

Smart Detection System of Safety Hazards in Industry 5.0

Stavroula Bourou ^{1,*}, Apostolos Maniatis ¹, Dimitris Kontopoulos ¹ and Panagiotis A. Karkazis ²

¹ Synelixis Solutions S.A., GR34100 Chalkida, Greece; maniatis@synelixis.com (A.M.); kontopoulos@synelixis.com (D.K.)

² Maggioli Group SPA, 47822 Santarcangelo di Romagna, Italy; panagiotis.karkazis@maggioli.gr

* Correspondence: bourou@synelixis.com

Abstract: Safety management is a priority to guarantee human-centered manufacturing processes in the context of Industry 5.0, which aims to realize a safe human–machine environment based on knowledge-driven approaches. The traditional approaches for safety management in the industrial environment include staff training, regular inspections, warning signs, etc. Despite the fact that proactive measures and procedures have exceptional importance in the prevention of safety hazards, human–machine–environment coupling requires more sophisticated approaches able to provide automated, reliable, real-time, cost-effective, and adaptive hazard identification in complex manufacturing processes. In this context, the use of virtual reality (VR) can be exploited not only as a means of human training but also as part of the methodology to generate synthetic datasets for training AI models. In this paper, we propose a flexible and adjustable detection system that aims to enhance safety management in Industry 5.0 manufacturing through real-time monitoring and identification of hazards. The first stage of the system contains the synthetic data generation methodology, aiming to create a synthetic dataset via VR, while the second one concerns the training of AI object detectors for real-time inference. The methodology is evaluated by comparing the performance of models trained on both real-world data from a publicly available dataset and our generated synthetic data. Additionally, through a series of experiments, the optimal ratio of synthetic and real-world images is determined for training the object detector. It has been observed that even with a small amount of real-world data, training a robust AI model is achievable. Finally, we use the proposed methodology to generate a synthetic dataset of four classes as well as to train an AI algorithm for real-time detection.



Citation: Bourou, S.; Maniatis, A.; Kontopoulos, D.; Karkazis, P.A. Smart Detection System of Safety Hazards in Industry 5.0. *Telecom* **2024**, *5*, 1–20.

<https://doi.org/10.3390/telecom5010001>

Academic Editor: Przemysław Falkowski-Gilski

Received: 31 October 2023

Revised: 7 December 2023

Accepted: 20 December 2023

Published: 22 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Industry 5.0; employee safety; intelligent detection system; synthetic dataset; virtual reality; AI object detector; PPE detection

1. Introduction

Over the years, numerous industrial revolutions took place, transforming the industrial scene. These evolutionary phases introduce the progressive integration of novel technologies into manufacturing procedures, aiming to enhance efficiency, productivity, and economic growth. Nowadays, Industry 5.0 is emerging, placing humans at the center of production. Specifically, the human-centered Industry 5.0 emphasizes the integration of smart technologies, automation, and data exchange in manufacturing. Beyond the human-centric focus of the Fifth Revolution, it also induces increased resilience and an improved emphasis on sustainability.

The human-centric Industry 5.0 places a high priority on employee safety in manufacturing, emphasizing a knowledge-driven approach to human–machine–environmental safety. Intelligent safety management that goes beyond conventional measures becomes necessary to deal with complex human–machine–environment interactions. Moreover, based on the capabilities and opportunities of Industry 5.0, it becomes pivotal to design and develop safety management strategies that not only address the unique challenges of each manufacturing setting but are also robust enough to be adapted across varying operational landscapes.

By leveraging advanced technologies, an efficient, flexible, and adaptable monitoring system can be established. Specifically, object detection algorithms can be integrated into video surveillance systems to analyze the footage in real-time and identify potential hazards within an industrial environment. However, the accuracy of such a complex process, based on AI methods, is strongly dependent on datasets.

The acquisition of real data poses significant challenges both in terms of cost and safety, specifically in cases where human participation is required. Especially in hazardous environments, such as manufacturing, obtaining real data involves significant risks, mainly concerning the safety of personnel involved in data collection. Those challenges can be addressed by utilizing game development platforms, such as Unity, to generate high-quality synthetic data via virtual reality (VR). This not only diminishes the time spent on data collection and annotation but also substantially reduces the requisite human effort and cost. Moreover, synthetic datasets can be created to accommodate the specifics of various industrial scenarios. VR acts as an immersive simulation tool, offering safe and controlled environments that mimic real-world scenarios. VR is utilized in various fields, such as education, healthcare, engineering, etc., with potential enhancements through sensors available in VR systems [1].

One of the fundamental methods to protect workers is to monitor and control their exposure to hazards, as well as to detect and identify potential risks in the workplace. According to the National Institute for Occupational Safety and Health (NIOSH) in the U.S., the sequence of control measures begins with the utilization of personal protective equipment (PPE), which refers to specialized gear or clothing intended to protect individuals from potential hazards in the workplace. PPE is used to minimize the risk of injury or exposure to various physical or other types of hazards. Some examples of PPE include safety helmets, vests and other protective clothing, safety goggles, etc.

In this research paper, we delve into the importance of a flexible and adjustable detection system of safety hazards for Industry 5.0. Initially, we analyze the five industrial revolutions. More focus is given to Industries 4.0 and 5.0, presenting the benefits of them as well as the needs that pushed the advent of the fifth revolution. Additionally, considering the human-centric nature of Industry 5.0, we examine the necessity of flexible and adaptable safety management methods in manufacturing, leveraging advanced technologies, and highly respecting human-machine-environment interactions.

Moreover, we propose a flexible and adjustable detection system that can be exploited by factory safety management to detect hazards in real-time. Considering the importance of PPE utilization regarding personnel safety, we focus on the detection of that equipment to enhance industrial safety. However, the proposed system is independent of the use case and can be applied to various scenarios and environments. The first stage of our system is the synthetic data generation methodology, which involves the creation of large-scale annotated datasets using 3D software, such as Blender, and a game development platform, like Unity. The various steps to achieve this are detailed, and all the mentioned information can be adjusted. The generated data can be modified and restructured to suit evolving requirements or to simulate new environments, enhancing the adaptability of the proposed methodology. The second stage concerns the training and evaluation of a training model that can be deployed on video surveillance systems to identify the target hazards in real-time. We evaluate the methodology in a practical scenario in which the performance of AI object detection models trained both on real-world data from the CHV dataset and on our synthetic data is compared. Additionally, we conduct a series of experiments to determine the optimal ratio of synthetic and real data for constituting the training set of object detectors, aiming to achieve the highest possible performance with the minimum number of real-world samples. Finally, by utilizing the proposed methodology, we create a synthetic dataset of four PPE classes, namely vest, helmet, glove, and goggle. We train an object detector on this dataset, and we employ real-world images for inference, managing to achieve real-time detections, proving that the detection system can be exploited for real-time applications. The detection system is illustrated in Figure 1. To the best of

our knowledge, this is the first study to apply synthetic dataset generation methodology, utilizing a game development platform for PPE detection in manufacturing and offering insights about the capabilities and limitations of exploiting synthetic data for real-world applications in this domain.

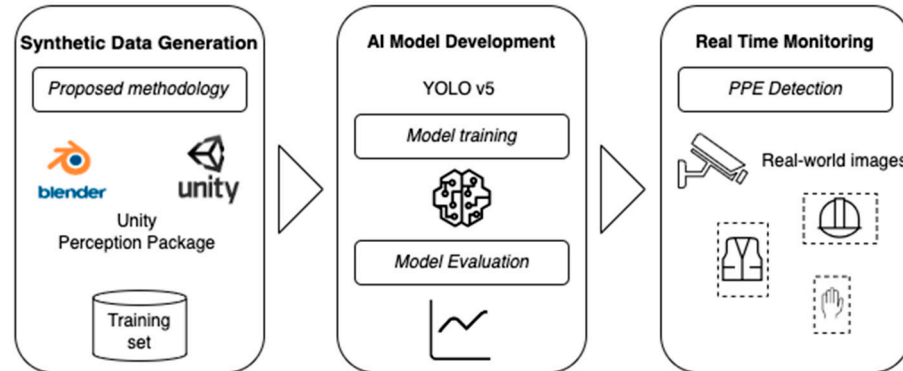


Figure 1. The proposed flexible and adjustable detection system for real-time hazard detection.

The remainder of the paper is organized as follows: Section 2 presents the different industrial evolutions, focusing on Industry 5.0 and its requirements for flexible and adaptable safety management methods. Section 3 provides the related work and background for synthetic dataset generation methods as well as AI/ML object detection algorithms. Section 4 introduces our proposed smart detection system, including the generation of annotated data and the creation of an AI object detection model, while Section 5 presents the experiments and the results. Finally, the paper concludes in Section 6.

2. Industry 5.0: Building a Human-Centric Industry

The evolution of industrial practices over the past centuries has been marked by successive revolutions, each introducing a new era of production and efficiency. Industry 5.0 is the newest one, aiming to place humans at the heart of production processes. This section presents the various industrial revolutions, focusing on Industry 5.0. Moreover, it describes the need for flexible and adaptable safety management methods for Industry 5.0.

2.1. Evolution of Industrial Revolutions and Industry 5.0

The first industrial revolution, Industry 1.0, rooted in the late 18th century, marked the transition from manual production methods to mechanized ones through the utilization of water and steam power. This shift brought about an increase in production and was primarily driven by the need for greater output and advancements in engineering. Progressing to Industry 2.0 in the early 20th century, the focus transitioned to mass production and assembly line techniques powered mainly by electricity. This change was necessitated by the growing demands of the growing global population and was characterized by the assembly lines of the automotive industry. Industry 3.0 came about in the late 20th century and was focused on the integration of computers and automation into the production process. The motivation for this transition was the rapid developments in electronic technology and the need for more precision, speed, and efficiency in production.

Industry 4.0 began in 2011 with an initiative in the high-tech strategy of the German government [2]. It describes the transformation of traditional industries through the integration of digital technologies, automation, and AI. It refers to the digitization of the manufacturing sector, which is driven by the rise of data, connectivity, analytics, and advancements in robotics. The integration of digital technologies into manufacturing reduces setup and processing times and labor and material costs, resulting in higher productivity in production [3]. Furthermore, this integration mitigates energy consumption and minimizes waste generation [4]. The capabilities of Industry 4.0 solutions allow the collection, analysis, and interpretation of a vast amount of data in real-time, empowering the rapid acquisition

of actionable insights and enhancing decision-making accuracy. However, from a socially sustainable perspective, the technological changes associated with Industry 4.0 should carefully recognize the central importance of human participation in the loop [5].

The idea of Industry 5.0 appears as an extension of Industry 4.0, originating from the observation that Industry 4.0 places less emphasis on the principles of social fairness and sustainability but more on digitalization and AI-driven technologies. Consequently, the concept of Industry 5.0 was introduced in 2017, emphasizing the importance of research and innovation to support the industry in its long-term service to humanity and adding humans into the equation while respecting planetary boundaries [6] and social aspects. Specifically, the Fifth Industrial Revolution emphasizes the necessity of placing humans at the heart of production processes, involving them in every step. This means that smart machines and robots are working together with humans to improve the efficiency of industrial production, considering the environmental impact. Additionally, Industry 5.0 identifies the ability of industry to fulfill social objectives by taking into account the limitations of our planet's resources as well as placing importance on employee health [7].

Industry 5.0 can yield numerous advantages for the manufacturing sector. Firstly, by combining advanced technologies with human intelligence, manufacturers can attain significant improvements in productivity and efficiency, resulting in enhanced competition while reducing costs. The enhanced collaboration between humans and machines holds the potential to establish manufacturing processes where machines handle routine tasks while humans undertake more complex tasks demanding advanced skills like innovation, decision-making, and problem-solving. Therefore, the accuracy and speed of manufacturing activities can be improved [8]. Industry 5.0 can also improve efficiency via collaborative robots, known as cobots, which can safely work together with human employees, offering assistance in activities like assembling, packaging, and ensuring quality [9,10]. Moreover, Industry 5.0 focuses on the creation of a safe and secure working environment to prioritize physical and mental health as well as the wellbeing of workers within the production process while protecting fundamental rights of workers, such as dignity and privacy.

2.2. Employee Safety in Industry 5.0 Manufacturing via Safety Management

In human-centered Industry 5.0, ensuring employee safety in manufacturing is of high importance. The manufacturing sector has historically been characterized by elevated injury rates stemming from the complexities and risks associated with its operations. Specifically, in 2020, the EU reported that manufacturing had the highest number of non-fatal accidents (18.2% of the total) and was the sector with the second-highest number of fatal accidents (14.6% of the total) [11]. By establishing protective measures, employees can concentrate on their tasks rather than being preoccupied with potential risks and hazards in their environment. Studies indicate that employees who perceive their workplace as secure and safe tend to exhibit enhanced performance compared to those who feel insecure [12,13].

Since Industry 5.0 places humans at the center of the new-generation manufacturing system, emphasis should be given to human-machine-environmental safety based on knowledge-driven approaches. One fundamental component to ensure human-centered manufacturing towards Industry 5.0 is the establishment of intelligent factory safety management [14]. Conventional approaches to safety management involve employee training, routine inspections, the utilization of warning signs, etc. Those preventive measures hold significant importance for proactively mitigating safety hazards and risks. However, those traditional measures ignore the complexity of human-machine-environment coupling, which requires more sophisticated approaches able to provide automated, reliable, real-time, and cost-effective safety management methods.

In the context of Industry 5.0, advanced technologies can be utilized to create intelligent factory safety management. This can be achieved by establishing an efficient monitoring mechanism to identify potential hazards in manufacturing. The system could comprise cameras strategically positioned in the space to capture real-time footage, which is then processed by object detection algorithms trained specifically on application-related data.

Real-time detection can trigger alerts to relevant personnel when non-compliance events occur. The robustness and efficiency of the monitoring system heavily depend on the available data for object detection training.

However, the complexity of the interaction between humans, machines, and the environment imposes challenges in formulating robust safety management strategies. A significant challenge lies in the adaptability of these safety methods; while one approach can be suitable for a particular case, it might not be relevant or applicable to another [15]. Therefore, it is crucial to develop safety management approaches that can be adaptable for exploitation in various manufacturing environments. This can be achieved by generating data tailored to specific applications and scenarios.

Flexible and adaptable safety management for Industry 5.0 manufacturing can be designed by employing our proposed system. Initially, high-quality synthetic datasets can be generated via virtual environments (VEs) by exploiting the proposed synthetic data generation methodology. These data are artificially produced instead of being acquired from real-world events, which offers numerous notable advantages, including the possibility for adaptability across diverse environments. The methodology of this paper provides the steps to create synthetic datasets that can simulate a wide range of conditions, variables, scenarios, and settings, which becomes particularly valuable in environments where a large amount of real-world data acquisition is challenging or costly. Afterward, AI object detectors can be trained on synthetic datasets and utilized for real-time monitoring based on data streams from cameras. Additionally, innovative architecture optimizes network and service performance [16]. For the networking of the monitoring system, the capabilities of Virtual Network Embedding (VNE) can also be explored [17].

3. Related Work and Background

In the rapidly evolving landscape of AI/ML applications, numerous studies have explored the possibilities and opportunities of creating and utilizing synthetic datasets for a range of ML tasks, especially object detection. Over the years, advancements in synthetic dataset generation have led to the development of more powerful ML models capable of accurately processing complex tasks. This section contains foundational knowledge in the field of synthetic dataset generation as well as in ML object detection.

3.1. Synthetic Datasets

The acquisition of comprehensive and high-quality datasets is a cornerstone of modern research and innovation across scientific and industrial domains. In several fields, such as computer graphics, datasets have stopped containing exclusively real data for over a few decades, with the appearance of synthetic datasets for graphics rendering and animation purposes [18]. The term synthetic dataset refers to a structured collection of data generated artificially through computational methods, mathematical models, or simulations to simulate real-world scenarios or phenomena. These datasets are intentionally created to mimic the characteristics and statistical properties of real data, but they are not derived from measurements. Their creation involves the application of algorithms, statistical techniques, and domain knowledge to replicate the complexity and diversity of authentic datasets, making them essential in situations where obtaining real data is challenging.

On the contrary, traditional data collection approaches often impose substantial financial burdens, technical complexities, and time constraints. In sectors like Industry 5.0, where manufacturing obstacles multiply, data gathering can pose risks to human safety and the integrity of industrial processes. In scientific fields such as disease diagnosis, real patient data address privacy and ethical concerns [19], and in disaster detection systems, real data may not contain enough examples of rare and extreme conditions [20]. Synthetic data generation offers a transformative solution that transcends these challenges, revolutionizing dataset utilization by providing a safe, scalable, and cost-efficient means of obtaining data that aligns with the requirements of modern research and industrial applications.

Synthetic data play a vital role in the development of AI models, especially when the real-time data are either insufficient or their acquisition is an expensive and risky procedure. This kind of dataset enables robust training, validation, and testing of AI models in various scenarios. Moreover, synthetic datasets can be designed to address diverse AI tasks, ensuring customized data availability for diverse applications. Specifically, researchers have turned their focus on game development platforms such as Unity [21] and Unreal Engine [22] for data provision by leveraging their capabilities for 3D modeling, simulation, and rendering. These engines are supplied with tools for generating virtual environments, objects, and characters, enabling the creation of highly realistic and controllable scenarios. By manipulating the game engines' parameters, researchers can simulate a wide range of real-world conditions, including lighting, weather, terrain, and physics. This flexibility is invaluable for generating diverse and complex datasets for tasks like autonomous vehicle training, robotics, and computer vision. Incorporating the functionality for camera placement and movement, capturing images and videos can happen from various viewpoints, a factor crucial for training machine learning models to understand and navigate real-world scenes, while the process of annotating, which is conventionally labor-intensive and time-consuming, can be automated through utilization of available information about object's 3D and 2D bounding boxes [23].

The use of synthetic datasets has seen a big surge in object detection. For instance, Boyong He et al. [24] turned to Unity 3D to create a dataset for training algorithms to spot ships in aerial images. This move cut down on the high costs and labor of obtaining and annotating real aerial images. In a similar vein, Kai Wang et al. [25] tackled the challenge of limited data for building smart vending machines by whipping up synthetic images using Unity, boosting the dataset available for smart vending machine development. On a different front, Jonathan Tremblay et al. [22] contributed to object detection by crafting a synthetic dataset focused on household objects for the purpose of improving robotic scene understanding algorithms. Notably, they built this synthetic dataset in a virtual environment using the Unreal Engine platform.

Focusing on the domain of facilities management, the authors in [26] propose an approach to enhance the performance of object detection algorithms. Specifically, this approach involves generating synthetically labeled images by leveraging pre-existing 3D building models and inserting them into a graphic engine. For the industrial case, the work in [27] presents a dataset with 200 K synthetic images for object detection purposes. The dataset is generated using a renderer, and the scenes are created in collaboration with different 3D experts. In addition to object detection, synthetic data also finds application in semantic segmentation. The paper in [28] demonstrates an approach that generates synthetic data to effectively train semantic segmentation AI algorithms in urban environments.

Synthetic data can be exploited for construction site safety management. The work [29] presents a data-driven worker detection approach that generates synthetic data from the virtual environment. These samples are used to train object detection AI models to detect workers in sites from load-view crane cameras. The task of detecting and tracking construction workers is also tackled in [30]. The paper presents the process for creating synthetic scenes with a 3D creation suite, while these data are used to train an object detection algorithm. Moreover, the authors in [31] create a synthetic dataset that can be utilized for worker fall detection based on a virtual environment. Table 1 summarizes the above-mentioned research works.

Although a valuable tool, the usage of synthetic datasets comes with certain drawbacks. One prominent limitation is the potential for a lack of fidelity to real-world data. While synthetic data can be meticulously crafted to replicate real-world scenarios, they may not fully capture the complexity, variability, and nuances of genuine environments [32]. As a result, when applied to real-world settings with unexpected challenges and unmodeled factors, models trained on synthetic datasets may struggle to generalize effectively. Furthermore, the accuracy of object annotations in synthetic datasets is heavily dependent on the

quality of the modeling and annotation processes, while annotation errors or discrepancies can impede model training and evaluation.

Table 1. Research works on synthetic data creation using virtual environments.

Research Paper	Domain	Task
Boyong He et al. [24]	Maritime surveillance	Ship recognition in aerial images
Kai Wang et al. [25]	Robot scene understanding	Object detection in vending machine
Tremblay et al. [22]	Facilities management	Objects detection for the household environment
Rampini and Re Cecconi [26]	Industry	Facility management component object detection
Akar et al. [27]	Urban scene understanding	Dataset for object detection
Saleh et al. [28]	Construction site safety management	Semantic Segmentation
Sutjaritvorakul et al. [29]		Worker detection
Neuhausen et al. [30]		Worker detection and tracking
Lee and Lee [31]		Worker fall detection

3.2. Object Detection Algorithms

Object detection is a fundamental computer vision task aiming to identify and localize instances of specific objects within images or video frames. This task demands both classification, where the class of the object must be identified, and localization, where the position of this object within the image must be determined. The rapid developments of Deep Learning techniques, particularly convolutional neural networks (CNNs), which hierarchically learn the relevant features from raw image data, have played a crucial role in revolutionizing the way object detection is developed and performed.

Object detection algorithms can be divided into two main approaches, namely two-stage and one-stage methods. Two-stage algorithms, such as R-CNN (region-based CNN) [33] and its advanced variants like Fast R-CNN [34] and Faster R-CNN [35], first propose a series of bounding boxes and then classify them. One-stage methods, like SSD (Single Shot Multibox Detector) [36] and YOLO (You Only Look Once) [37], detect objects in a single pass through the network. Two-stage methods are often more accurate but computationally expensive, while one-stage algorithms are generally faster and, therefore, suitable for real-time applications; however, they sacrifice some accuracy. In the case of real-world applications, in which the inference time is important, YOLO is the preferable choice, allowing faster processing without significant loss of accuracy. Throughout the years, YOLO experienced a series of iterative enhancements, progressing from the initial YOLOv1 [37] to a more sophisticated YOLOv4 [38] and YOLOv5 [39]. Each iteration has introduced improvements in speed, accuracy, and functionality.

Object detection is widely used in various applications, including autonomous driving, smart agriculture, healthcare, industrial automation, and safety. Specifically for industrial safety, by employing object detection, workplace safety can be enhanced, preventing accidents and ensuring regulatory compliance. In industrial environments, the capabilities of object detection systems to instantly identify objects are valuable. By continuously monitoring the surroundings, these systems can promptly detect potential hazards such as moving machinery, falling objects, or the absence of PPE.

Numerous research studies have extensively explored Deep Learning object detection algorithms to detect PPE in different environmental conditions. Several studies explicitly focus on safety helmet detection in various hazardous environments. The study in [40] presents considerable advantages for detecting safety helmet wear on construction sites using convolutional neural networks for face detection and bounding box regression. However, this work does not consider variations in environmental conditions. The capabilities of YOLOv4 and YOLOv5 are explored for accurate and real-time helmet detection in the construction site [41] and the power industry [42], respectively. The results of those studies indicate that they can meet the requirements of real-world scenarios, but the detection accuracy should be further improved.

Several studies focus on the exploitation of Machine Learning or Deep Learning methods to detect more than one object of PPE. The authors in [43] present three DL models based on YOLO architecture to identify if the workers are wearing hard hats and vests based on the Pictor-v3 dataset, a self-obtained and annotated image dataset that is created by crowd-sourcing and web-mining techniques. One of the main limitations of vision-based detection methods is that they are prone to occlusion, poor illumination, and image blurriness. The work in [44] introduces a system designed to inform supervisors when a worker is not wearing a helmet or vest. Another relevant paper in this field is demonstrated in [45], in which the authors develop a detector that detects in real time if people wear face masks, face shields, and hand gloves. The detector is trained on a combination of both captured and collected images from Google. In both [44,45] papers, YOLO v4 is used as an architecture model with satisfactory results.

The work described in [46] presents a system that enhances workers' safety by employing a camera that detects the utilization of PPE using the YOLO v4 algorithm. To achieve this, the researchers collect data from public resources and manually annotate it. The created dataset consists of five classes, namely hardhats, safety vests, safety gloves, safety glasses, and hearing protection. The system provides good detection accuracy; however, it shows poor performance in uncontrolled environments and in the detection of small and occluded objects. One relevant paper in that field is [47], which proposes a system for real-time detection of PPE, namely helmets, vests, and gloves, using deep neural networks on video streams. The authors of this work also enhance public datasets with manually annotated ones while they deploy the application on a low-cost embedded system located near the camera and directly connected to it, adopting the edge computing paradigm.

4. Flexible and Adjustable Detection System

In this section, we introduce an innovative methodology designed for the generation of a large-scale annotated dataset using a game development platform. Our proposed workflow offers a versatile approach to creating annotated data for object detection algorithms without constraints related to object properties or specific fields of application. The result of this generation process is a comprehensive set of RGB synthetic images accompanied by corresponding text files containing the coordinates of bounding boxes outlining objects. The synthetic dataset sets the foundation for training the AI model. Specifically, the proposed detection system includes the methodology for the generation of synthetic datasets as well as the creation of an AI model consisting of model training and evaluation. The workflow comprises two main stages:

1. **Synthetic Dataset Creation:** The initial stage contains the processes to generate a dataset of synthetic images via a virtual environment;
2. **AI Model Creation:** The second stage focuses on the definition of the AI model architecture for object detection tasks as well as the training of the model. Additionally, the evaluation of model performance is performed based on appropriate metrics to gain insights into how well the model identifies objects.

4.1. Synthetic Dataset Creation

The proposed methodology has been designed to ensure both effectiveness and target-oriented outcomes. It produces a rich and diverse dataset that can cover a wide range of scenarios, making it a valuable asset for a variety of object detection applications. Figure 2 illustrates the proposed methodology.

The first step of the pipeline is the in-depth analysis and understanding of the application characteristics. This step involves a comprehensive analysis of potential variations in geometry, appearance, and utility of the 3D models. This analysis guides the definition of all possible scenarios to be represented within the dataset and provides insights on how parameters such as size, rotation, and hue should be manipulated to achieve this diversity.

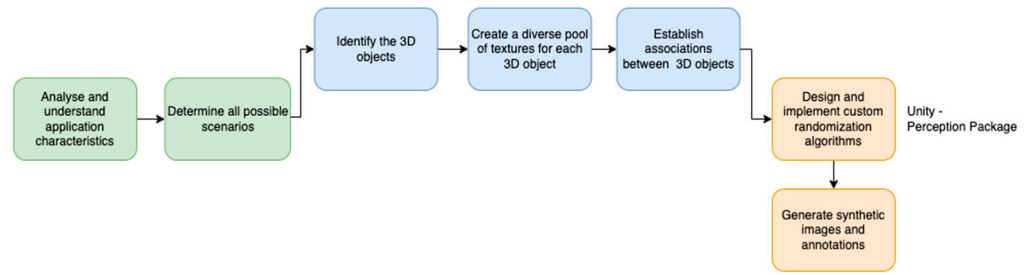


Figure 2. Proposed methodology for synthetic data generation.

Subsequently, considering the application’s purpose, the dataset requirements, and the insights extracted from the previous analysis, the 3D models that will be utilized for the generation of synthetic datasets are identified and selected. For our case, 3D models that represent the human body as well as personal protective equipment (PPE) are considered, specifically including vests, helmets, gloves, and goggles. Then, leveraging GIMP 2.0 [48], we are equipped to create a diverse pool of textures for each object, enhancing the dataset’s robustness. Subsequently, we leverage Blender [49] to establish associations between the object of interest and other objects closely linked to it in real-life scenarios, thereby improving the overall lifelikeness of the dataset. Figure 3 shows various texture variations in a 3D object, specifically for the vest (right) as well as the defined association between 3D objects (left).

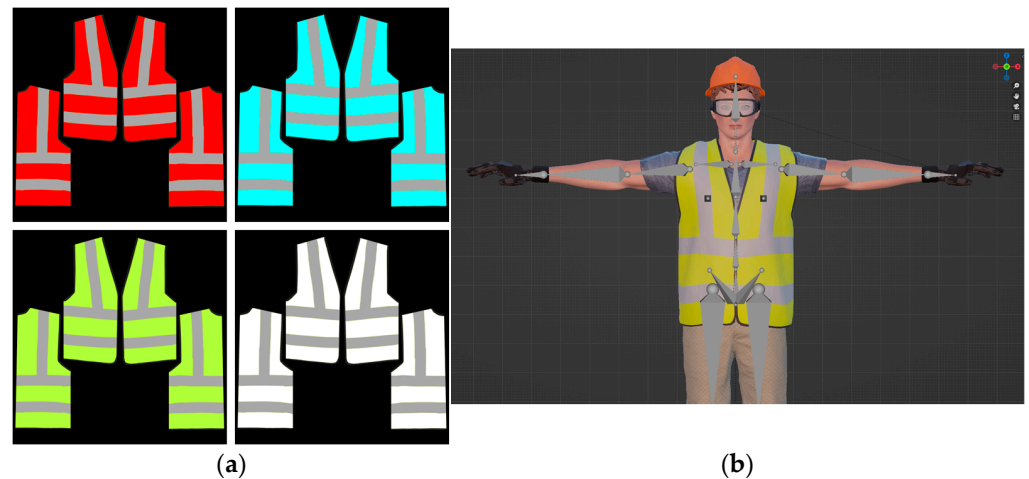


Figure 3. (a) Different variations in vest texture; (b) associations between 3D objects in a Blender environment.

After preprocessing the various 3D objects, they are integrated into the Unity 3D gaming engine, where the actual dataset generation unfolds. The dataset generation is performed using Unity Perception [23], which is a set of tools and packages provided by Unity Technologies designed to aid in the creation of synthetic datasets for the development of ML models. It streamlines the critical processes of generating and annotating large amounts of data by simulating various scenarios. Utilizing this toolkit, the user can create diverse and rich datasets by controlling the content of the scene, such as active lighting sources, visible 3D models, and their properties. Moreover, users can design and implement randomization algorithms tailored to their specific requirements. These algorithms serve to depict the desired scenarios while introducing the necessary noise to enhance dataset robustness. In our scenario, multiple custom algorithms are designed and implemented to change the scene’s parameters and options during the generation process. Table 2 presents all custom randomizer functionalities that are utilized for our case.

Table 2. Custom randomization algorithms and their functionality.

Algorithm ID	Functionality
<i>GridPicker / GridEnabler</i>	Picks/Enables random layout to be displayed
<i>SeatPicker</i>	Picks and enables random human postures
<i>ForegroundObjectRandomizer</i>	Randomly changes rotation and scale parameters of human 3D model
<i>WearablesRandomizer</i>	Picks and enables random 3D PPE on each human 3D model
<i>HueRandomizer</i>	Randomly changes the hue of the 3D object
<i>CustomTextureRandomizer</i>	Randomly changes the texture of the 3D object

To provide a realistic context, we chose to display the 3D objects as being worn by human models, with each model assuming its own unique body posture. To ensure diversity, we develop various custom randomization scripts inspired by the principles of perception, enabling us to create eight distinct layouts, each featuring a portion of the final samples. The layouts are grids of fixed positions where 3D models are being placed. In every frame, a human model equipped with safety wearables is randomly selected and placed in each corresponding position of the active layout. The selection of both the human model and the equipment category follows a stochastic pattern. It is noteworthy that each spawn point imposes constraints on the models, governing their size and rotation within well-defined boundaries that respect the objects' geometry and utility. In Figure 4, a layout of 2 and 8 people is shown, respectively. Each individual is depicted in a different body posture and is equipped with some or all of the PPE.

**Figure 4.** Layout of 2 (left) and 8 (right) employees.

To further increase the robustness of the AI model trained on the synthetic data, we insert background noise into the virtual scene. This takes place further away on the depth axis so as not to restrict foreground objects' visibility. There, we place a 3D plane to surround the camera's viewport area, depicting real-life scenes, focusing exclusively on industrial spaces to match our objective. The chosen image changes throughout the sampling process, enhancing the noise resistance of the model during training. Figure 5 depicts the unity of the 3D space, showcasing the arrangement of background noise alongside the foreground 3D objects.

The ground truth bounding boxes are provided by Perception's [23] labeling method, which allows quick annotation of our 3D models. The created virtual scene is equipped with an orthographic camera able to capture bounding boxes of all objects of interest within its viewport. During the generation process, for every sample, the ecosystem's randomization pipeline applies all randomizing scripts, resulting in a unique frame being sampled. It should be mentioned that the number of generated images, as well as the classes of the dataset, are adjustable and can be chosen by the user. The process resulted in a pool of images, each of them having multiple instances of the foreground 3D models, along with the same amount of JSON files containing the desired annotations for the object of interest. Figure 6 illustrates a synthetic image generated following the proposed methodology.

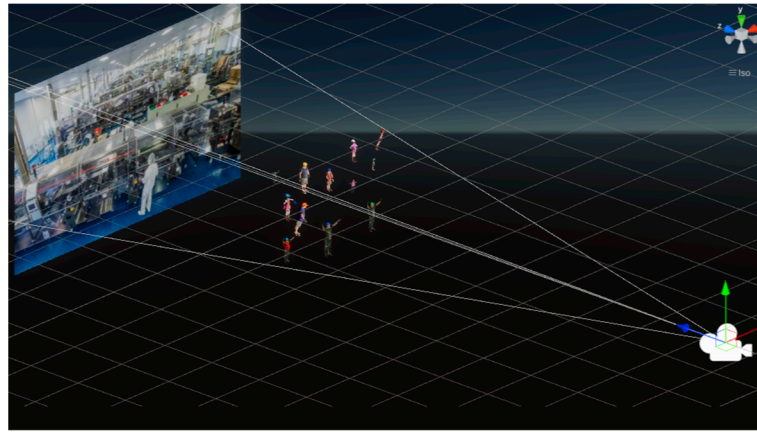


Figure 5. Unity 3D Space. From left to right, a real-life scene as background noise, foreground 3D objects, and the camera's viewpoint with the colored arrows to represent the X, Y, Z axis is 3D space.



Figure 6. Synthetic Image.

4.2. AI Model Creation

The synthetic dataset can be used to create robust object detection algorithms. Depending on the specific requirements and objectives of each application, various ML models can be considered. It is important to select an architecture that aligns well with the needs and performance criteria of the application.

4.2.1. AI Model Definition and Training

In this study, the YOLO architecture is chosen as an object detector due to its high accuracy and quick inference speed. Specifically, we focus on the YOLOv5 family [39], which offers various versions of pretrained models. Since we are interested in deploying the trained AI model for real-time applications, we decided to utilize the YOLOv5s. This version represents the “small” variation in this architecture, while it stands out with its optimized design, making it highly efficient and suitable when real-time inference is demanded. YOLOv5 employs the Generalized Intersection over Union (GIoU) loss as a bounding box regression function [50], Equation (1). GIoU can solve the inaccurate computation associated with non-overlapping bounding boxes.

$$L_{GIoU} = 1 - IoU + \frac{|C - B \cup B^{gt}|}{|C|} \quad (1)$$

where B is the predicted box, B^{gt} is the ground truth bounding box, and $IoU = \frac{B \cap B^{gt}}{B \cup B^{gt}}$ and C is the smallest box covering B and B^{gt} .

Rather than initiating the training of the YOLOv5s model from scratch, which demands a large volume of data and considerable computational resources, we utilize the YOLOv5s model pretrained on the COCO dataset [51] and then fine-tune it. By doing so, it is possible to achieve robust object detection capabilities with less data and in fewer epochs. The model is trained for 15 epochs with a batch size set at 32. The training of a model is executed on a computer with an Nvidia RTX 4500 graphic card (GPU) (<https://www.nvidia.com/en-eu/design-visualization/rtx-a4500/>, accessed on 19 December 2023), an Intel Core i9-12900 processor (CPU) (<https://www.intel.com/content/www/us/en/products/sku/134597/intel-core-i912900-processor-30m-cache-up-to-5-10-ghz/specifications.html>, accessed on 19 December 2023), and 32 GB of memory.

4.2.2. AI Model Evaluation

One of the most crucial steps in the creation of an AI model is to evaluate the performance of the trained model. The evaluation not only measures the efficacy of the trained model but also can provide insights about the quality of training data as well as indications for potential enhancements regarding the dataset generation procedures. The main goal of our evaluation process is to determine the suitability of the synthetic dataset for training an AI model effectively. Specifically, via the evaluation procedure, we can discover whether the synthetic dataset can be used to successfully train a model.

The evaluation of AI models can be achieved by quantitative metrics as well as by qualitative processes. Metrics such as Intersection over Union (*IoU*), Precision, Recall, and Average Precision (*AP*) procure quantitative insights, while the visual inspections of predicted bounding boxes provide qualitative indications. In the case of object detection tasks, the predictions of the AI model are bounding boxes accompanied by class labels. Having this in mind, it becomes essential to adopt appropriate evaluation metrics to assess the performance of object detectors and the utility of a synthetic dataset for this task. In this study, the evaluation metrics that are utilized to determine the effectiveness of trained models on synthetic datasets are Average Precision (*AP*) and Mean Average Precision, as well as visual inspection of predictions.

The Intersection over Union (*IoU*) measures the localization accuracy of an AI object detector. Specifically, the *IoU* calculates the amount of overlap between the predicted and the ground truth bounding boxes. In object detection, the AI model may predict multiple bounding boxes for each object. Based on the *IoU* and a chosen threshold, predictions can be categorized as true positives (*TP*), false positives (*FP*), or false negatives (*FN*). In practice, the *IoU* threshold is often set at 0.5. Utilizing this categorization, the precision and recall values are calculated based on Equations (2) and (3), respectively.

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

Consequently, based on those values, the precision–recall curve is designed. The numerical representation of this curve is given by the Average Precision (*AP*) in Equation (4). *AP* is computed for each object class and represents the weighted mean of precisions at each threshold with the increase in recall. A high *AP* value indicates both high precision and recall, while a lower *AP* suggests a decline in either precision or recall across a range of confidence thresholds.

$$AP = \int_{r=0}^1 p(r) dr \quad (4)$$

where r is the recall and $p(r)$ is the precision as a function of recall.

An extension of the Average Precision is the Mean Average Precision (mAP), which gives the precision for the entire model. The mAP is determined by taking the average of the AP for all the classes N , Equation (5).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

5. Experiments and Results

The proposed system is evaluated in a series of different experiments to identify its robustness in generating synthetic data that can be utilized to train AI object detectors for real-time applications.

5.1. Real-World and Synthetic Datasets Description

To evaluate the efficacy of the synthetic dataset generated following our proposed methodology, it is crucial to contrast it with an appropriate real-image dataset, which can be considered a benchmark. This can be achieved by comparing the performance of trained AI models. The models are trained on both real and synthetic datasets, and they are evaluated on the same test set to achieve comparable results. Therefore, a properly selected open dataset is crucial to ensuring a fair and insightful comparison. The samples in the synthetic dataset represent various scenarios and conditions of industrial setup. On the contrary, we selected a publicly available real dataset consisting of images acquired from real-world site environments. Both of those datasets represent the challenges encountered in PPE detection, ensuring equitable comparison by evaluating the trained AI models' effectiveness. The detailed description of those datasets is vital for providing a basis for the subsequent analysis and evaluation of AI models trained on them.

Utilizing the methodology for creating a synthetic dataset, as presented in Section 3.1, we generate a dataset of two classes, namely helmets and vests, aimed at PPE detection in industrial environments. For ease of reference throughout this paper, we named our synthetic dataset SYN_HV. SYN_HV comprises 600 images with a total of 7071 instances; among these, 3510 belong to the vest class, while 3561 belong to the helmet class. We choose to have an approximately equal number of instances in each class to prevent class imbalance, ensuring that the trained AI model has an equal opportunity to learn each class, therefore achieving more reliable and generalized performance across all classes.

Color Helmet and Vest (CHV) [52] is a novel, open-source dataset for PPE detection in construction environments. The dataset consists of 6 PPE classes, including helmets of four different colors, vests, and persons. Specifically, the CHV categorizes helmets based on their colors, namely blue, yellow, white, and red, as different colors indicate different roles on construction sites. It contains 1300 images, considering the site background, different gestures, varied angles and distances, and 9209 instances in total. The CHV dataset is split into training, validation, and test sets, comprising 80% (1064 images), 10% (133 images), and 10% (133 images) of the total dataset, respectively.

To ensure comparability among the AI models trained on both datasets, some modifications should be performed on the CHV dataset. Specifically, the different color categories of the CHV 'helmet' class might be useful in some situations; however, in our study, we are interested in recognizing the presence of a helmet, regardless of its color. For this reason, we combine the color-specific categories of CHV into a single 'helmet' category. Additionally, the 'person' category of the CHV dataset is removed since it does not align with our core focus. Regarding the 'vest' class, it is retained as it is without any modification. After the aforementioned modifications, the 'helmet' category of the CHV dataset contains 2762 instances, while the 'vest' class consists of 1396 instances. Figure 7 shows the number of instances per class for both the SYN_HV and CHV datasets.

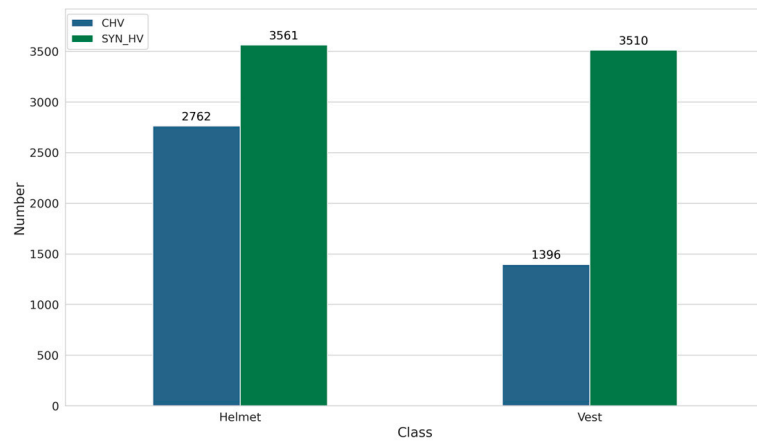


Figure 7. Number of instances per class for real-world and synthetic datasets.

It is worth noting that the validation and testing of both AI models, trained on real and synthetic datasets, are performed using the same validation and test set as proposed in the CHV dataset. This approach ensures consistency in our assessment and allows a direct comparison between the performance of the AI models, enabling us to derive valid conclusions about the usage of synthetic datasets during the training of AI models.

5.2. AI Model Training on Real-World and Synthetic Datasets

To assess the efficacy of our synthetic dataset SYN_HV, we utilize it to train a YOLOv5s model. The performance of this model is compared with the performance of YOLOv5s trained on a real-world dataset, CHV. To achieve a fair comparison, a common test set composed of real images of CHV is used. The performance of each trained model is quantified by precision, recall, AP, and mAP metrics.

At each training round, we adopt a learning rate of 0.01, combined with a momentum of 0.937. During each iteration, the model processed batches of 32 images, all resized to 640×640 pixels. To regulate the adjustment of the model weights, we apply a weight decay of 0.0005 and employ the Stochastic Gradient Descent (SGD) as our optimizer. Part of our training strategy is the use of early stopping. This precautionary measure ensured that if the model's performance did not improve over five consecutive epochs, the training would be terminated. Such an approach not only helps conserve computational resources but also acts as a deterrent to potential overfitting.

Table 3 provides a comparison of performance metrics of object detectors trained on a real-world dataset, CHV, as well as on our synthetic dataset, SYN_HV, that contains only synthetic images. Both of those models are evaluated on the same test set of real images. It is observed that the AI model trained on real images demonstrates superior performance across all metrics than that trained solely on synthetic datasets. The superior performance of the trained model on real data can be explained by its familiarity with the real-world distributions and complexities encapsulated in the test set. On the other hand, the model trained on synthetic images may have difficulties understanding the real images and predicting the correct bounding boxes. Hence, the discrepancy in performance underscores the importance of training data that closely aligns with the conditions under which the AI model will be evaluated.

Table 3. Performance metrics comparison for trained AI models on CHV and SYN_HV.

Trained Dataset	Precision	Recall	AP Vest	AP Helmet	mAP
CHV	89.6%	84.8%	86.4%	91.4%	88.9%
SYN_HV	77.5%	67.8%	67.6%	75.5%	71.6%

5.3. Experimenting with Synthetic and Limited Amount of Real-World Data

As described in Section 4.2, an AI model trained exclusively on a synthetic dataset cannot achieve as high performance as one trained on a real-world dataset. However, in many real-world scenarios, the acquisition of large-scale amounts of real data is infeasible, posing a significant challenge to the development of an effective AI model. Therefore, the combination of real and synthetic data that constitutes the training dataset seems like a promising solution, addressing the problem of limited real data. Therefore, it is crucial to define the optimal ratio of real-world and synthetic data to maximize the model's performance.

Through a series of experiments, we aim to examine the potential benefits and limitations of combining real and synthetic images for the creation of a training set that can be used to train AI models for real-world applications. To achieve this, we design a series of experiments where we begin with a relatively small number of real images and progressively augment the dataset with synthetic images. The real images are sourced from the CHV dataset, of which 50 are allocated for the training set and 25 for the validation set. In each experiment, we trained the YOLOv5s model on the combined training dataset while we evaluated it on the CHV test set, comprising 133 images, to ensure a consistent benchmark for evaluation.

Table 4 presents the number of real and synthetic images that constitute the training set for each experiment. In the first one, working under the constraints of limited data availability, the model is trained exclusively on a set of 50 real images. In all the following six experiments, the number of real images remains the same while we incrementally increase the number of synthetic images.

Table 4. Number of real-world and synthetic images for each experiment.

Experiment ID	Real Images Number	Synthetic Images Number
E1_50_0	50	0
E2_50_50	50	50
E3_50_100	50	100
E4_50_150	50	150
E5_50_300	50	300
E6_50_600	50	600
E7_50_1200	50	1200

YOLOv5s is trained on training datasets of the various experiments, and it is evaluated at the same CHV test set. Table 5 presents the mAP values of trained models for each experiment. In the initial experiment, the model, trained with only a small number of real images, achieved a mAP of 14.3%. In the second experiment, the incorporation of synthetic data led to a slight improvement in the model's performance, increasing overall mAP to 16.3%. In the third one, a significant rise in mAP is observed, escalating to 71.2%. The highest mAP value of 84.1% is achieved in the sixth experiment, in which the training set contains 50 real images and 600 synthetic ones.

Table 5. mAP values of object detector for each experiment.

	E1_50_0	E2_50_50	E3_50_100	E4_50_150	E5_50_300	E6_50_600	E7_50_1200
mAP	14.3%	16.3%	17.3%	71.2%	79.5%	84.1%	81.0%

In the last experiment, where we doubled the number of synthetic images in the training set to 1200, it was observed that the mAP value slightly decreased to 81.0%. One explanation for this decrease in performance is the model's over-adaptation to the characteristics of the synthetic data. When exposed to a significant volume of synthetic data, the model develops a bias, becoming particularly adept at recognizing objects in virtual conditions. Consequently, its ability to generalize and recognize objects in real-world scenarios could be compromised. This trend suggests a potential saturation point beyond

which adding more synthetic data may not necessarily lead to performance gains and, in fact, could risk the model's efficacy in real-world conditions, leading to diminished returns.

To ensure the robustness and reliability of our findings, we repeat the experiments listed in Table 4 two more times, each time selecting a different random set of 75 images from the CHV dataset. The methodology remains consistent. Specifically, we begin with 50 real images and then gradually introduce synthetic data to observe the impact on performance. As illustrated in Figure 8, the performance of the models showcases similar trends across all experiments, validating the authenticity of our initial results. It should be mentioned that the best performance of all models is observed when combining 50 real training images with 600 synthetic images. This configuration yields an average mAP score of 84.3%, with a standard deviation of 0.4%.

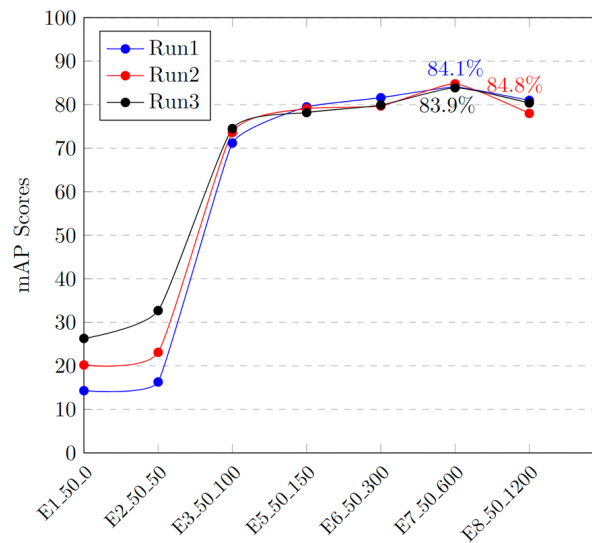


Figure 8. Model's performance across different ratios of real-world and synthetic images.

Remarkably, by employing a training set comprising just 50 real images combined with synthetic data to train an object detector, we manage to bridge the performance gap, achieving a result that is only a slight 4.7% behind the scenario of having only real images.

5.4. Synthetic Dataset of Four PPE Classes

The proposed methodology provides users and researchers with the ability to design synthetic datasets tailored to the specific objectives of their applications. Taking advantage of this, we utilized the proposed methodology to expeditiously create a dataset of four PPE classes, namely vests, helmets, gloves, and goggles. By equipping our virtual workers with this extensive set of safety equipment, we made our synthetic images more closely reflect real-world settings where workers typically wear multiple pieces of PPE simultaneously. This synthetic dataset is used to train YOLOv5s.

Having trained the model in these four distinct classes—vests, helmets, gloves, and goggles—we face a challenge in its evaluation. The scarcity of open-source datasets that encompass all these PPEs prevents the quantitative evaluation of the model. However, we perform a qualitative evaluation by inspecting the predicted bounding boxes on the images. We tested our model with real-world images, as shown in Figure 9, observing its behavior and ability to accurately detect the four different PPE classes. It should be noted that the average inference time of the model on images is 29.6 ms, demonstrating the availability of the system to identify different PPE for real-time applications.



Figure 9. Visual representation of the model's predictions on real-world images of the CHV test set.

Our future plans include enhancing the synthetic dataset with a small number of real-world images depicting instances of the four PPE classes, training the model, and integrating it into intelligent factory safety management. This aims to establish an efficient monitoring mechanism to detect non-compliance with protective measures in industrial environments. The system could comprise surveillance cameras to capture real-time footage, which is then processed by the trained object detection model. Real-time detection can trigger alerts to inform supervisors when non-compliance events occur, i.e., when employees are not equipped with the appropriate PPE.

6. Conclusions

This paper presents a robust detection system to identify hazards in manufacturing processes and enhance the safety management of Industry 5.0. Considering the characteristics of the Fifth Industrial Revolution, the proposed detection system takes advantage of more sophisticated approaches and digital methodologies, such as game development platforms and advanced AI algorithms, to provide automated, reliable, real-time, and cost-effective safety management capabilities. In addition, the system is flexible and adjustable to meet the requirements imposed by the complexity of the human-machine-environment coupling. Specifically, the synthetic dataset creation methodology provides all the methodologies required to generate data via a VR environment, ensuring that the data can be modified and restructured to simulate new scenarios and environments.

The proposed system is evaluated through a series of experiments focusing on the detection of PPE. The conducted experiments highlight the potential of generated synthetic datasets, especially in contexts where authentic data might be lacking or challenging to acquire. Specifically, YOLOv5s is trained on real-world data from the CHV dataset as

well as on synthetic data, with the model trained on the former dataset to achieve higher performance. A possible explanation for this could be that models might develop a bias towards synthetic data generated in VR environments. This underlines the necessity to include some real-world samples in the training set as well as to establish a balanced ratio between synthetic and real-world data to achieve optimal model performance.

Another notable observation from the experiments is that by using just 50 real images supplemented with 600 synthetic ones, we enrich our model's performance, bringing it to within just 4.7% of the result achieved with a dataset of 1064 real images. This is a valuable observation since it proves that an AI model can be trained with a small amount of real data, thereby significantly reducing the human effort and cost associated with data acquisition and annotation procedures. Moreover, by exploiting the capabilities of synthetic dataset creation methodology, we generate another dataset consisting of four PPE classes, namely vests, helmets, gloves, and goggles, based on which we train YOLOv5s. The model is quantitatively evaluated on real-world images since there is no openly available dataset containing those classes. The inference time of this model is 29.6 ms, illustrating its usability in real-time monitoring.

Author Contributions: Conceptualization, P.A.K.; methodology, S.B. and P.A.K.; software, A.M. and D.K.; validation, A.M. and D.K.; formal analysis, S.B., A.M. and D.K.; investigation, S.B., A.M., D.K. and P.A.K.; data curation, D.K.; writing—original draft preparation, S.B., A.M., D.K. and P.A.K.; writing—review and editing, S.B. and P.A.K.; visualization, A.M. and D.K.; supervision, P.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by the European Union's Horizon HADEA research and innovation program under grant Agreement 101092851 for the XR2LEARN project.

Data Availability Statement: The dataset used for this article is available at <https://drive.google.com/file/d/1fdGn67W0B7ShpBDbbQpUF0ScPQa4DR0a/view> (accessed on 10 August 2023).

Acknowledgments: We acknowledge the equal contribution of all the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AI	Artificial Intelligence	R-CNN	Region-based Convolutional Neural Network
AP	Average Precision	SGD	Stochastic Gradient Descent
CHV	Color Helmet and Vest	SSD	Single Shot Multibox Detector
FN	False Negative	SYN_HV	Synthetic Helmet and Vest
FP	False Positive	TP	True Positive
IoU	Intersection over Union	VE	Virtual Environment
mAP	Mean Average Precision	VNE	Virtual Network Embedding
ML	Machine Learning	VR	Virtual Reality
NIOSH	National Institute for Occupational Safety and Health	YOLO	You Only Look Once
PPE	Personal Protective Equipment		

References

1. Vretos, N.; Daras, P.; Asteriadis, S.; Hortal, E.; Ghaleb, E.; Spyrou, E.; Leligou, H.C.; Karkazis, P.; Trakadas, P.; Assimakopoulos, K. Exploiting sensing devices availability in AR/VR deployments to foster engagement. *Virtual Real.* **2019**, *23*, 399–410. [CrossRef]
2. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, conception and perception. *J. Manuf. Syst.* **2021**, *61*, 530–535. [CrossRef]
3. Koch, V.; Kuge, S.; Geissbauer, R.; Schrauf, S. Industry 4.0: Opportunities and challenges of the industrial internet. *Strategy PwC* **2014**, *13*, 5–50.
4. Waibel, M.W.; Steenkamp, L.P.; Moloko, N.; Oosthuizen, G.A. Investigating the Effects of Smart Production Systems on Sustainability Elements. *Procedia Manuf.* **2017**, *8*, 731–737. [CrossRef]
5. Kong, X.T.R.; Luo, H.; Huang, G.Q.; Yang, X. Industrial wearable system: The human-centric empowering technology in Industry 4.0. *J. Intell. Manuf.* **2019**, *30*, 2853–2869. [CrossRef]

6. European Commission. Directorate General for Research and Innovation. In *Industry 5.0: Towards a Sustainable, Human Centric and Resilient European Industry*; Publications Office: Luxembourg, 2021. Available online: <https://data.europa.eu/doi/10.2777/308407> (accessed on 10 August 2023).
7. Akundi, A.; Euressti, D.; Luna, S.; Ankobiah, W.; Lopes, A.; Edinbarough, I. State of Industry 5.0—Analysis and Identification of Current Research Trends. *Appl. Syst. Innov.* **2022**, *5*, 27. [[CrossRef](#)]
8. George, A.S.; George, A.H. Revolutionizing Manufacturing: Exploring the Promises and Challenges of Industry 5.0. *Partn. Univers. Int. Innov. J.* **2023**, *1*, 22–38. [[CrossRef](#)]
9. Østergaard, E.H. The “Human Touch” Revolution Is Now under Way. Available online: <https://industrialmachinerydigest.com/industrial-news/white-papers/welcome-industry-5-0-human-touch-revolution-now-way/> (accessed on 20 October 2023).
10. Simões, A.C.; Pinto, A.; Santos, J.; Pinheiro, S.; Romero, D. Designing human-robot collaboration (HRC) workspaces in industrial settings: A systematic literature review. *J. Manuf. Syst.* **2022**, *62*, 28–43. [[CrossRef](#)]
11. Eurostat—Statistics Explained Accidents at Work—Statistics by Economic Activity. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_-_statistics_by_economic_activity#Developments_over_time (accessed on 20 October 2023).
12. Karanikas, N.; Melis, D.J.; Kourousis, K.I. The Balance between Safety and Productivity and its Relationship with Human Factors and Safety Awareness and Communication in Aircraft Manufacturing. *Saf. Health Work* **2018**, *9*, 257–264. [[CrossRef](#)]
13. Shikdar, A.A.; Sawaqed, N.M. Worker productivity, and occupational health and safety issues in selected industries. *Comput. Ind. Eng.* **2003**, *45*, 563–572. [[CrossRef](#)]
14. Lu, Y.; Zheng, H.; Chand, S.; Xia, W.; Liu, Z.; Xu, X.; Wang, L.; Qin, Z.; Bao, J. Outlook on human-centric manufacturing towards Industry 5.0. *J. Manuf. Syst.* **2022**, *62*, 612–627. [[CrossRef](#)]
15. Wang, H.; Lv, L.; Li, X.; Li, H.; Leng, J.; Zhang, Y.; Thomson, V.; Liu, G.; Wen, X.; Sun, C.; et al. A safety management approach for Industry 5.0’s human-centered manufacturing based on digital twin. *J. Manuf. Syst.* **2023**, *66*, 1–12. [[CrossRef](#)]
16. Karkazis, P.A.; Railis, K.; Prekas, S.; Trakadas, P.; Leligou, H.C. Intelligent Network Service Optimization in the Context of 5G/NFV. *Signals* **2022**, *3*, 587–610. [[CrossRef](#)]
17. Prekas, S.; Karkazis, P.; Nikolakakis, V.; Trakadas, P. Comprehensive Comparison of VNE Solutions Based on Different Coordination Approaches. *Telecom* **2021**, *2*, 390–412. [[CrossRef](#)]
18. Raghunathan, T.E. Synthetic Data. *Annu. Rev. Stat. Appl.* **2021**, *8*, 129–140. [[CrossRef](#)]
19. Chen, R.J.; Lu, M.Y.; Chen, T.Y.; Williamson, D.F.K.; Mahmood, F. Synthetic data in machine learning for medicine and healthcare. *Nat. Biomed. Eng.* **2021**, *5*, 493–497. [[CrossRef](#)] [[PubMed](#)]
20. Pérez-Porras, F.-J.; Triviño-Tarradas, P.; Cima-Rodríguez, C.; Meroño-de-Larriva, J.-E.; García-Ferrer, A.; Mesas-Carrascosa, F.-J. Machine Learning Methods and Synthetic Data Generation to Predict Large Wildfires. *Sensors* **2021**, *21*, 3694. [[CrossRef](#)]
21. Öztürk, A.E.; Erçelebi, E. Real UAV-Bird Image Classification Using CNN with a Synthetic Dataset. *Appl. Sci.* **2021**, *11*, 3863. [[CrossRef](#)]
22. Tremblay, J.; To, T.; Birchfield, S. Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation. *arXiv* **2018**, arXiv:1804.06534. [[CrossRef](#)]
23. 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* **2021**, arXiv:2107.04259. [[CrossRef](#)]
24. He, B.; Li, X.; Huang, B.; Gu, E.; Guo, W.; Wu, L. UnityShip: A Large-Scale Synthetic Dataset for Ship Recognition in Aerial Images. *Remote Sens.* **2021**, *13*, 4999. [[CrossRef](#)]
25. Wang, K.; Shi, F.; Wang, W.; Nan, Y.; Lian, S. Synthetic Data Generation and Adaption for Object Detection in Smart Vending Machines. *arXiv* **2019**. [[CrossRef](#)]
26. Rampini, L.; Re Cecconi, F. Synthetic images generation for semantic understanding in facility management. *Constr. Innov.* **2023**. [[CrossRef](#)]
27. Akar, C.A.; Tekli, J.; Jess, D.; Khoury, M.; Kamradt, M.; Guthe, M. Synthetic Object Recognition Dataset for Industries. In Proceedings of the 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Natal, Brazil, 24–27 October 2022; IEEE: New York, NY, USA, 2022; pp. 150–155.
28. Saleh, F.S.; Aliakbarian, M.S.; Salzmann, M.; Petersson, L.; Alvarez, J.M. Effective Use of Synthetic Data for Urban Scene Semantic Segmentation. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 11206, pp. 86–103. ISBN 978-3-030-01215-1.
29. Sutjaritvorakul, T.; Vierling, A.; Berns, K. Data-Driven Worker Detection from Load-View Crane Camera. In Proceedings of the 37th International Symposium on Automation and Robotics in Construction, Kitakyushu, Japan, 26–30 October 2020.
30. Neuhausen, M.; Herbers, P.; König, M. Using Synthetic Data to Improve and Evaluate the Tracking Performance of Construction Workers on Site. *Appl. Sci.* **2020**, *10*, 4948. [[CrossRef](#)]
31. Lee, J.; Lee, S. Construction Site Safety Management: A Computer Vision and Deep Learning Approach. *Sensors* **2023**, *23*, 944. [[CrossRef](#)]
32. Koo, S.; Park, C.; Lee, S.; Seo, J.; Eo, S.; Moon, H.; Lim, H. Uncovering the Risks and Drawbacks Associated with the Use of Synthetic Data for Grammatical Error Correction. *IEEE Access* **2023**, *11*, 95747–95756. [[CrossRef](#)]
33. 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, Columbus, OH, USA, 23–28 June 2014.

34. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: New York, NY, USA, 2015; pp. 1440–1448.
35. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
36. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9905, pp. 21–37.
37. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: New York, NY, USA, 2016; pp. 779–788.
38. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection 2020. *arXiv* **2020**, arXiv:2004.10934. [[CrossRef](#)]
39. Ultralytics YOLOv5: A State-of-the-Art Real-Time Object Detection System 2021. Available online: <https://docs.ultralytics.com> (accessed on 20 October 2023).
40. Shen, J.; Xiong, X.; Li, Y.; He, W.; Li, P.; Zheng, X. Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 180–196. [[CrossRef](#)]
41. Benyang, D.; Xiaochun, L.; Miao, Y. Safety helmet detection method based on YOLO v4. In Proceedings of the 2020 16th International Conference on Computational Intelligence and Security (CIS), Guangxi, China, 27–30 November 2020; IEEE: New York, NY, USA, 2016; pp. 155–158.
42. Fu, D.; Gao, L.; Hu, T.; Wang, S.; Liu, W. Research on Safety Helmet Detection Algorithm of Power Workers Based on Improved YOLOv5. *J. Phys.Conf. Ser.* **2022**, *2171*, 012006. [[CrossRef](#)]
43. Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep learning for site safety: Real-time detection of personal protective equipment. *Autom. Constr.* **2020**, *112*, 103085. [[CrossRef](#)]
44. Ahmed Al Daghan, A.T.; Vineeta; Kesh, S.; Manek, A.S. A Deep Learning Model for Detecting PPE to Minimize Risk at Construction Sites. In Proceedings of the 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 9–11 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
45. Protik, A.A.; Rafi, A.H.; Siddique, S. Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow. In Proceedings of the 2021 IEEE Region 10 Symposium (TENSYPMP), Jeju, Republic of Korea, 23–25 August 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
46. Karlsson, J.; Strand, F.; Bigun, J.; Alonso-Fernandez, F.; Hernandez-Diaz, K.; Nilsson, F. Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers. In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods*; SCITEPRESS—Science and Technology Publications: Lisbon, Portugal, 2023; pp. 395–402.
47. Gallo, G.; Rienzo, F.D.; Garzelli, F.; Ducange, P.; Vallati, C. A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge. *IEEE Access* **2022**, *10*, 110862–110878. [[CrossRef](#)]
48. GIMP—GNU Image Manipulation Program. Available online: <https://www.gimp.org/> (accessed on 20 October 2023).
49. Blender. Available online: <https://www.blender.org/> (accessed on 20 October 2023).
50. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; IEEE: New York, NY, USA, 2019; pp. 658–666.
51. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; Volume 8693, pp. 740–755. ISBN 978-3-319-10601-4.
52. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.