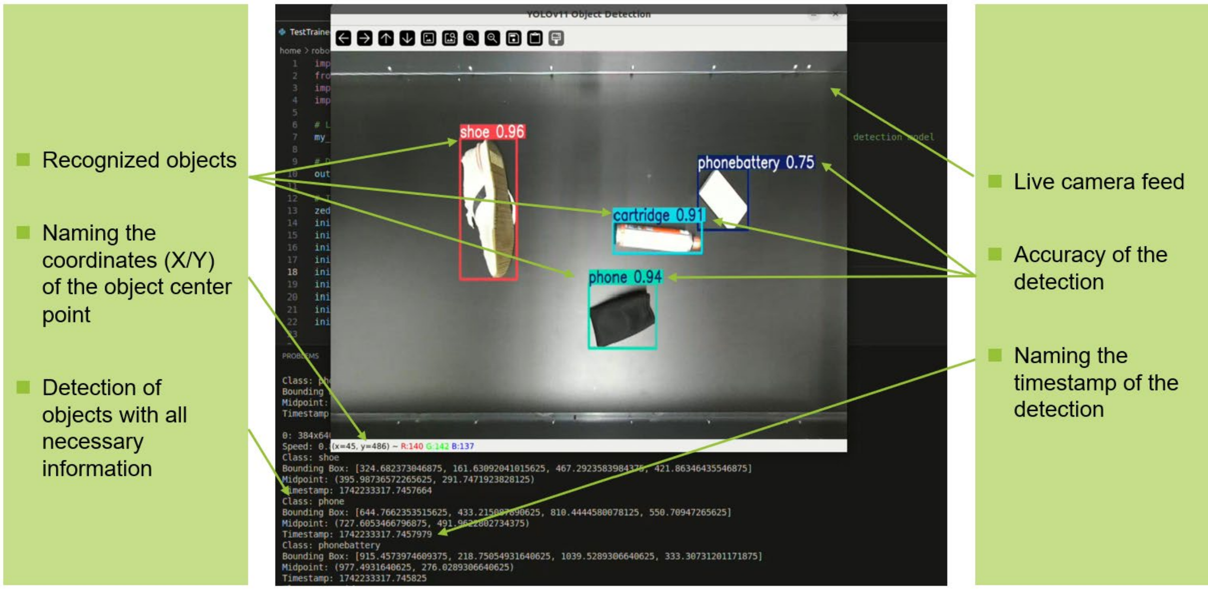

Figure 2: Evaluation of segmentation performance under realistic recycling conditions. Accuracy is assessed via live camera input on synthetic and real-world datasets of heterogeneous LWP waste. The setup replicates operational conditions, including variable backgrounds, occlusions, and low-contrast features.

By using different data sets, the models can be evaluated in a controlled environment, and their strengths and weaknesses can be analyzed in detail. The models YOLOv8-seg, YOLOv9-seg, RTMDet-Ins, and YOLOv11 were trained in different sizes and with two input resolutions of 640x640 and 1280x1280 pixels. The evaluation is done using different metrics. The primary metric is mAP50:95, which is used to evaluate the performance of the models in segmenting instances. AP50:95 (Average Precision) is also used. This evaluates the accuracy of the models with respect to individual classes. In addition to accuracy, inference time is also used for evaluation. It characterizes the suitability of the models for real-time applications.

The evaluation of different YOLO models on the ALPHA-Dataset shows a high and consistent accuracy (AP50:95) on synthetic data, with little variance between classes (Figure 3). On real data, however, there is more variation. While larger models and higher input resolutions increase accuracy on synthetic data, this effect is not consistently observed on real data. Inference times increase with model size, with individual models such as YOLOv8m-seg or RTMDet-Ins-m offering a good balance between accuracy and efficiency. Overall, the models provide more reliable results on synthetic data than on real images.

The findings derived from the BETA-Dataset demonstrate elevated AP50:95 metrics on synthetic imagery, with the notable exception of the "Battery" category, which exhibits an average of 67.36%, a substantial decline compared to the other classes that reach an average of 91.90% (Figure 4). It is noteworthy that augmenting the input size to 1280x1280 enhances performance, particularly for the "Battery" class. However, analysis of real data reveals significant fluctuations across multiple classes, including "Notebook battery", "Cell phone battery", and "Shoe". The efficacy of larger models does not consistently translate into superior outcomes. The RTMDet-Ins-m 640x640 model offers a satisfactory balance between accuracy and speed, with 57.1% mAP50:95 at a mere 3.7 ms. Overall, the real data set continues to pose a significant challenge.

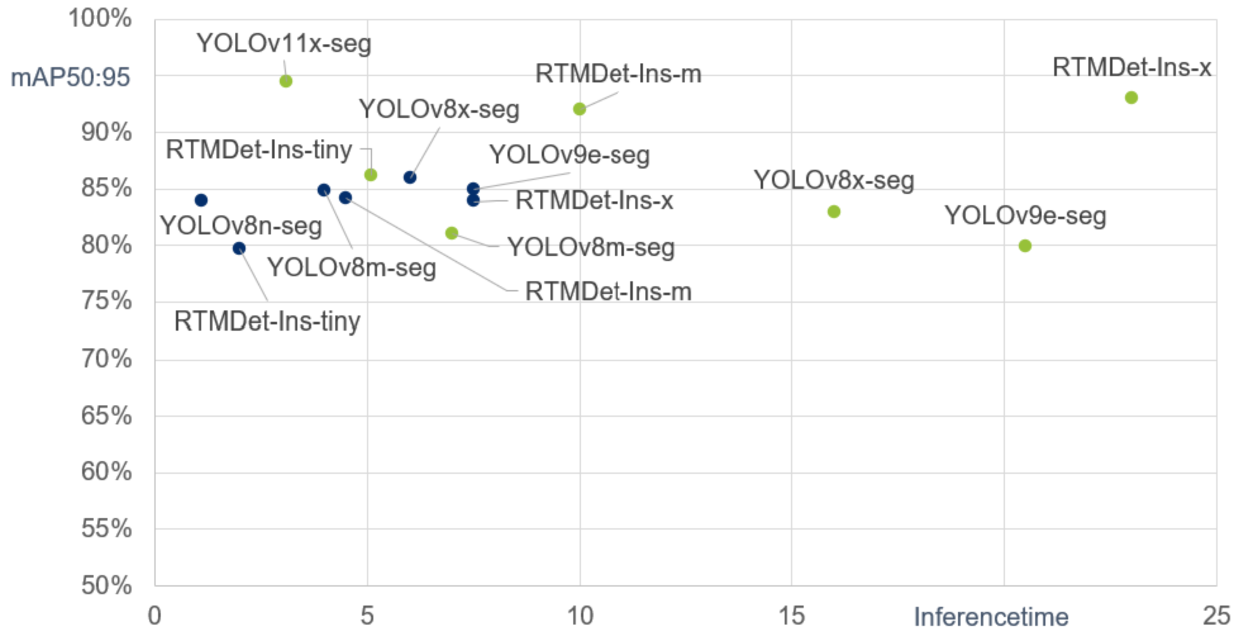

Figure 3: Comparative evaluation of YOLO-based segmentation models on the ALPHA-Dataset. On synthetic data, models achieve high accuracy (mAP50:95) with low class-wise variance, whereas real-world data shows greater variability. YOLOv11x-seg (1280 × 1280, green pentagon) exceeds most models in accuracy (>97%) and inference time (<4 ms), outperforming many 640 × 640 variants in speed. X-axis shows inference time, y-axis mAP50:95.

The experimental findings show that the RTMDet-Ins-x model achieves high segmentation accuracy on synthetic data at 1280×1280 resolution (mAP50:95 = 92.2%) but declines on real material streams (mAP50:95 = 59.9%) due to generalization challenges (Figure 4). At this resolution, the model attains a mAP50:95 of 92.2% on the synthetic BETA dataset and 59.9% on real data, with an inference time of 23 ms. The YOLOv8x-seg model, also at 1280×1280 resolution, reaches a higher 94.6% on synthetic data but only 41.0% on real data, with an inference time of 17 ms.

The GAMMA-Dataset is based on real-world images and incorporates cut-and-paste objects to replicate the complex visual conditions of recycling environments. With the utilization of this dataset, the incorporation of YOLOv11 architecture resulted in a substantial enhancement in object detection performance. Smaller objects are now detected more reliably, even under changing lighting conditions. This is supported by optimized image processing, variable input resolutions (e.g., 1080p/HDK2), and an increased confidence threshold of 80%, which together reduce misclassifications.

YOLOv11's multi-label architecture enables the simultaneous detection of multiple features across different object classes. After fewer than 50 training epochs, the model achieves a mAP50:95 of up to 90%. However, certain background objects such as shoes and cartridges remain difficult to identify. In contrast, batteries are accurately detected, unless they are visually obstructed. The model's inference speed is optimized for real-time use, making it well suited for high throughput sorting systems.

While synthetic datasets yield consistently high segmentation performance, results on real-world data remain more variable. Nevertheless, all evaluated models including YOLOv11 demonstrate real-time processing capabilities. These findings highlight the progress achieved in realistic object detection scenarios and point toward further potential for AI-assisted sorting under real-world conditions.

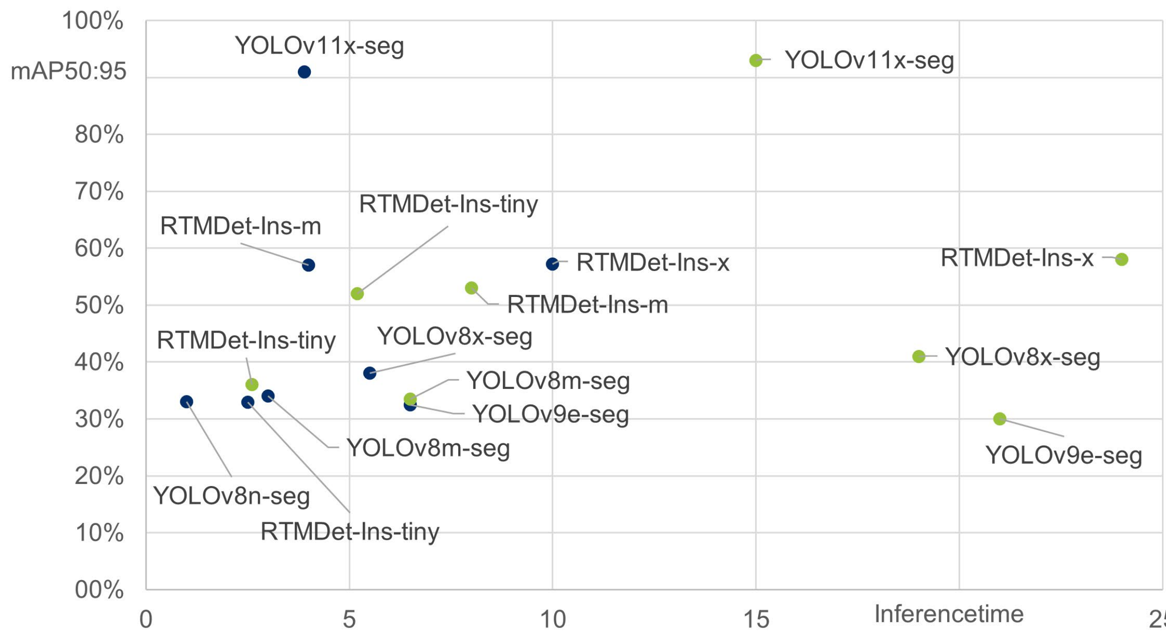

Figure 4: Segmentation performance of YOLOv11x-seg on the synthetic BETA-Dataset. YOLOv11x-seg (1280 × 1280, green pentagon) is benchmarked by mean average precision (mAP50:95) and inference time. The x-axis shows inference time (ms), the y-axis accuracy (mAP50:95, %). It achieves 90% accuracy with <4 ms inference, outperforming all other 1280 × 1280 models and most 640 × 640 models in speed.

The developed simulation platform facilitates a pragmatic evaluation of disparate sorting systems by integrating physical and algorithmic components. The integration of Unity, ROS2, and MoveIt2 ensures a realistic simulation of the gripping processes and allows for efficient optimization of motion planning. Notably, the incorporation of YOLOv11 is expected to lead to a substantial enhancement in recognition accuracy, consequently facilitating more precise object detection within the simulation environment. This enhancement is instrumental in facilitating data-driven decision-making processes, thereby enhancing the efficacy of sorting strategies in real-world implementations. In the long term, simulation offers a resource-saving option for developing new sorting concepts and further automating recycling processes.

Within the simulation, a rudimentary material flow model was initially formulated, wherein small objects are transported at a sufficient distance from each other. The simulation was configured to administer approximately 30 target objects per minute, thereby enabling the assessment of the performance of disparate robot systems. The ensuing tests yielded discernible disparities in the efficacy of the robot types examined (Figure 5).

A Delta robot was identified as the most efficient solution. Its configuration, which positions it directly above the conveyor belt, utilizes its limited working area. The lightweight design of the Delta robot facilitates high acceleration and maximum speeds, thereby maximizing the pick rate. During the simulation, the Delta robot successfully picked and sorted all objects without errors, making it a particularly efficient choice for highly dynamic sorting processes.

A Selective Compliance Assembly/Articulated Robot Arm (SCARA) robot also demonstrated commendable performance. Its positioning adjacent to the conveyor belt resulted in the influence of its semi-cylindrical working area on the objects to be sorted. This configuration constrained the robot's temporal window for material collection on the opposite side of the conveyor belt. Nevertheless, the SCARA robot was determined to be an efficient alternative, especially for scenarios with moderate sorting speeds.

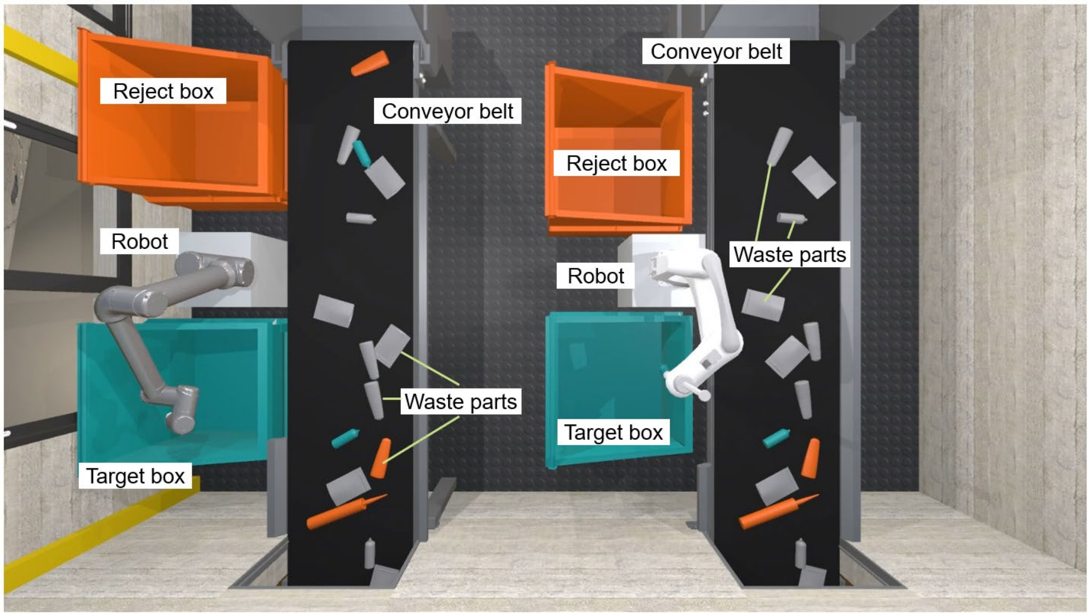

Figure 5: Simulation-based performance comparison of robotic systems in controlled material flow. A simplified material flow model was implemented in the simulation, with small objects conveyed at irregular intervals avoiding overlapped objects. Approximately 30 target objects per minute were introduced to evaluate the performance of various robotic systems. The simulation revealed marked differences in the effectiveness of the tested robot types under identical throughput conditions.

On the other hand, an articulated industrial robot, which served as the third experimental device, has a much wider workspace coverage compared to the other robot types evaluated in the study. However, its high dead weight resulted in reduced acceleration and a lower maximum speed. Consequently, the pick rate was constrained to a maximum of 15 objects per minute, which was considerably lower than the performance of the other robot systems evaluated.

6 CONCLUSION

This study develops a simulation platform for robotic sorting systems that enables efficient testing and optimization of grasping and recognition processes. The integration of Unity, ROS2, and MoveIt2 provides a realistic, modular environment that supports the evaluation of grasping strategies and informs decision making for real-world applications. Within this framework, the study demonstrates the potential of modern AI-based segmentation models that show strong performance in sorting recyclables. Although synthetic datasets have enabled significant progress, generalization to real-world material flows remains a challenge. The integration of YOLOv11 led to notable improvements in detection accuracy and robustness, mainly due to its multi-label detection capabilities and advanced image processing algorithms. Future research will extend robotic sorting by incorporating AI-based manipulation strategies, such as reinforcement learning, to actively modify material arrangements. In addition, the development of a digital twin will support detailed analysis of system behavior, particularly the synchronization between sensors and actuators. In conclusion, this study highlights the importance of advanced segmentation models and simulation environments for optimizing recycling processes and provides a solid foundation for further advances in robotic material separation. It also underscores the key role of AI-driven methods in promoting sustainable and efficient resource use.

7 OUTLOOK

The research focuses on the acquisition and detailed characterization of material flows as a basis for the simulation of robotic sorting processes. Building on this, the simulation environment is currently being extended to include a full cycle time evaluation. In contrast to conventional models that only consider pick rates, this approach also considers electrical and mechanical response times, allowing for a more accurate assessment of system performance and targeted process optimization.

Future research will focus on challenging material types, particularly large-area packaging materials such as thin films and plastic bags. These often overlap during sorting and hide items underneath - something that current technologies cannot reliably detect or handle. In practice, such tasks are performed by human operators who rely on their dexterity and adaptability. To replicate these capabilities in an automated environment, the use of two collaborative robotic arms offers a promising solution. However, existing robots are often too heavy to efficiently manipulate lightweight packages. Future developments in lightweight and agile robotic systems could therefore play a critical role in enabling fully automated handling and sorting of such complex materials.

REFERENCES

Bircanoglu, C., M. Atay, F. Beser, O. Genc, and M. A. Kizrak. 2018. "RecycleNet: Intelligent Waste Sorting Using Deep Neural Networks.":1–7.

Blank, A., M. Hiller, S. Zhang, A. Leser, M. Metzner, M. Lieret, J. Thielecke, and J. Franke. 2019. "6DoF Pose-Estimation Pipeline for Texture-less Industrial Components in Bin Picking Applications.":1–7.

Bolya, D., C. Zhou, F. Xiao, and Y. J. Lee. 2018. "YOLACT: Real-Time Instance Segmentation." *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*:9156–9165.

Cai, Z., and N. Vasconcelos. 2017. "Cascade R-CNN: Delving into High Quality Object Detection.".

Dirr, J., J. C. Bauer, D. Gebauer, and R. Daub. 2024. "Cut-paste image generation for instance segmentation for robotic picking of industrial parts." *The International Journal of Advanced Manufacturing Technology* 130:191–201.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. 2020. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.".

Europäische Kommission. 2020, Ein neuer Aktionsplan für die Kreislaufwirtschaft Für ein saubereres und wettbewerbsfähigeres Europa. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0098, accessed 20th March 2025.

Fischer, M., D. Foord, J. Frecè, K. Hillebrand, I. Kissling-Näf, R. Meili, M. Peskova, D. Risi, R. Schmidpeter, and T. Stucki. 2023, *Sustainable Business. Managing the Challenges of the 21st Century*, Springer Nature: Cham.

GitHub. 3/20/2025, ultralytics/docs/en/models. https://github.com/ultralytics/ultralytics/blob/main/docs/en/models, accessed 20th March.

Graffieti, G., G. Borghi, and D. Maltoni. 2022. "Continual Learning in Real-Life Applications." *IEEE Robotics and Automation Letters* 7:6195–6202.

He, Z., W. Feng, X. Zhao, and Y. Lv. 2021. "6D Pose Estimation of Objects: Recent Technologies and Challenges." *Applied Sciences* 11:228.

Kiyokawa, T., J. Takamatsu, and S. Koyanaka. 2024. "Challenges for Future Robotic Sorters of Mixed Industrial Waste: A Survey." *IEEE Transactions on Automation Science and Engineering* 21:1023–1040.

Koskinopoulou, M., F. Raptopoulos, G. Papadopoulos, N. Mavrakis, and M. Maniadakis. 2021. "Robotic Waste Sorting Technology: Toward a Vision-Based Categorization System for the Industrial Robotic Separation of Recyclable Waste." *IEEE Robotics & Automation Magazine* 28:50–60.

Labbé, Y., J. Carpentier, M. Aubry, and J. Sivic. 2020. "CosyPose: Consistent Multi-view Multi-object 6D Pose Estimation.":574–591.

Lubongo, C., and P. Alexandridis. 2022. "Assessment of Performance and Challenges in Use of Commercial Automated Sorting Technology for Plastic Waste." *Recycling* 7:11.

Lyu, C., W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen. 2022. "RTMDet: An Empirical Study of Designing Real-Time Object Detectors.".

Macenski, S., T. Foote, B. Gerkey, C. Lalancette, and W. Woodall. 2022. "Robot Operating System 2: Design, architecture, and uses in the wild." *Science robotics* 7:eabm6074.

Manakitsa, N., G. S. Maraslidis, L. Moysis, and G. F. Fragulis. 2024. "A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision." *Technologies* 12:15.

Mao, W.-L., W.-C. Chen, H. I. K. Fathurrahman, and Y.-H. Lin. 2022. "Deep learning networks for real-time regional domestic waste detection." *Journal of Cleaner Production* 344:131096.

Mao, W.-L., W.-C. Chen, C.-T. Wang, and Y.-H. Lin. 2021. "Recycling waste classification using optimized convolutional neural network." *Resources, Conservation and Recycling* 164:105132.
Marculescu, D., D. Stamoulis, and E. Cai. 2018. "Hardware-aware machine learning." *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*:1–8.
MEILO. 3/18/2025, MEILO Gesellschaft zur Rückgewinnung sortierter Werkstoffe mbH & Co. KG - MEILO Gesellschaft zur Rückgewinnung sortierter Werkstoffe mbH & Co. Kg. https://www.meilo-gernsheim.de/leistungen/hochwertiger-output, accessed 18th March.
Raptopoulos, F., M. Koskinopoulou, and M. Maniadakis. 2020. "Robotic Pick-and-Toss Facilitates Urban Waste Sorting.":1149–1154.
Ren, S., K. He, R. Girshick, and J. Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE transactions on pattern analysis and machine intelligence* 39:1137–1149.
Robben, C., and H. Wotruba. 2019. "Sensor-Based Ore Sorting Technology in Mining—Past, Present and Future." *Minerals* 9:523.
Tremblay, J., T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield. 2018. "Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects.".
Vo, A. H., Le Hoang Son, M. T. Vo, and T. Le. 2019. "A Novel Framework for Trash Classification Using Deep Transfer Learning." *IEEE Access* 7:178631–178639.
Wieser, M., S. H. Unterberger, and R. Lackner. 2022. "Application of Hyperspectral Imaging for identification of aging state of Styrene-Butadiene-Styrene." *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* 271:120918.
Yang, Y., J. Yang, H. Zeng, and J. Li. 2021. "A 3D Vision-Based Conveyor Tracking System for Pick-and-Sort Robotic Applications.":231–236.
Zhang, Q., Q. Yang, X. Zhang, W. Wei, Q. Bao, J. Su, and X. Liu. 2022. "A multi-label waste detection model based on transfer learning." *Resources, Conservation and Recycling* 181:106235.
Zhao, B., H. Zhang, X. Lan, H. Wang, Z. Tian, and N. Zheng. 2020. "REGNet: REgion-based Grasp Network for End-to-end Grasp Detection in Point Clouds.":13474–13480.
Zhou, Y., and K. Yang. 2017. "Exploring TensorRT to Improve Real-Time Inference for Deep Learning.":2011–2018.
Zitnick, C. L., and P. Dollár. 2014. "Edge Boxes: Locating Object Proposals from Edges." 8693:391–405.

AUTHOR BIOGRAPHIES

ALEXANDER SCHLOSSER is research associate at the Institute for Factory Automation and Production Systems (FAPS) at the Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg. His research focuses on control of tensegrity-inspired robot kinematics using self-learning algorithms; His email address is alexander.schlosser@faps.fau.de.

TAKERU NEMOTO is a mechatronics engineer in Research & Predevelopment at Siemens AG. His research focuses on advanced kinematics for efficient mechatronic systems. His email address is takeru.nemoto@siemens.com.

JONAS WALTER is the director of the research sector Medical Technology at the Institute for Factory Automation and Production Systems (FAPS) at the Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg. His research focuses on tensegrity-inspired robots and neuromechatronics. His email address is Jonas.Walter@faps.fau.de

SEBASTIAN AMON was a student at the Institute for Factory Automation and Production Systems (FAPS) at the Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg. His master thesies focused on real-time detection of contaminants in material streams using a deep learning pipeline. His email address is sebastian.amon@fau.de

JÖRG FRANKE heads the Chair of the Institute for Factory Automation and Production Systems (FAPS) at the Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg since 2009. At FAPS, he focuses his research on the manufacturing, planning and simulation of complex mechatronic systems. Previously, he held various management positions with global responsibility in industry. In addition to other honorary activities, he is a member of the International Academy for Production Engineering (CIRP) and the Scientific Society for Production Engineering (WGP). His email address is joerg.franke@faps.fau.de

SEBASTIAN REITELSHÖFER is the director of the research sector Robotics at the Institute for Factory Automation and Production Systems (FAPS) at the Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg. His main area of research is social robotics and smart mechatronics. His email address is sebastian.reitelshoefer@faps.fau.de.