From Medical Imaging to Robotic Perception: Real-Time Depalletising with Instance Segmentation

József Lénárt[1] and László Rónai[1] and
György Hegedűs[1]

[1] University of Miskolc, Miskolc-Egyetemváros, 3515, Hungary
jozsef.lenart@uni-miskolc.hu

Abstract. This paper presents a real-time robotic depalletising system that combines *YOLOv8-seg*-based instance segmentation with *RGB-D* depth sensing. Motivated by the successful transfer of segmentation paradigms from medical imaging to industrial robotics, the proposed system performs pixel-accurate object detection and three-dimensional localisation of cardboard boxes on the top layers of pallets without requiring a predefined pallet model. The software architecture, implemented in *C++17* with *OpenCV*, *ONNX Runtime*, and the *Orbbec SDK*, consists of four modular components: image acquisition, instance segmentation, 3D coordinate calculation, and TCP/IP-based robot communication. Experimental evaluation in a real industrial environment demonstrated a detection rate of 98%, a localisation accuracy of ±3 mm, and a processing throughput of 10 FPS. The results confirm that instance segmentation significantly outperforms traditional object detection methods in providing geometric information for robotic grasping, and that architectures originally developed for medical image analysis can be effectively deployed in logistics automation.

Keywords: Instance segmentation, robotic depalletising, YOLOv8-seg, RGB-D sensing, real-time 3D localisation.

1 Introduction

With the development of industrial automation, robotic material handling systems are playing an increasingly important role in logistics and manufacturing processes. During depalletising, products stacked on pallets must be automatically recognised, localised, and removed. Although the mechanical precision of modern industrial robots now enables reliable manipulation, the system's performance is still primarily limited by the sensing subsystem [1]. The depalletizing task is significantly complicated by the diverse appearance of cartons, different surface characteristics, varying illumination, and contact and partial coverage between objects. It is not enough for a robot to recognise the presence of objects; their precise separation and spatial identification are also required for reliable grasping [1, 2]. Traditional image processing methods, such as edge detection, contour extraction, or template matching, work well in structured environments, but the changing requirements of modern logistics systems increasingly exceed their capabilities [3]. The rise of deep learning has enabled semantic interpretation of

© The Author(s) 2026
S. Árvai-Homolya et al. (eds.), *Proceedings of the 2nd Innovation in Engineering Education Conference (IEE 2026)*, Atlantis Highlights in Social Sciences, Education and Humanities 58,
https://doi.org/10.2991/978-94-6239-823-8_11

scenes and pixel-accurate object delineation. The development of segmentation has been particularly significant in the field of medical imaging. *U-Net* and its improved versions [4–7] have laid new foundations for semantic segmentation, while later instance segmentation models – *Mask R-CNN*, *YOLACT* and *YOLOv8-seg* – have also enabled the separation of individual objects [8–17]. In recent years, transformer-based architectures and basic models such as *SAM* and *MedSAM* have further expanded the application possibilities of segmentation [18–20]. There are several methodological connections between medical imaging and industrial robotics. In both fields, accurate separation of relevant objects and reliable interpretation of noisy, partially obscured elements are fundamental tasks. As a result, several architectures originally developed for medical applications have also appeared in machine vision. The aim of the research was to develop an image processing system capable of automatically recognising cardboard boxes on the top layer of a pallet, separating objects that are in contact with each other, and localising them in three dimensions. To this end, a *YOLOv8-seg*-based instance segmentation model and an *Orbbec Gemini 435LE RGB-D* camera were used. This paper reviews the evolution of segmentation paradigms from medical imaging to industrial robotics applications and presents a practical depalletising system that implements these methods for real-time object recognition and three-dimensional localisation.

2 The applied methodology

2.1 Development and applicability of segmentation paradigms

During the development of modern machine vision, object detection, semantic segmentation, instance segmentation, and panoptic segmentation have become the most important paradigms for scene interpretation. Object detection describes the location of objects in terms of bounding boxes [8], while semantic segmentation assigns a class label to each pixel [4–7]. Instance segmentation combines these approaches and assigns a separate pixel-accurate mask to each object instance [9–17].

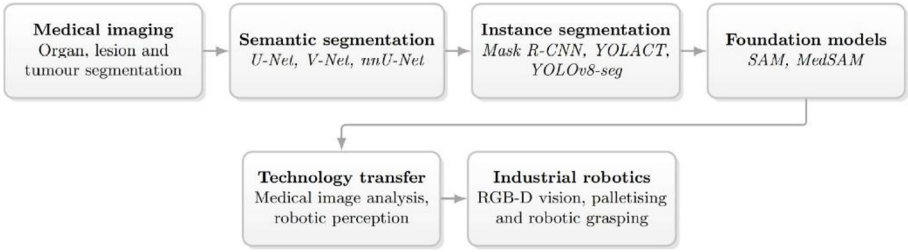


Fig. 1. Evolution of segmentation paradigms from medical imaging to industrial robotics.

The most recent development direction is represented by transformer-based architectures and basic models such as *SAM* and *MedSAM* [18–20]. In robotic depalletising tasks, instance segmentation is the most suitable approach, as it enables the separation of objects that are in contact and the determination of geometric information necessary

for grasping. The development of segmentation paradigms and their relationship to medical imaging is illustrated in Fig. 1. Considering the requirements of the depalletising task, instance segmentation proved to be the most suitable solution, as it simultaneously provides object separation and the geometric information required for robot control.

2.2 Selection of the YOLOv8-seg model

The main considerations for selecting the segmentation architecture were real-time operation, recognition accuracy, and industrial integrability. A comparison of the widely used *Mask R-CNN*, *YOLACT*, *SOLOv2*, and *YOLOv8-seg* models is shown in Table 1. *YOLOv8-seg* provides a favourable balance between accuracy and processing speed while supporting *ONNX*-based industrial deployment [17, 21].

Table 1. Comparison of representative instance segmentation models.

Model	Speed	Accuracy	Real-time applicability
Mask R-CNN	Low	Very high	Limited
YOLACT	High	Moderate	Good
SOLOv2	Moderate	High	Good
YOLOv8-seg	High	High	Excellent

Since the application under study requires recognition of a single object class, the model is well suited to the requirements of the depalletising environment.

2.3 Hardware and software architecture

The system sensor is an *Orbbec Gemini 435LE RGB-D* camera, which provides synchronised colour and depth data. The software was written in *C++17* using *OpenCV*, *ONNX Runtime* and *Orbbec SDK*. The processing chain consists of four main modules:

1. *RGB-D* image capture,
2. *YOLOv8-seg*-based instance segmentation,
3. 3D coordinate calculation,
4. robot communication via TCP/IP connection.

This modular structure not only allows easy expansion of the system and adaptation to different robot platforms, but also ensures that the neural network, coordinate calculation, and robot communication operate as independent components. Building on this foundation, the next section describes how three-dimensional localisation is achieved.

2.4 Three-dimensional localisation

The spatial position of the detected objects is determined using the masks generated by *YOLOv8-seg* and the depth image. Based on the depth values assigned to the centroids of the objects, the coordinates can be calculated using the classic depth camera model:

$$x = \frac{(u - c_x)}{z} f_x, y = \frac{(v - c_y)}{z} f_y, z = \text{depth}(\text{u}, \text{v}), \quad (1)$$

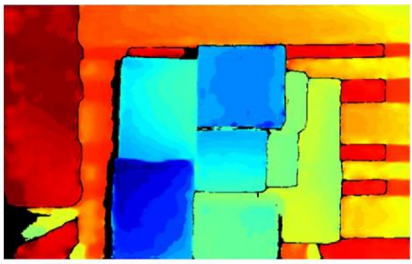
where (u, v) are the coordinates of the pixel, (c_x, c_y) are the coordinates of the principal point, and (f_x, f_y) are the focal lengths of the camera. The points defined in the camera coordinate system are transferred to the robot coordinate system using a calibrated transformation. Based on depth information, the system can automatically determine the objects on the top layer of the pallet, thereby supporting the selection of the next grasping target.

3 Results

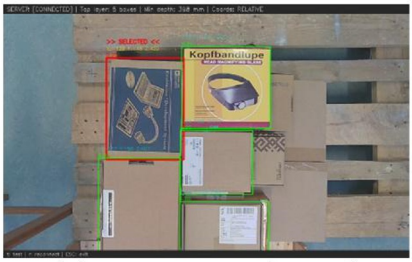
The performance of the developed depalletising system was tested in a real industrial environment using cartons of different sizes and appearances. The tests also included objects with different surfaces, labelled objects, and partially covered objects, as these pose significant challenges for robotic depalletising.



a) *RGB-D* camera image.



b) Depth map of camera image.



c) Instance segmentation result

Fig. 2. Application of instance segmentation in an industrial depalletising scenario. The *RGB-D* image **a)** and the corresponding depth map **b)** are processed by an instance segmentation model to identify individual packages and estimate their spatial locations for robotic grasping **c)**.

3.1 Instance segmentation results

The *YOLOv8-seg*-based system reliably separated cartons that were in contact with each other and delivered stable segmentation results across varying illumination and surface conditions. Pixel-accurate masks allowed for the determination of geometric features of the objects and supported 3D localisation (see Fig. 2). The results showed that instance segmentation significantly improves the quality of geometric information available to the robot compared to traditional object detection methods.

3.2 Performance metrics

To evaluate the system's performance, the detection rate, position accuracy, and processing speed were examined. The system successfully recognised 98% of the visible objects. False detections typically occurred in cases of significant occlusion or when the geometry differed from that in the training data. The accuracy of the three-dimensional localisation was ±3 mm, while the entire processing chain operated at an average speed of 10 FPS.

3.3 Three-dimensional localisation and industrial applicability

The integration of depth information enabled the determination of the objects' spatial coordinates and the automatic selection of targets located in the top layer of the pallet. The system did not require a predefined pallet model or a known number of layers, since target selection was based solely on real-time geometric measurements. The modular architecture allows for easy integration with different robot platforms. The coordinates of the detected objects can be transmitted to the robot controller via standard TCP/IP communication, while the built-in annotation environment supports the rapid adaptation of the system to new products and logistics environments. Based on the results, it can be concluded that the combination of instance segmentation and *RGB-D* sensing is suitable for supporting real-time industrial depalletising tasks while providing a favourable compromise between accuracy, processing speed, and integrability [1, 13, 17, 21, 22].

4 Discussion

The results show that segmentation methods developed in medical imaging can be successfully adapted to industrial robotics applications. In both areas, a fundamental task is to accurately separate relevant objects in noisy, partially obscured or changing environments [4–7]. The research clearly illustrates the importance of technological knowledge transfer between the two disciplines. Based on the studies, instance segmentation proved to be the most suitable approach for the depalletising task. In contrast to object detection, it provides not only the object's position but also its precise geometric description and can separate objects that are in contact [9–13]. This directly supports robotic manipulation tasks and the determination of grasping points. The use of

YOLOv8-seg represented a suitable compromise between accuracy and real-time operation. Although higher-accuracy architectures exist, in an industrial environment, processing speed and system integration are often as important as maximum recognition performance [17, 22]. The integration of *RGB-D* sensing has allowed the segmentation results to be directly linked to three-dimensional localisation and robot control. The current directions of segmentation development offer further opportunities for industrial robotics. Transformer-based architectures and basic models such as *SAM* or *MedSAM* [18–20] are expected to reduce the need for application-specific training data while improving generalisability. In addition, an increasingly important area of research is the integration of segmentation and robotic decision-making, in which segmentation masks can be used not only for object recognition but also to support manipulation planning. The presented system is currently designed to handle a single object class; therefore, an important direction for future research is investigating multi-class depalletising tasks and comparing different segmentation architectures in an industrial environment. Overall, it can be concluded that segmentation technologies derived from medical imaging are playing an increasingly decisive role in the development of intelligent robotic systems.

5 Conclusions

The study reviewed the development of segmentation methods from medical imaging to industrial robotics applications and presented a depalletising system based on *YOLOv8-seg* and *RGB-D* sensing. The results demonstrated that the segmentation paradigms developed in medical imaging can be effectively applied in industrial environments. Based on the studies, instance segmentation proved to be the most suitable approach for supporting robotic depalletising tasks, as it can separate objects that are in contact with each other and provide the geometric information necessary for manipulation. The presented system achieved a detection rate of 98%, positional accuracy of ±3 mm, and a processing speed of 10 FPS, providing a suitable basis for real-time industrial applications. One of the most important results of the research is the demonstration that the development of medical imaging and industrial robotics is closely linked. From *U-Net*-based models to instance segmentation architectures to the latest basic models, a technological development chain is emerging that is of decisive importance for both areas. In the future, the role of transformer-based architectures and universal models such as *SAM* or *MedSAM* is expected to continue to grow, thereby increasing the generalisability and applicability of robotic vision systems. Overall, segmentation plays a key role in the development of next-generation intelligent robotic systems.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. D. Buongiorno, D. Caramia, L. Di Ruscio, N. Longo, S. Panicucci, G. Di Stefano, V. Bevilacqua, A. Brunetti, Object detection for industrial applications: Training strategies for ai-based depalletizer, Applied Sciences 12 (2022) 11581. doi:10.3390/app122211581.
2. World Congress on Electrical Engineering and Computer Systems and Science, Logistics Box Recognition in Robotic De-Palletizing System with Combination of Cycle-GAN and Mask-RCNN, Avestia Publishing, 2023. doi:10.11159/mvml23.113.
3. D. Katsoulas, L. Bergen, L. Tassakos, A versatile depalletizer of boxes based on range imagery, in: Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292), Vol. 4, 2002, pp. 4313–4319. doi:10.1109/ROBOT.2002.1014438.
4. R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, D. Merhof, Medical image segmentation review: The success of u-net, IEEE Transactions on Pattern Analysis and Machine Intelligence (2024) 1–20. doi:10.1109/tpami.2024.3435571.
5. N. Siddique, S. Paheding, C. P. Elkin, V. Devabhaktuni, U-net and its variants for medical image segmentation: A review of theory and applications, IEEE Access 9 (2021) 82031–82057. doi:10.1109/access.2021.3086020.
6. 2016 Fourth International Conference on 3D Vision (3DV), V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. doi:10.1109/3dv.2016.79.
7. 7. F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, K. H. Maier-Hein, nnU-net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2020) 203–211. doi:10.1038/s41592-020-01008-z.
8. I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, P. Ilono, Deep convolutional neural networks in medical image analysis: A review, Information 16 (2025) 195. doi:10.3390/info16030195.
9. M. Süleyman Yıldırım, E. Dandıl, Automatic detection of multiple sclerosis lesions using Mask R-CNN on magnetic resonance scans, IET Image Processing 14 (12 2020). doi:10.1049/iet-ipr.2020.1128.
10. 2024 Asian Conference on Intelligent Technologies (ACOIT), Detection and segmentation of oral lesion using Mask R-CNN, IEEE, 2024. doi:10.1109/acoit62457.2024.10939183
11. Y. Zhang, S. Chan, V. Y. Park, K.-T. Chang, S. Mehta, E. K. Kim, F. J. Combs, P. R. Chang, D., R. Parajuli, R. S. Mehta, C.-Y. Lin, S.-H. Chien, J.-H. Chen, M.-Y. Su, Automatic detection and segmentation of breast cancer on MRI using Mask R-CNN trained on non–fat-sat images and tested on fat-sat images, Academic Radiolog 29 (2022) S135–S144. doi:10.1016/j.acra.2020.12.001.
12. J. Yoon, J. Han, T. P. Nguyen, Logistics box recognition in robotic industrial depalletising procedure with systematic rgb-d image processing supported by multiple deep learning methods, Engineering applications of artificial intelligence 123 (2023) 106311–106311. doi:10.1016/j.engappai.2023.106311.
13. S. Kim, V.-D. Truong, K.-H. Lee, J. Yoon, Revolutionizing robotic depalletizing: Ai-enhanced parcel detecting with adaptive 3D machine vision and RGB-D imaging for automated unloading, Sensors 24 (2024) 1473. doi:10.3390/s24051473.
14. Lecture Notes in Computer Science, Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation, Vol. 13803. doi:10.1007/978-3-031-25066-8_9.
15. IEEE Xplore, UNETR: Transformers for 3D Medical Image Segmentation. doi:10.1109/WACV51458.2022.00181.

16. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. doi:10.1007/978-3-031-08999-2_22.
17. 2025 4th International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology (AIoTC), YOLOv8n-Seg-ELA: Real-Time Instance Segmentation of Industrial Packages With Efficient Local Attention, IEEE, 2025. doi:10.1109/aiotc66747.2025.11198641.
18. X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Medical Image Analysis 58 (2019) 101552. doi:10.1016/j.media.2019. 101552.
19. J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, Nature Communications 15 (1) (2024) 654. doi:10.1038/s41467-024-44824-z.
20. M. A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, Y. Zhang, Segment anything model for medical image analysis: An experimental study, Medical Image Analysis 89 (2023) 102918. doi:10.1016/j.media.2023.102918.
21. 2023 27th International Conference on Methods and Models in Automation and Robotics (MMAR), Depalletisation humanoid torso: Real-time cardboard package detection based on deep learning and pose estimation algorithm, IEEE, 2023. doi:10.1109/mmar58394.2023.10242581.
22. R. Sapkota, D. Ahmed, M. Karkee, Comparing yolov8 and mask r-cnn for instance segmentation in complex orchard environments, Artificial intelligence in agriculture 13 (07 2024). doi:10.1016/j.aiia.2024.07.001.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

