Using Video Segmentation for Unsupervised Industrial Opitcal Production Control – Case Study for the SmartFactory@OST

Valmir Bekiri, Stefan Stöckler

OST – Eastern Switzerland University of Applied Sciences Institute for Information and Process Management, Rosenbergstrasse, St. Gallen, Switzerland
valmir.bekiri@ost.ch, stefan.stoeckler@ost.ch

This paper explores the application of video segmentation techniques for unsupervised industrial optical production control. It discusses the limitations of traditional manual inspection methods and the growing need for automated, efficient, and accurate quality control in modern manufacturing. The study focuses on unsupervised methods, particularly Meta's Segment Anything Model 2 (SAM-2), for their adaptability to new product types without extensive labeled training data. One use case from the SmartFactory@OST is presented in detail: video segmentation for monitoring a cobot handling floorballs. The implementation demonstrates the potential of these techniques to enhance efficiency, reduce costs, and improve quality assurance in industrial settings. The paper concludes by highlighting the advantages of unsupervised segmentation methods and suggesting areas for future research and improvement.

DOI
https://doi.org/10.18690/um.fov.4.2025.8

ISBN
978-961-286-998-4

Keywords:
computer vision, smart factory, video segmentation, industry 4.0, AI

University of Maribor Press

1 Introduction

Quality control is a critical process in industrial manufacturing to ensure that products meet required specifications and standards. Traditionally, quality inspection has relied heavily on manual visual inspection by human operators, which can be time-consuming, labor-intensive, and prone to errors due to fatigue or inconsistency. With the increasing demands for efficiency, accuracy, and automation in modern manufacturing, there is a growing need for more advanced and reliable quality control methods. (Raisul Islam et al., 2024; Vergara-Villegas et al., 2014; Villalba-Diez et al., 2019)

Newly developed quality assurance methods often rely on supervised machine learning techniques, which require extensive labeled datasets and are time-consuming to implement. This has led to a growing interest in unsupervised methods, particularly image and video segmentation, for industrial production quality control. (Ettalibi et al., 2024; Villalba-Diez et al., 2019)

These technologies enable machines to divide visual data into distinct segments, which correspond to different parts or features of an object, allowing for precise identification of defects and irregularities. (Ettalibi et al., 2024; Villalba-Diez et al., 2019)

This paper explores the application of video segmentation techniques for unsupervised industrial production control, examining recent advances, key challenges, and promising future directions in this important area.

1.1 SmartFactory@OST as a production facility

The SmartFactory@OST realizes a manufacturing company that exemplifies a comprehensive learning environment specifically designed to bridge theoretical education with practical implementation of Industry 4.0 concepts across multiple disciplines. Within this innovative framework, the floorball manufacturing process serves as a particularly effective pedagogical model for demonstrating advanced manufacturing concepts and digital integration techniques, see Figure 1. (Stefan Stöckler et al., 2021)

Figure 1: Floorball production cell
Source: https://www.ost.ch/smartfactory

The floorball production system represents a sophisticated implementation of configurable product manufacturing that addresses multiple aspects of modern industrial processes. Each ball consists of two independently colored hemispheres, with nine distinct color options available per hemisphere, yielding 45 possible color combinations. This configuration presents an ideal use case for teaching critical concepts such as product variant management, manufacturing execution system (MES) integration, and real-time production planning. The process flow incorporates a collaborative robot (cobot) that handles the ball halves and transfers them to a specialized welding machine, where the hemispheres are melted together under precisely controlled conditions, see Figure 1. (Stefan Stöckler et al., 2021)

The system architecture incorporates multiple data acquisition points, real-time monitoring capabilities, and comprehensive process integration with the central SAP S/4HANA ERP system. Students engage with complex implementation challenges including material master data configuration, production and sales bill of materials differentiation, variant configuration methodology, and integration of order management with manufacturing execution. This interconnected approach facilitates a comprehensive understanding of vertical and horizontal system

integration within manufacturing environments, allowing students to gain hands-on experience with practical industry applications of digital transformation concepts. The floorball production cell thus serves as a microcosm of larger industrial environments, enabling students to develop transferable skills applicable to full-scale manufacturing operations while remaining accessible for educational purposes. (Stefan Stöckler et al., 2021)

Although it is a laboratory environment, the production cell including the cobot of the SmartFactory@OST offers ideal conditions as a case study for optical production control through computer vision.

1.2 Image and Video Segmentation

Computer vision and image processing techniques have emerged as powerful tools for automated visual inspection in industrial settings. In particular image and video segmentation approaches offer significant potential for unsupervised quality control by automatically partitioning images or video frames into meaningful segments that can be analyzed to detect defects or anomalies. Unlike supervised methods that require extensive labeled training data, unsupervised segmentation techniques can adapt to new product types and defect patterns with minimal manual configuration like reference points. (Raisul Islam et al., 2024; Vergara-Villegas et al., 2014)

Image segmentation involves dividing an image into multiple segments or objects, typically to identify boundaries and objects within images. Common segmentation approaches include thresholding, edge detection, region-based methods, and clustering algorithms. Recent advances in artificial intelligence have also enabled more sophisticated segmentation models that can learn complex visual patterns. (Villalba-Diez et al., 2019; Zhou et al., 2021)

In addition, video-based segmentation, which captures temporal information, is becoming increasingly important in real-time quality control applications. Video segmentation allows for continuous monitoring of production processes, enabling the detection of transient defects or handling errors that may not be visible in static images. (Ettalibi et al., 2024; Ravi et al., 2024)

Unsupervised quality control by image segmentation is already well researched in the health care sector see B. Audelan and H. Delingette (Audelan & Delingette, 2019). Villalba-Diez et. al. have also shown the possibilities of computer vision in the printing industry for optical quality control see (Villalba-Diez et al., 2019). Furthermore M. R. Islam et al. is showing advanced methods for quality control in different industries like agriculture, electronics, automotive and more. (Raisul Islam et al., 2024)

Proving that machine learning and computer vision has huge opportunity in quality control.

1.3 Meta's Segment Anything Model 2 (SAM-2)

SAM-2 is an advanced AI model for image and video segmentation, released by Meta AI in 2024. It improves upon the original SAM with enhanced capabilities. It can segment any object in an image without prior training on specific categories, so it works effectively on new, unseen objects, which is crucial for the industrial quality control.

The model accepts various input methods like points and boxes to guide the segmentation and therefore to produce more precise object masks. And most importantly it can technically operate in real-time on consumer hardware. (Ravi et al., 2024)

The Segment Anything Video dataset, which will be used in the prototype implementation, was trained on 50.9K videos. 35.5M masks are trained in the dataset which is 53 times more masks than any other existing video segmentation dataset available in July 2024. One of the most interesting features of SAM-2 is that segmented objects in a video can be occluded by other objects and reappear again and still be tracked by the model. (Ravi et al., 2024)

This makes it an ideal solution for our cobot case study where the robot arm could cover on some occasions the object we are looking for.

2 Design and Implementation

The SmartFactory@OST delivers the perfect base for an proof-of-concept implementation on unsupervised quality control methodologies, with the principal objective of demonstrating the feasibility and efficacy of routine video surveillance integration within industrial manufacturing environments. A specific production case involving floorball manufacturing is the experimental testbed for this prototypical implementation, offering both practical complexity and operational relevance. The error case were manually triggered in the labortory environment to test if the proof-of-concept is working properly.

The technical architecture of this implementation framework employs Python as the primary programming language, using Meta's Segment Anything Model 2 (SAM-2) as the core computer vision foundation. This state-of-the-art model facilitates advanced segmentation capabilities without requiring extensive labeled training datasets, a significant advantage in manufacturing environments where product variations are common and comprehensive data labeling would be prohibitively resource intensive. Only two reference points were necessary for the system to detect the floorball and its two parts correctly. The implementation additionally incorporates OpenCV libraries for fundamental image processing operations, visualization rendering, and supplementary computer vision functionalities.

From a hardware perspective, the implementation deliberately utilizes cost-effective equipment configurations, demonstrating that advanced computer vision applications can be deployed without substantial infrastructure investments. A standard high-definition webcam provides sufficient video capture resolution and frame rate capabilities to support the analytical requirements of the system. This hardware minimalism represents an important consideration for practical industrial adoption, as it substantially reduces implementation barriers and enhances return-on-investment projections for manufacturing facilities considering similar quality control automation.

2.1 Production control with a cobot using video segmentation

In this use case the cobot is grabbing two ball halves and putting them into the welding machine. The welding machine fuses them together and after cooling down the cobot is grabbing the finished ball from the welding machine and transporting it to the other end of the production cell to release it into the dispensing tube. This grabbing and transporting process is an unsafe operation for the cobot. The floorball may slip through the gripper see Figure 1, the ball could be malformed due to issues in the welding process and many different issues could occur during the production. That's exactly where the video segmentation with SAM-2 comes in handy to check if the cobot's handling is failure free. The route that the robot is taking from the welding machine to the dispensing tube could vary based on which dispensing tube is set for the respective order. There are three ways on how the ball can be dispensed. Therefore there are also three different paths the cobot could potentially take. And this is where the charm of our approach comes in: The system does not need to be adapted since it can track basically all current and potential new paths automatically as long as it is somewhere in the viewport visible to the camera.

SAM-2 can technically not process a video file or stream as is. It needs to be converted to images based on the frames of the video. This gives us the additional opportunity to make the system more efficient since we do not need to analyze all frames per second. In the current solution every 15th frame is taken from the video. This setting is tested specifically for this application and needs to be configured and tested separately for every use case, based on how fast and far away the objects are moving in between frames and is also dependent on how many frames the used camera is delivering.

Figure 2: Floorball with different colours
Source: Created by the authors

In the next step an input information needs to be passed to SAM-2 to define which part of the video or respectively the frame is going to be segmented and therefore tracked. Since the floorball's can have a different colour (see Figure 2) per ball halve, two points are passed referencing each ball halve.

In the video segmentation workflow exemplified by the SAM-2 implementation for floorball detection, a sophisticated processing pipeline is initiated upon point-based reference specification. When the two reference points, each corresponding to a distinct ball half, are supplied to the model, SAM-2 executes a complex series of computational operations. The foundation of this process involves the model's attention mechanism analyzing pixel-level relationships surrounding the specified points, thereby establishing a probabilistic mask boundary delineation. (Ravi et al., 2024)

This segmentation approach is particularly significant since it demonstrates robust object tracking persistence even during partial occlusion by the cobot arm, a critical capability for continuous quality monitoring and therefore it enables completely unsupervised quality control, eliminating the resource-intensive requirement for extensive labeled training datasets.

Figure 3: Tracking the floorball parts through the production and visualization through masks

Source: Created by the authors

The resulting segmentation mask, as visualized in Figure 3, provides precise boundary detection of the floorball. With this technique the floorball can be basically tracked within the whole production cell.

Figure 4: Mask of the detected floorball for additional verification and quality control
Source: Created by the authors

The created mask allows us to perform subsequent comparative analysis against sample images to verify production integrity and cobot handling efficacy throughout the manufacturing process. Using a sample image of a floorball (Figure 2) we can therefore additionally quality control the segmented video and check if the detected part is really a floorball, see Figure 4. By passing this check we can be sure that the floorball is produced correctly, and the production line is working as intended.

As the floorball has a simple shape of an circle it is pretty simple to calculate if the created mask represents the mask of the sample image.

By performing quality assurance during the natural movement of products, you can eliminate the need for dedicated inspection time or additional handling. Quality control becomes seamlessly integrated into the existing production flow, with no separate stations or extra steps required. While the current implementation focuses on basic presence and shape detection for floorballs, this same methodology could extend to more complex products where multiple quality parameters could be verified simultaneously during handling or process steps.

An potential improvement for the current implementation would be to calculate the perimeter, area and aspect ratio of the mask and check if those fit to the expected values. This would further improve the quality assurce, since we could technically also know exactly in what distance the floorball should be in each production step and therefore we could not only detect if its an floorball but also if the size of it is genuine and eliminate potential false positives.

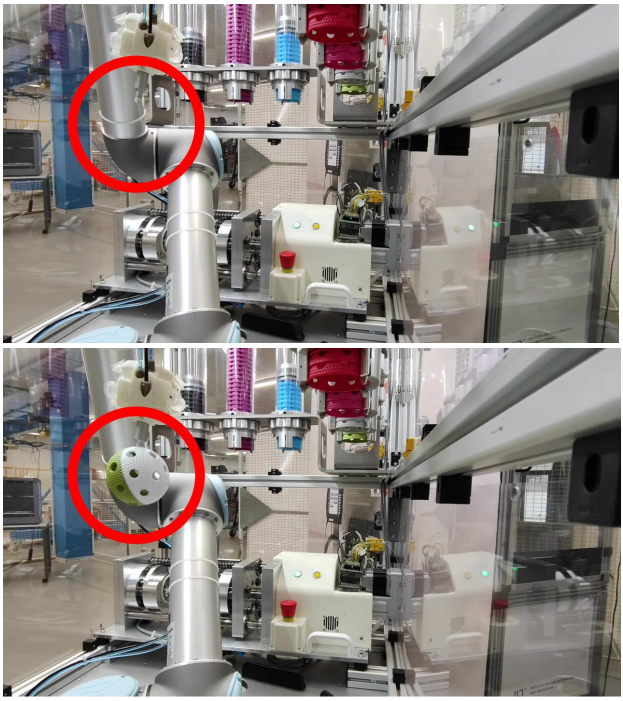

Figure 5: The difference between a wrong and a correct production line
Source: Created by the authors

In Figure 5 two images are shown which describe the difference of an error prune production line and the correct production line. As visible in the above picture there is no floorball in the gripper component. That might happen after the welding process is done and the gripper tries to fetch the now fully cooled down floorball. This state represents a significant production anomaly that traditionally required human supervision for detection or an aditional sensor for the specific use case.

Using the implemented proof-of-concept we can now immediatly detect that the ball is not in the expected position of the gripper. The implementation therefore demonstrates the efficacy of unsupervised segmentation techniques for industrial quality control applications without requiring extensive labeled training datasets, thereby significantly reducing implementation overhead while maintaining high detection reliability.

3 Conclusion and Outlook

Implementing video segmentation for unsupervised industrial optical process control offers several advantages. Image and video segmentation can adapt to different products and defects without retraining on new labeled data and only using minimal manual input. The automation reduces the time and labor costs associated with manual inspections. Machine-based inspections provide consistent results, minimizing human error without fatigue for 24/7 production lines. Last but not least it allows for immediate detection and correction of defects, reducing waste and downtime. (Ettalibi et al., 2024)

As industries strive for higher efficiency and lower operational costs, the adoption of unsupervised image and video segmentation for quality control is not just advantageous but necessary. It represents a significant step toward fully automated and intelligent manufacturing systems capable of self-monitoring and adaptation. (Ettalibi et al., 2024)

The implementation demonstrates the efficacy of Meta's Segment Anything Model 2 (SAM-2) for unsupervised industrial optical quality control in manufacturing environments. The case study validates several key advantages of this approach within the SmartFactory@OST production facility.

A particularly notable characteristic of the implementation is its path-independence during object tracking. The system successfully maintains continuous object identification regardless of the cobot's movement during transport operations. This intrinsic adaptability eliminates the need for explicit path-specific training, enabling flexible production line reconfiguration without system recalibration.

The model exhibits robust tracking persistence during object manipulation. Product rotation and occlusion by the cobot's gripper or arm presents no significant challenges to the tracking algorithm. Once initial object recognition is established, the system maintains comprehensive product tracking as a cohesive entity despite orientation changes or temporary visual obstruction.

In conclusion, Meta's Segment Anything Model offers a powerful solution for image and video segmentation tasks in industrial quality control. Its ability to operate without extensive supervised training data makes it particularly well-suited for unsupervised applications, where it can enhance efficiency, reduce costs, and improve the overall quality assurance process.

Despite demonstrable efficacy, several technical constraints remain in current implementations. The most significant limitation of the SAM-2 model is the prerequisite for manual specification of initial reference parameters, either points, bounding boxes, or preliminary masks, to effectively segment targeted objects within visual data. This necessitates a priori domain knowledge transfer to the system during configuration phases. A fully autonomous implementation would require supplementary preprocessing algorithms capable of automatic object detection prior to segmentation. So, this knowledge has to be passed first to the system before it can work properly. It would be a big improvement if a processing step could be put in which could autodetect the object.

Therefore, several promising research trajectories emerge from this implementation. The development of automated object detection preprocessing modules would constitute a significant advancement, enabling fully autonomous segmentation without manual reference specification. Additionally, the integration of feedback-driven corrective mechanisms represents an important evolution toward closed-loop quality control systems. For instance, when the system detects a floorball dislocation from the gripper, it could theoretically initiate automated recovery procedures by tracking the object's position and implementing appropriate remediation protocols.

The integration of multimodal imaging technologies, particularly hyperspectral imaging as proposed by De Ketelaere et al. (2022), offers substantial potential for enhancing detection capabilities beyond visible spectrum limitations. Hyperspectral approaches enable material composition analysis and defect identification that

remain imperceptible under standard RGB imaging conditions. (De Ketelaere et al., 2022)

A noteworthy advantage of the proposed methodology is the elimination of specialized sensor arrays traditionally employed in quality control systems. The camera-centric approach enables comprehensive process monitoring which significantly reduces the system complexity and maintenance requirements. This architectural simplification represents an important advancement in industrial monitoring system design, providing both cost efficiency and integration flexibility for existing production environments.

It is also important to mention that the developed use case was build upon an controlled environment within the SmartFactory@OST. The current solution therefore represents an demonstration of an potential professional implementation. Many different egde case would need further tests to gain statistics of the success-rate.

In conclusion, the integration of unsupervised segmentation methodologies, particularly those leveraging foundation models like SAM-2, presents a transformative approach to industrial quality control that aligns with Industry 4.0 paradigms of intelligent, adaptive manufacturing systems. Future work will focus on addressing the identified limitations while expanding the technological capabilities toward increased autonomy.

References

Audelan, B., & Delingette, H. (2019). *Unsupervised Quality Control of Image Segmentation Based on Bayesian Learning* (S. 21–29). https://doi.org/10.1007/978-3-030-32245-8_3

De Ketelaere, B., Wouters, N., Kalfas, I., Van Belleghem, R., & Saeys, W. (2022). A fresh look at computer vision for industrial quality control. *Quality Engineering*, *34*(1), 152–158. https://doi.org/10.1080/08982112.2021.2001828

Ettalibi, A., Elouadi, A., & Mansour, A. (2024). AI and Computer Vision-based Real-time Quality Control: A Review of Industrial Applications. *Procedia Computer Science*, *231*, 212–220. https://doi.org/10.1016/j.procs.2023.12.195

Raisul Islam, M., Zakir Hossain Zamil, M., Eshmam Rayed, M., Mohsin Kabir, M., Mridha, M. F., Nishimura, S., & Shin, J. (2024). Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. *IEEE Access*, *12*, 121449–121479. https://doi.org/10.1109/ACCESS.2024.3453664

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár,

P., & Feichtenhofer, C. (2024). *SAM 2: Segment Anything in Images and Videos* (No. arXiv:2408.00714). arXiv. https://doi.org/10.48550/arXiv.2408.00714

Stefan Stöckler, Roman Hänggi, Raphael Bernhardsgrütter, & Christoph Baumgarten. (2021, September 13). (PDF) SAP Academic Community Conference 2021 DACH Industrie 4.0 begreifbar machen -Die SmartFactory@OST. *SAP Academic Community Conference 2021 DACH*. SAP Academic Community Conference 2021 DACH, München. https://doi.org/10.14459/2021md1622154

Vergara-Villegas, O. O., Cruz-Sánchez, V. G., de Jesús Ochoa-Domínguez, H., de Jesús Nandayapa-Alfaro, M., & Flores-Abad, Á. (2014). Automatic Product Quality Inspection Using Computer Vision Systems. In J. L. García-Alcaraz, A. A. Maldonado-Macías, & G. Cortes-Robles (Hrsg.), *Lean Manufacturing in the Developing World* (S. 135–156). Springer International Publishing. https://doi.org/10.1007/978-3-319-04951-9_7

Villalba-Diez, J., Schmidt, D., Gevers, R., Ordieres-Meré, J., Buchwitz, M., & Wellbrock, W. (2019). Deep Learning for Industrial Computer Vision Quality Control in the Printing Industry 4.0. *Sensors (Basel, Switzerland)*, *19*(18). https://doi.org/10.3390/s19183987

Zhou, L., Zhang, L., & Konz, N. (2021). *Computer Vision Techniques in Manufacturing*. https://doi.org/10.36227/techrxiv.17125652.v1