Object Detection for Human–Robot Interaction and Worker Assistance Systems

Hooman Tavakoli, Sungho Suh, Snehal Walunj, Parsha Pahlevannejad, Christiane Plociennik, and Martin Ruskowski

1 Introduction: Why Object Detection in the Industrial Environment is Helpful?

The integration of Object Detection (OD) technology into industrial environments offers significant changes and improvements by addressing critical challenges and optimizing operations. By accurately detecting and recognizing objects in real time, OD systems play an important role in ensuring safety, streamlining workflows, and efficient assistance for humans. In today's complex industrial landscape, accurate OD is pivotal as it serves as the foundation for various safety mechanisms, such as identifying obstacles and hazardous materials, thereby reducing accidents and downtime.

Efficiency, productivity, and worker safety are paramount in industrial settings, making accurate object detection an essential component of worker assistance systems. Leveraging advanced algorithms, sensor fusion techniques, and machine learning methodologies, industries can achieve improved automation, enhanced human–robot interaction, and optimized processes. OD technology enables indus-

H. Tavakoli (✉) · S. Walunj · P. Pahlevannejad · C. Plociennik · M. Ruskowski
German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

Technologie-Initiative SmartFactory, Kaiserslautern, Germany
e-mail: hooman.tavakoli_ghinani@dfki.de; snehal.walunj@dfki.de; parsha.pahlevannejad_chaleshtori@dfki.de; christiane.plociennik@dfki.de; martin.ruskowski@dfki.de

S. Suh
German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

Department of Computer Science, RPTU Kaiserslautern-Landau, Kaiserslautern, Germany
e-mail: sungho.suh@dfki.de

J. Soldatos (ed.), *Artificial Intelligence in Manufacturing*,
https://doi.org/10.1007/978-3-031-46452-2_18

trial modules and agents to perceive and analyze their surroundings, leading to optimized operations and safeguarding the well-being of workers.

By integrating OD technology with human assistance and collaboration mechanisms, safer and more productive interactions are fostered in industrial environments. Collaborative robots, or Cobots, are increasingly deployed to augment the capabilities of human operators and improve overall productivity. Leveraging object detection technology, Cobots accurately perceive and respond to the presence of humans, ensuring safe and seamless cooperation within shared workspaces. This prioritizes safety and productivity in human–robot interactions, driving innovation and efficiency in the industrial domain.

Assistance systems play an important role where manual work is prevalent. In industrial processes such as assembly, training, or maintenance processes, these systems support minimizing the workload of humans. There are different types of assistance systems, ranging from manual workstations equipped with cameras and displays to immersive assistance systems over head-mounted devices or smart devices. Computer vision techniques such as object detection help understand the worker's environment from visual data. This understanding can be used to enrich existing software systems with object information to achieve multiple goals. With applications ranging from healthcare to the automotive industry, object detection models such as Yolo [22] and Faster RCNN [7] have gained popularity on account of their real-time detection performance. For example, in [15], an Advanced Driver Assistance System (ADAS) is equipped with real-time object detection to provide safety and a better driving experience. In use cases such as ADAS systems encountering moving objects, real-time object detection becomes even more important. Object detection also finds applications for solving industrial problems such as quality inspection.

An Augmented Reality (AR)-based assistance system uses the context of the existing reality or real environment and intends to augment it with useful information. In order to capture the real-world context, camera sensors play an integral part in the AR system. They provide real-world data, and the display serves as the counterpart helping the user visualize the system together with augmented information. The input visual data are understood using deep learning methods such as object detection and pose estimation, the output of which can be used for solving a large spectrum of problems in the form of an assistance system.

Object detection outputs can be utilized to create effective assistance systems that provide real-time recommendations and guidance. By integrating OD with head-worn devices, workers can receive valuable information and instructions related to their tasks. For instance, an OD system can detect and recognize objects in the assembly place, or potential hazards in the worker's surroundings. The system can then analyze these data and provide recommendations, such as the next step in the assembly pipeline. It can also alert the worker to the presence of a hazardous object. These recommendations are displayed on the worker's head-worn device, providing immediate and personalized assistance. By leveraging OD in human assistance systems, workers can benefit from increased safety, improved efficiency, and enhanced situational awareness, ultimately leading to a more productive and secure work environment.

Supervised learning approaches such as object detection are highly dependent on data. A large amount of image data are required for training. Synthetically generated and labeled dataset offers several advantages in this regard. It simplifies and economizes the generation of large datasets, eliminating manual labeling and reducing human errors. Its flexibility allows for easy manipulation according to specific requirements, enabling researchers to control factors such as lighting, camera angles, and object placements. This enhances the model's ability to learn from diverse scenarios and improves its generalization capabilities.

Furthermore, synthetic data are well-suited for various computer vision tasks, including the detection of small or rare objects that may be challenging to capture in real-world datasets. Its ability to simulate objects of any scale or size makes it invaluable in training models. Synthetic data can be generated in large quantities, providing ample training samples for deep learning models that require extensive labeled data. Moreover, it allows for easy augmentation, increasing dataset diversity and variability. These advantages contribute to the development of more accurate and robust deep learning models in computer vision.

In the subsequent sections, we will delve into the following aspects:

1. Background: We will explore the utilization of object detection in industrial environments and its diverse range of applications.
2. Scenarios: We will discuss two pivotal scenarios within the industrial environment setting where object detection is employed for human–robot interaction and worker assistance systems. Moreover, the respective methodology for the worker assistance systems scenario is discussed, and the results are presented.
3. Ongoing and Future Work: We will highlight currently ongoing research that involves the exploration of continual learning techniques and dataset optimization through the combination of real and synthetic datasets.
4. Conclusion: The chapter will conclude by summarizing the key findings and contributions to the topic.

2 Background

The industrial environment is undergoing a profound transformation toward greater intelligence and autonomy, thanks to the emergence of machine learning (ML) approaches. With access to abundant data and powerful hardware resources, deep learning techniques and artificial intelligence methods have become increasingly valuable. In this context, leveraging AI-based methods within industrial settings offers significant potential for reducing human errors and enhancing safety [4], particularly in scenarios involving human–robot collaboration or shared workspaces with close proximity between these two crucial agents. The exploration of human–robot interaction in complex and unpredictable environments has become a prominent research area, where AI methods can be effectively harnessed [13]. Within this landscape, deep models serve as powerful tools that excel in addressing intricate challenges encountered in the industrial domain. Their ability to learn hierarchical representations directly from raw data positions them as ideal solutions for a wide range of applications, encompassing object detection, anomaly detection, predictive maintenance, quality control, and optimization.

Deep models, capable of addressing diverse challenges encountered in industrial settings, have emerged as highly effective tools. These encompass vision-based approaches [25], natural language processing (NLP) techniques [1], as well as human wearable sensor-based approaches for HRI [16], among others. Notably, both conventional and deep-learning-based ML models have recently been harnessed for video streaming analysis with the aim of object detection [6]. Deep models, characterized by their end-to-end nature, aim to address the challenge of laborious and time-consuming feature extraction from data [11].

Vision-based AI approaches have become increasingly valuable in addressing industrial problems due to their ability to interpret and analyze visual data in real time. These approaches leverage deep learning algorithms and computer vision techniques to process images or video streams captured by cameras or sensors installed in industrial settings. By employing vision-based AI, various industrial challenges can be effectively tackled. Hence, vision-based approaches are widely recognized for their significant value across diverse scenarios. One prominent vision-based approach is object detection, which involves the classification and localization of multiple objects in the target frames or images. It has proven to be highly applicable in various scenarios, contributing to tasks such as improving safety, facilitating human–robot interaction [25], aiding in error identification for workers, and optimizing task completion time.

Given the wide range of challenges and numerous common use cases encountered by vision-based AI approaches in industrial settings, it is imperative to thoroughly investigate various aspects of AI-based approaches, especially in industrial environments. These aspects encompass examining the architecture of deep learning models, conducting comprehensive analysis and assessment of the data sources, and addressing other significant factors to ensure optimal performance and effectiveness. In the following, we dig into some of these challenges and solutions.

2.1 Dataset

In the case of industrial scenarios, there is difficulty in terms of real-world data collection. Especially in vision problems, collecting images with diverse conditions and viewpoints, and also labeling them, is an effort-intensive as well as a time-consuming task. In some cases, the frames used as the source of input can originate from various cameras within the environment. These cameras can include ceiling-mounted cameras, robot cameras, as well as head-worn cameras of mixed reality devices like HoloLens that capture workers' point of view [24]. Also in some cases, the data collection cameras are different from the edge devices on which an object detection model needs to be deployed. In such a situation, synthetic data generation becomes inevitable. Also, CAD data play an eminent role in the complete product life cycle of a manufacturing product, and for the product as well as the machine, it is readily available. These CAD models can be exploited to generate synthetic datasets. However, using CAD data directly cannot serve as a reliable solution. The CAD data only resemble the real objects in geometry; however, they lack materials and texture. Thus, there exists a large amount of distinction in the appearance of the real and CAD data. If we use CAD-based synthetic images for training, which is to be tested on real-world objects, there is a problem of domain difference, i.e., the real and synthetic data domain. In an attempt to solve this issue, the technique of domain randomization is demonstrated in [19]. Domain randomization can be achieved by randomizing various aspects of the simulation scene such as the backgrounds, illumination, orientation of the objects, etc. [3].

2.2 Architectures

Numerous prominent object detection architectures, including the R-CNN pipeline [7], Fast-RCNN [8], and Faster-RCNN [9], SSD [12], You Only Look Once (YOLO) approaches [22], and its following versions such as YOLOv7 [28], provide well-defined pipelines for detecting objects across different scenarios. These architectures offer robust methodologies for accurately identifying and localizing objects within the visual data. State-of-the-art techniques, such as YOLO object detection, have particularly excelled in enabling real-time or near-real-time object detection. This capability is of utmost importance in industrial environments, where the ability to detect objects promptly is crucial for effective use cases. By leveraging these advanced object detection methods, industries can enhance their operational efficiency, safety, and decision-making processes.

2.3 Application in Industrial Environment

Object detection is highly practical and beneficial in various broad use cases. For instance, in [21], object detection pipelines were employed to automate logistic processes within industrial environments. Saeed et al. [23] addressed the challenge of detecting faults in industrial product images, particularly focusing on small-object detection. Another case study is conducted by Usamentiaga et al. [27] for evaluating the state-of-the-art deep-based object detection models as well as semantic segmentation in the scenario of automated surface inspection in metals. In the field of robotics, [2] researched the integration of an object detection CNN-based model to leverage a robot in a sorting task.

2.4 Challenges

It is obvious that object detection in the industrial environment can facilitate many tasks especially in which a robot and a human need to collaborate and work in a shared workplace, and promotes safety by facilitating interaction between humans and robots. However, object detection in the industrial environment does face some challenges. Object detection in a complex and unpredictable industrial environment, in which interference objects similar to the goal objects can be found easily due to the similarity in shape, size, and color, and random positioning and orientation of different objects make this detection much more difficult [2]. Also, detecting small objects poses a significant challenge due to their limited representation of features.

An overview of concepts and terminology related to artificial intelligence (AI) is given in ISO/IEC 22989:2022 Information Technology—Artificial intelligence—Artificial intelligence concepts and terminology" [10].

3 Scenarios

In this section, we explore the scenarios in both projects, STAR and InCoRAP for object detection in the factory environment with specific emphasis on safety in human–robot collaboration and interaction, and human assistance systems. We study the use cases in which object detection can be utilized for safety in the industrial environment. Additionally, we investigate the application of object detection in human assistance systems. Furthermore, we will explain our recent work on context-based object detection methods, particularly for small objects in the assembly use case within the industrial environment. We dive into the details of how context-based approaches can effectively detect and identify small objects in assembly scenarios.

3.1 Object Detection for Human–Robot Interaction in the STAR Project

Object detection for human detection can play a crucial role in enhancing safety in industrial environments. Although human detection can be considered as a subtask of object detection, it is facing some more complexity due to the wide range of possible appearances on account of the articulated pose, and clothing, to name a few [20]. Hence, studying human detection is vital from a safety perspective, especially in a complex, unpredictable, and dangerous industrial environment. Here are a few ways human detection will be utilized:

- Enhancing Worker Safety: Implementing object detection systems that are capable of accurately identifying and tracking human presence can greatly contribute to worker safety. These systems enable the implementation of proactive safety measures to prevent accidents or potentially dangerous situations. For instance, if a worker enters a restricted area or approaches a hazardous machine, the object detection system can promptly detect their presence and trigger warnings that can be raised in the HoloLens head-worn device or automatically shut down the equipment to mitigate potential accidents.

- Collision Avoidance: Object detection can be utilized to detect the presence of humans in the vicinity of moving machinery or vehicles. This information can be used to alert operators or autonomous systems to slow down, change direction, or stop to avoid collisions and ensure worker safety.

  Our pilot study aims to explore the practical application of ceiling cameras or robot cameras in hazardous accident scenarios occurring within an industrial environment. Our focus is specifically directed toward examining the collaboration between a moving robot and human workers. The moving robot's primary role is to assist in the transportation of objects from the warehouse to the workstation, as well as facilitating the transfer of objects between various production lines, which often occurs in an unpredictable timescale.

  To ensure the safety of workers and prevent any potential accidents, we employ localization and classification from object detection techniques for both the robot and human agents. These techniques enable us to effectively localize the position of the robots and humans within the workspace through the ceiling cameras in the environment. By real-time monitoring and analyzing of the robot and human locations, we can promptly identify and mitigate potential collisions or unsafe situations, which leads to a secure and accident-free working environment.

Overall, object detection for human detection in industrial environments brings safety by enabling proactive measures and collision avoidance. It helps create a safer working environment, reduces the risk of accidents, and enhances the well-being of workers.

3.2 Object Detection for Manual Assembly Assistance System in InCoRAP

In the InCoRAP use case, the AR-based Assistance system observes the egocentric point of view of the worker. The goal is to observe worker activity in the assembly process in order to support them through a mobile robot collaboration. The object detection model is a part of the assistance used to observe the assembly state based on the detected objects. For AR applications, the head-mounted device: HoloLens2 is used. The detected objects are the observations that correspond to the assembly steps, and these observations are later used to support the worker.

Research focusing on the evaluation of AR systems has proven AR-based assistance systems advantageous over conventional instruction manuals [5]. Thus,

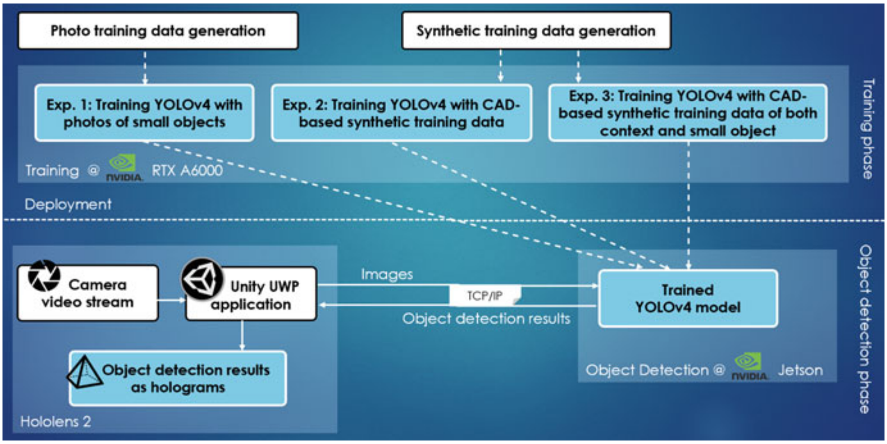
Photo training data generation
Synthetic training data generation
Exp. 1: Training YOLOv4 with photos of small objects
Exp. 2: Training YOLOv4 with CAD-based synthetic training data
Exp. 3: Training YOLOv4 with CAD-based synthetic training data of both context and small object
Training @ NVIDIA RTX A6000
Training phase
Deployment
Camera video stream
Unity UWP application
Images
TCP/IP
Object detection results
Trained YOLOv4 model
Object detection results as holograms
Object Detection @ NVIDIA Jetson
Hololens 2
Object detection phase

Fig. 1 The pipeline of the 3 experiments within the training phase with testing on the HoloLens2 frames transferred to the edge server for detecting target objects [25]

AR-based assistance places less workload on the worker or user compared to document-based guidance.

The field of small-object detection has garnered considerable research interest and has become increasingly popular. In vision-based object detection methods, the texture and arrangement of objects play a pivotal role in enabling the object detection pipeline to extract relevant features. It is worth noting that smaller objects may pose a challenge, as their size can be reduced significantly during the feature extraction process. For example, an object with dimensions of 32 × 32 pixels, after passing through five pooling layers in the VGG16 model, would be represented as a mere 1-pixel [17]. The concept of small objects, as defined in [26], encompasses objects sized at 32 × 32 pixels within the context of image analysis.

We specifically focus on small-object detection within an assembly scenario, where workers assemble various electrical components (e.g., buttons, resistors, LED, buzzers) on a breadboard to create a final product (Fig. 3b) [25]. In this scenario, a robot assists the worker by delivering parts from the warehouse, and an assistance system detects the current assembly steps and suggests the next probable part to be installed on the breadboard. The frames captured from the worker's point of view (POV) are seized using the HoloLens2, which the worker wears during the assembly process. The pipeline for this approach is illustrated in Fig. 1. In the testing phase, the frames are transmitted to the server where the object detection model performs inference. The detected objects provide information such as class identification and bounding box coordinates. This information is then communicated from the server to the HoloLens2 device using Unity communication. Subsequently, the HoloLens2 device generates holographic representations, displaying the class identification and bounding box information as augmented data. Workers in the environment can observe these holograms when they focus their gaze on the respective objects.

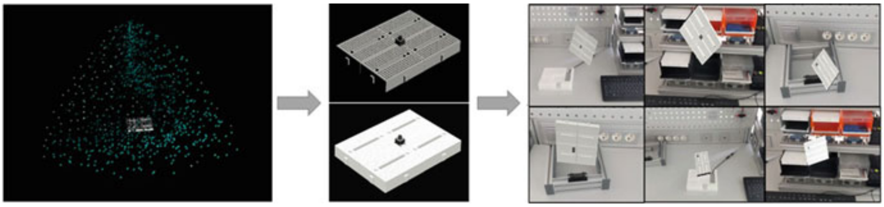

Fig. 2 Synthetic dataset generation pipeline [25]

3.3 Methodology: Context-Based Two-Step Object Detection

For computer vision problems such as image classification or object detection, the foremost task is the collection of suitable image data and then labeling it. However, this process can be automated using a game engine such as Unity [3]. In a Unity renderer space, there is a game camera that consists of the attributes of a physical camera. There are two types of views obtained in such a game development environment, one being the developer view and the other being the game view. These features allow for creating a simulation scene for synthetic image data generation. As introduced in [3], the Unity Perception allows for customization and user-defined feature development with the goal of synthetic data generation in Unity. We developed a scene similar to that of the Unity Perception Package that allows for multi-object detection dataset generation. The 3D models of the desired object are CAD models converted into Unity-compatible format and imported into the scene.

The synthetic dataset generation pipeline shown in Fig. 2 shows the CAD model of the object imported into a Unity scene. On importing into the scene, the CAD models are processed to achieve a photo-realistic appearance using the renderer features. The scene is developed in such a way that it simulates a systematic image-capturing process [3] to deliver the desired dataset. A script has been written for the game camera to capture the object images from various viewpoints and distances to the object during the simulation. From a set of background images, the backgrounds of the objects are randomized in the simulation. By rendering multiple views of the scenes, we generate a diverse set of synthetic images that mimic real-world conditions. Additionally, data augmentation techniques such as object rotation, scene illumination, and object occlusion are applied. Furthermore, the labeling tool within this Unity scene automatically annotates the generated images with bounding boxes, providing ground truth information for training and validation.

In Table 1, the benefits of using synthetic data were demonstrated in the context of installing small objects on a breadboard as an assembly process. Additionally, utilizing synthetic data presents advancements in the two-step detection approach. Table 1 illustrates the Mean Average Precision (mAP) for the detection of the small button on the breadboard for our three experiments. The mAP is a widely used evaluation metric in object detection tasks. It measures the accuracy and precision

Table 1 The mAP results for all 3 experiments for different Intersection of Union (IoU) are illustrated and confirmed that the two-step detection improved the mAP for the buttons, the target small object in the frames [25]

	mAP (full size)						mAP (cropped)					
IoU	**0.01**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.01**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**
Exp. 1	0.3%	0%	0%	0%	0%	0%	2.6%	0%	0%	0%	0%	0%
Exp. 2		44%	26%	4%	0.6%	0.03%						
Exp. 3		44%	26%	4%	0.6%	0.03%		70%	69%	58%	27%	8.5%

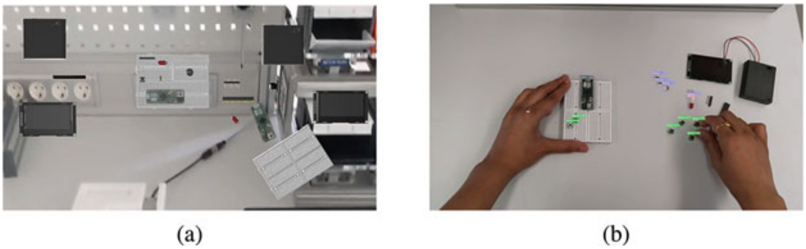

Fig. 3 (**a**) Synthetic image for assembly objects, (**b**) object detection on a video captured from HoloLens2 for the corresponding real objects

of object detection algorithms by calculating the average precision for each class and then taking the mean across all classes. The mAP is calculated according to [18]. The first experiment is trained on the 295 conventional images from buttons and tested on the 90 images containing 221 tiny buttons. The same setup is applied in the 2nd experiment, but in the training phase, we utilized the 1300 synthetic images of the breadboard and 2500 images of tiny buttons installed on the breadboard. It is clear that the mAP for the second experiment is considerably higher than the first experiment, which is around 0% for almost all different IoUs. In the third experiment, we utilized the same training and testing data as in the second experiment. However, during the inference phase, we adopted a different approach. Instead of resizing the entire frame, we first detected the breadboard and then cropped it based on the context (specifically, the Breadboard). The resulting cropped frame was then forwarded to the YOLOv4 object detection pipeline to detect the buttons as small objects. Through the third experiment, we demonstrated how employing a context-based cropping approach on the frames leads to a significant improvement in mAP (Table 1). More details can be found in [25] (Fig. 3).

4 Ongoing Research

In this section, we present the current research endeavors conducted at our research center in the field of object detection. We will highlight the ongoing projects and studies that are directly aligned with our focus on advancing object detection techniques.

4.1 Hybrid Dataset

Considering the advantages that synthetic data offer with respect to reduced human effort and time consumption, it has gained research interest. However, it can encounter challenges related to optimizing various aspects of the scene, including CAD models, lighting conditions, backgrounds, and object textures. In our recent research, we are specifically investigating the utilization of real data in combination with synthetic data to enhance the precision of object detection. By leveraging real data in this manner, we aim to optimize the object detection performance and facilitate the creation of datasets that meet our specific requirements.

4.2 Continual Learning (CL)

An additional challenge in this chapter is the need to update previously trained object detection models to accommodate new tasks, rather than retraining the model from scratch with both old and new data. This process of continual learning aims to address the issue of catastrophic forgetting, where the model's performance on previous tasks significantly declines as it is trained on new tasks [29]. This presents an interesting avenue for further exploration within the field of study.

Both domain incremental learning and task incremental learning [14] offer potential research approaches that can be applied to our specific environmental scenarios. These methods enable the model to adapt to new tasks while retaining knowledge from previous tasks. Investigating and leveraging different continual learning techniques can contribute to the development of a more flexible and efficient object detection approach.

Figure 4 depicts the utilization of continual learning to optimize the object detection process using the synthetic data in our scenarios. The procedure begins by importing CAD models of new objects and subsequently generating synthetic image datasets. These datasets are then passed to the next phase, where the previously trained YOLOv7 model is retrained using the replay approach of continual object detection. Finally, in the evaluation phase, frames captured from HoloLens are processed by the new object detection model, and the performance of this model is assessed. Thus it is an iterative process, starting from importing synthetic data to dataset generation, model retraining, and evaluation phase, aiming to expand the range of OD classes by updating the previously trained model.

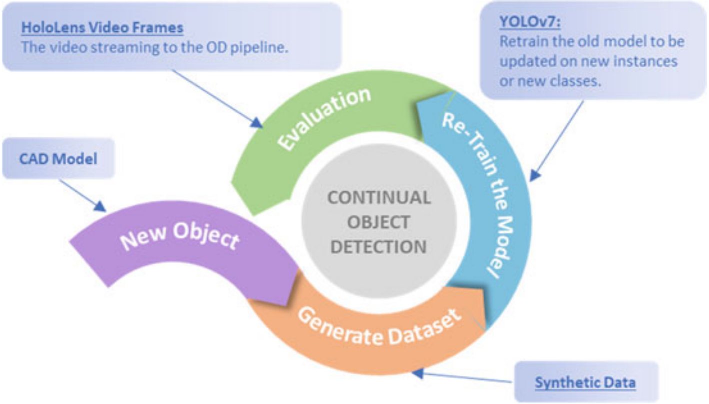
HoloLens Video Frames
The video streaming to the OD pipeline.
YOLOv7:
Retrain the old model to be updated on new instances or new classes.
CAD Model
Evaluation
Re-Train the Model
CONTINUAL OBJECT DETECTION
New Object
Generate Dataset
Synthetic Data

Fig. 4 The pipeline demonstrates the sequential process of generating synthetic data, conducting training and testing using the HoloLens2 World Camera, with the aim of establishing a continual learning pipeline for model updates

5 Conclusion

This chapter provides an exploration of object detection in industrial environments, specifically focusing on various scenarios involving human assistance systems, and safety requirements for human–robot collaboration. Additionally, we examine the application of object detection in conjunction with augmented reality devices, which offer intuitive communication interfaces for workers in these environments.

Moreover, we emphasize the substantial advantages gained from the incorporation of synthetic data in expediting the laborious data generation process and enhancing object detection outcomes. Through the utilization of synthetic data, we achieve improved efficiency and precision in object detection results. Additionally, we present our recent research focusing on the detection of small objects within industrial environments, which poses a significant challenge. We demonstrate that our approach significantly enhances the detection of small objects in manual assembly scenarios, resulting in notable improvements in performance.

Ultimately, this chapter not only sheds light on some solutions to object detection barriers in industrial settings but also paves the way for further exploration in related fields. This includes the development of sustainable practices within our scenarios and the establishment of a more generalized process that encompasses data generation and the preparation of object detection models.

Acknowledgments This work was supported by the European Union's Horizon 2020 program project STAR, under grant agreement numbers H2020-956573, and by the BMBF (German Federal Ministry of Education and Research) project InCoRAP (0IW19002).

References

1. Agnello, P., Ansaldi, S.M., Lenzi, E., Mongelluzzo, A., Roveri, M.: RECKONition: a NLP-based system for industrial accidents at work prevention. arXiv preprint arXiv:2104.14150 (2021)
2. BinYan, L., YanBo, W., ZhiHong, C., JiaYu, L., JunQin, L.: Object detection and robotic sorting system in complex industrial environment. In: 2017 Chinese Automation Congress (CAC), pp. 7277–7281 (2017). https://doi.org/10.1109/CAC.2017.8244092
3. Borkman, S., Crespi, A., Dhakad, S., Ganguly, S., Hogins, J., Jhang, Y.C., Kamalzadeh, M., Li, B., Leal, S., Parisi, P., et al.: Unity perception: generate synthetic data for computer vision. arXiv preprint arXiv:2107.04259 (2021)
4. Chen, J.H., Song, K.T.: Collision-free motion planning for human-robot collaborative safety under cartesian constraint. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 4348–4354. IEEE, New York (2018)
5. Eversberg, L., Lambrecht, J.: Evaluating digital work instructions with augmented reality versus paper-based documents for manual, object-specific repair tasks in a case study with experienced workers. arXiv preprint arXiv:2301.07570 (2023)
6. Gallo, G., Di Rienzo, F., Ducange, P., Ferrari, V., Tognetti, A., Vallati, C.: A smart system for personal protective equipment detection in industrial environments based on deep learning. In: 2021 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 222–227 (2021). https://doi.org/10.1109/SMARTCOMP52413.2021.00051
7. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
8. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
9. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
10. ISO: Information technology—artificial intelligence—artificial intelligence concepts and terminology. Standard ISO/IEC 22989:2022, International Organization for Standardization (2022)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017). https://doi.org/10.1145/3065386
12. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single shot MultiBox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21–37. Springer, Berlin (2016)
13. Liu, Z., Liu, Q., Xu, W., Liu, Z., Zhou, Z., Chen, J.: Deep learning-based human motion prediction considering context awareness for human-robot collaboration in manufacturing. Procedia CIRP **83**, 272–278 (2019). https://doi.org/https://doi.org/10.1016/j.procir.2019.04.080. https://www.sciencedirect.com/science/article/pii/S2212827119306948, 11th CIRP Conference on Industrial Product-Service Systems
14. Menezes, A.G., de Moura, G., Alves, C., de Carvalho, A.C.: Continual object detection: a review of definitions, strategies, and challenges. Neural Netw. (2023)
15. Murthy, J.S., Siddesh, G.M., Lai, W.C., Parameshachari, B.D., Patil, S.N., Hemalatha, K.L.: ObjectDetect: a real-time object detection framework for advanced driver assistant systems using YOLOv5. Wirel. Commun. Mob. Comput. **2022**, 10 (2022). https://doi.org/10.1155/2022/9444360
16. Neto, P., Simão, M., Mendes, N., Safeea, M.: Gesture-based human-robot interaction for human assistance in manufacturing. Int. J. Adv. Manuf. Technol. **101**, 119–135 (2019)
17. Nguyen, N.D., Do, T., Ngo, T.D., Le, D.D.: An evaluation of deep learning methods for small object detection. J. Electr. Comput. Eng. **2020**, 1–18 (2020)

18. Padilla, R., Passos, W.L., Dias, T.L., Netto, S.L., Da Silva, E.A.: A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics **10**(3), 279 (2021)
19. Pasanisi, D., Rota, E., Ermidoro, M., Fasanotti, L.: On domain randomization for object detection in real industrial scenarios using synthetic images. Procedia Comput. Sci. **217**, 816–825 (2023). https://doi.org/https://doi.org/10.1016/j.procs.2022.12.278. https://www.sciencedirect.com/science/article/pii/S1877050922023560, 4th International Conference on Industry 4.0 and Smart Manufacturing
20. Paul, M., Haque, S.M., Chakraborty, S.: Human detection in surveillance videos and its applications-a review. EURASIP J. Adv. Signal Process. **2013**(1), 1–16 (2013)
21. Poss, C., Ibragimov, O., Indreswaran, A., Gutsche, N., Irrenhauser, T., Prueglmeier, M., Goehring, D.: Application of open source deep neural networks for object detection in industrial environments. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 231–236 (2018). https://doi.org/10.1109/ICMLA.2018.00041
22. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
23. Saeed, F., Ahmed, M.J., Gul, M.J., Hong, K.J., Paul, A., Kavitha, M.S.: A robust approach for industrial small-object detection using an improved faster regional convolutional neural network. Sci. Rep. **11**(1), 23390 (2021)
24. Su, Y., Rambach, J., Minaskan, N., Lesur, P., Pagani, A., Stricker, D.: Deep multi-state object pose estimation for augmented reality assembly. In: 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 222–227. IEEE, New York (2019)
25. Tavakoli, H., Walunj, S., Pahlevannejad, P., Plociennik, C., Ruskowski, M.: Small object detection for near real-time egocentric perception in a manual assembly scenario. arXiv preprint arXiv:2106.06403 (2021)
26. Torralba, A., Fergus, R., Freeman, W.T.: 80 million tiny images: a large data set for nonparametric object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **30**(11), 1958–1970 (2008)
27. Usamentiaga, R., Lema, D.G., Pedrayes, O.D., Garcia, D.F.: Automated surface defect detection in metals: a comparative review of object detection and semantic segmentation using deep learning. IEEE Trans. Ind. Appl. **58**(3), 4203–4213 (2022)
28. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696 (2022)
29. Wang, L., Zhang, X., Su, H., Zhu, J.: A comprehensive survey of continual learning: theory, method and application. arXiv preprint arXiv:2302.00487 (2023)