POLITECNICO DI TORINO
Repository ISTITUZIONALE

Machine Learning and Digital Twin for Production Line Simulation: A Real Use Case

Original
Machine Learning and Digital Twin for Production Line Simulation: A Real Use Case / Oriti, Damiano; Brizzi, Paolo; Giacalone, Giorgio; Manuri, Federico; Sanna, Andrea; Tovar Ordoñez, Orlando. - ELETTRONICO. - 314:(2022), pp. 814-821. (6th International Conference on Human Interaction & Emerging Technologies: Future Systems (IHIET-FS 2021) CHU-Université de Reims Champagne-Ardenne, France August 27-29, 2021) [10.1007/978-3-030-85540-6_103].

Availability:
This version is available at: 11583/2910759 since: 2022-03-10T11:27:00Z

Publisher:
Springer

Published
DOI:10.1007/978-3-030-85540-6_103

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature’s AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-030-85540-6_103

(Article begins on next page)

08 May 2026

Machine Learning and Digital Twin for Production Line Simulation: A Real Use Case

Damiano Oriti [1], Paolo Brizzi[2], Giorgio Giacalone[1],
Federico Manuri[1], Andrea Sanna[1] and Orlando Tovar Ordoñez[2]

[1] Dipartimento di Automatica e Informatica, Politecnico di Torino,
c.so Duca degli Abruzzi 24, 10129, Turin, Italy
{damiano.oriti, federico.manuri, andrea.sanna}@polito.it
{giorgio.giacalone}@studenti.polito.it
[2] Competence Industry Manufacturing 4.0 S.C.A R.L,
c.so Settembrini 178, 10135, Turin, Italy
{paolo.brizzi, orlando.tovar}@cim40.com

Abstract. The advent of Industry 4.0 has boosted the usage of innovative technologies to promote the digital transformation of manufacturing realities, especially exploiting the possibilities offered by cyber physical systems and virtual environments (VEs). Digital Twins (DTs) have been widely adopted to virtually reproduce the physical world for training activities and simulations, and today they can also leverage on the integration of Machine Learning (ML), which is considered a relevant technology for industry 4.0. This paper investigates the usage of a combination of DT and ML technologies in the context of a real production environment, specifically on the creation of a DT enhanced with YOLO (You only look once), a state-of-the-art, real-time object detection algorithm. The ML system has been trained with synthetic data automatically generated and labelled and its performance enables its usage in the VE for real-time users training.

Keywords: Machine Learning · Digital Twin · Virtual Reality · Industry 4.0 · Virtual Manufacturing · Synthetic Dataset

1 Introduction

The advent of the Fourth Industrial Revolution has boosted the usage of innovative technologies to promote the digital transformation of manufacturing realities, improving the automation of traditional industrial practices and processes. Nowadays, factories should leverage digitalization to become flexible towards strong products customization and adaptive to unexpected events.

Among these relevant technologies, Digital Twins (DTs) have been widely adopted to virtually reproduce the physical world and to integrate real environments with their digital counterpart [1]. DTs can enhance manufacturing systems in many ways: they can be used to improve the system's predictions of future states by combining simulation models and real data; their simulations can support decision-making tasks, process optimization and commissioning of production operations; a virtual simulation with a precise reproduction of a real-world process would improve remote assistance, could be adopted to train technicians, and could be exploited for maintenance procedures.

Machine Learning (ML) is another crucial technology to improve the automation of manufacturing systems: it can extract useful insights through a training process on large volume datasets and then exploit the acquired knowledge in real case situations [2]. A relevant challenge in the training process is the availability of huge, labelled datasets consisting of high-quality data. To overcome this challenge, simulation tools could be exploited to automatically generate large datasets of quality data for training ML models, both in a cost and time-effective way, alleviating the need for real world data. Nowadays, digital simulations which integrate ML mechanism based on synthetic data have not been widely researched and a few papers explore this scenario. This contribution describes the development and deployment of a DT of a production line for assembling skateboards. The physical line has been deployed at the Competence Industry Manufacturing 4.0 (CIM 4.0) Center, a research center located in Turin which aims to provide the strategic and operative support tools for manufacturing-oriented enterprises toward the digital transformation of industrial processes accordingly to the Industry 4.0 vision. The production line exploits an image-processing algorithm to assist a collaborative robot (cobot) in a pick and place task. Since this step of the procedure relies on an obsolete contour detection image processing algorithm, it turned out to be prone to errors. The DT is then exploited to simulate and evaluate a novel image-processing approach based on a convolutional neural network (CNN) trained with a synthetic dataset which can provide real-time performances, thus enabling the DT to be used for real-time immersive virtual reality simulations.

The remaining of this paper is organized as follows: section 2 explores the state of the art for the proposed research topic. Section 3 provides a detailed description of the design and development of the proposed Digital Twin. Section 4 illustrates the system evaluations, with a focus on the CNN performances assessment. Section 5 summarizes the paper contents and suggests possible future works.

2 Previous Works

Recent advancements in information technologies have been proven to have wide applications in smart manufacturing [4]. The adoption of artificial intelligence, enhanced algorithms and virtualization of production lines in manufacturing-oriented enterprises can lead to real-time, self-organized control and coordination of the industrial process through digitization and simulations [5]. DTs have been exploited in many ways to enforce the innovation process defined through the Industry 4.0 paradigm. Nikolakis et al. [6] proposed a DT framework to optimize the planning and commissioning of production process through simulation. Tao et al. [1] reported several DT applications which have been successfully applied in manufacturing processes to reduce costs and to improve performance. ML has been exploited as well as DT to improve manufacturing processes. CNNs, whose main purpose consists of image processing and analysis, have been widely adopted in industry for: automatic defect identification [7], system optimizations [8], product diagnostics [9], classification tasks [10], and for part detection and recognition in robotic manipulation tasks [11]. Moreover, advances in computer graphics are shifting the dataset creation paradigm for training ML models towards synthetic datasets: Gaidon et al. [12] generated a photo-realistic dataset of virtual environments for transport. Dekhtiar et al. [13] used CAD datasets to train a deep neural network for mechanical component recognition in industrial environments. Dahmen et al. [14] proposed a syntetic dataset to perform defects detection in optical inspection systems. Alexopoulos et al. [3] proposed a framework for implementing DT models for training ML models. However, they also highlight the lack of concrete implementations of real use cases investigating the validity of DT-driven ML systems. The topic of generating synthetic datasets from DTs is not widely discussed in literature, especially in association with manufacturing systems.

3 System Design and Development

The aim of this work is to create a DT of a demonstrative industry 4.0 production line for assembling skateboards. The purpose of the DT is manifold, since it will be used to 1) visualize, navigate and inspect the production line through immersive Virtual Reality (VR), 2) train technicians to correctly interact and operate the different technologies and their interaction interfaces and 3) test and evaluate possible upgrades to the physical line in a VE prior to implementing them in the real world. The line process consists of four macro steps: 1) the technician starts the process selecting through a user interface the color for the wheels, the trucks, and the board; 2) then he/she picks the components from an automatic warehouse and positions them on the production pallet; 3) a robotic manipulator performs a pick and place task to assemble the skateboard; 4) the technician completes the skateboard assembly fastening all the screws in the last step.

3.1 The Physical Production Line

The physical production line consists of three main assets which are shown in Fig. 1: 1) a Modular Intralogistic Organizer (MIO) by Comau, which operates as an automated warehouse, 2) a Virtual Guidance Interactive Learning (Vir.GIL) system by Comau, a complex system which combines together different technologies to provide digital guidance to the user, and 3) a Racer 5 by Comau, a 6-axis articulated robotic manipulator with a 5 kg payload, designed to ensure both industrial efficiency while providing safe, barrier-free operations.

The MIO rotates its eight shelves providing, one at a time, all the parts required to assemble the skateboard. The Vir.GIL system guides the operator through the preparation of the production pallet: a laser pointer and vocal hints suggest where and how to correctly place the components. For the third step of the procedure, the pallet is manually moved by the operator from the Vir.GIL workbench to the Racer 5, which starts picking each component and assembling the skateboard. Since the position of the wheels is not fixed and the cobot needs to detect their precise locations to correctly pick them up, an image processing algorithm tries to detect their position searching a circular shape in a given frame of the pallet. Then, if the precision is higher than 80%, the system is confident on the wheel position and proceeds with the pick and place operations. Finally, the operator moves back the pallet to the Vir.GIL station to fasten the screws.

Fig. 1. The MIO (left), the Vir.GIL (center), and the Racer 5 (right).

3.2 The System Architecture

The proposed DT was developed and deployed on a workstation equipped with an Intel Core i7 CPU, an NVIDIA GeForce Quadro 4000 Graphic Card and 128 GB RAM with dual boot for either Windows 10 or Ubuntu 20.04 LTS. The DT was developed in Unity 3D on Windows: the SteamVR plugin enables the application to run on the HTC VIVE Pro (2018) immersive Virtual Reality headset, whereas the Barracuda plugin is used to integrate the neural network, exported in the Open Neural Network Exchange (ONNX) format, into Unity 3D. The ML system was deployed on Ubuntu using Darknet, an open-source neural network framework written in C and CUDA. The Darknet framework was deployed to use You Only Look Once (YOLO), a state-of-the-art, real-time object detection system, to recognize the pose of the wheels for the pick and place task. To this end, it was necessary to install the Nvidia Cuda Toolkit, the Cuda Deep Neural Network library and the OpenCV library. The synthetic dataset used to train the Neural Network was generated through Blender, an open-source 3D modeling software. Unity 3D is a state-of-the-art game engine, widely used due to its cross-platform capabilities and its compatibility with novel technologies such as the HTC VIVE. Since the purpose of the DT is the training of an operator, Unity 3D provides a simple and intuitive framework for creating a complex VR interactive experience.

3.3 Digital Twin Development

The DT was developed retrieving the 3D models for all the physical assets and objects of the production line. The wheels, board, trucks, and screws of the skateboard have been modelled in Blender. The main goal of the DT is to offer an interactive experience; thus, it was necessary to recreate all the interactions and actions available in the real environment. The DT enables the user to navigate the virtual space, to grab the skateboard components and the pallet, and to operate the Racer 5, the MIO and the Vir.GIL. The MIO and the Racer 5 cobot behaviors have been virtualized through a combination of scripting and custom animations created in Blender. The vocal hints provided by the Vir.GIL were translated into visual hints in the form of 2D pop up displaying the instructions and bouncing 3D arrows to highlight the points of interest on the workbench. A snapshot of the final DT is shown in Fig. 2.

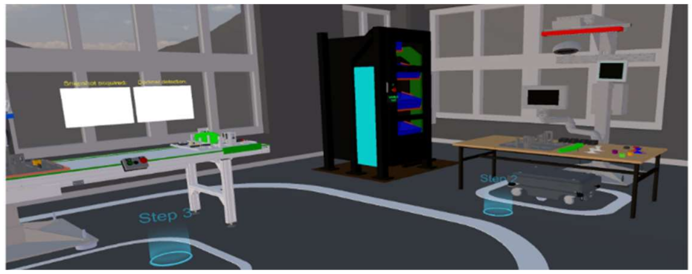

Fig. 2. A snapshot of the final digital twin production line.

3.4 Object Detection Through Real-Time CNN

The third step of the process involves the pick and place task performed by the Racer 5. The current vision system tries to detect the wheels position searching their shape in the framed image. However, the system cannot distinguish between face up and face down wheels, thus, when the pallet is placed under the cobot the wheels may be on the wrong side. To improve the pick and place task and make it robust to this type of errors, YOLO has been deployed to correctly detect the wheels. YOLO has been proven to be extremely fast and accurate [15] and it allows to easily tradeoff between speed and accuracy simply by changing the size of the model, without retraining the whole neural network. After deploying YOLO, the synthetic dataset used for this research was generated with Blender, using the wheel digital model created for the DT. A python script has been developed to automatically produce high-quality renders exploiting Blender's Eevee render engine. YOLO should be able to recognize 3 classes: wheel top, wheel back and wheel side. Thus, the script created 150 rendering for each class, for a total of 450 images with resolution 960 × 540 pixels (as depicted in Fig. 3). To make the detection system more reliable and flexible, variation have been introduced for each class in term of colors and camera perspective. For each render the script annotates (within a .txt file) the label for the class, the x-axis and y-axis coordinates, the width and height for each object.

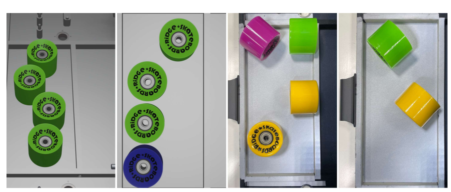

Fig. 3. Samples of the training dataset (left) and of the validation dataset (right).

To start the training phase, it is necessary to setup the configuration file, where it is possible to define the number of classes (3), epochs (6000), filters (24) and the size of the images taken as input (416 × 416 pixels). Those parameters have been fine-tuned through numerous tests on the specific dataset to achieve optimal performances. The input size is different from that of the synthetic dataset because the original renders are modified in the data augmentation step performed by the proposed detection system, which enlarges the training dataset by a large number of variations obtained through cropping, scaling and other visual artifacts and transformations applied to the original images. A validation dataset is defined to compute the mean Average Precision (mAP) of the neural network (two samples are depicted in Fig. 3). The training process is started through a shell prompt, and it takes about twelve hours to complete. Once the training phase was completed, the model was tested with both images and a video of the real production line. The YOLO system grabs each frame of the video, makes the prediction, and shows the result as a new, labelled video. To use the trained neural network in the DT Unity 3D application it was necessary to convert it to the Open Neural Network Exchange (ONNX) standard, an open format for ML models which allows to easily interchange models between various ML frameworks and tools. Through the ONNX, the DT can perform the object detection on a snapshot of the wheels taken with an invisible camera in the VE, at the same position and perspective of the real one. The result of the detection will be a tensor, a container that stores data in N-dimensions. The trained YOLO system generates an output of dimension 24 × 52 × 52 containing the results of the detection, given by the bounding boxes and the confidence of the detection. For each object detected in the input image, the output tensor stores the x-axis coordinate, the y-axis coordinate, width, and height. After filtering these results by selecting the ones with higher confidence, it is possible to draw the corresponding bounding boxes to the snapshot used as input, using different colors to distinguish the three classes (top, back, side). Fig. 4 shows an example of real-time detection.

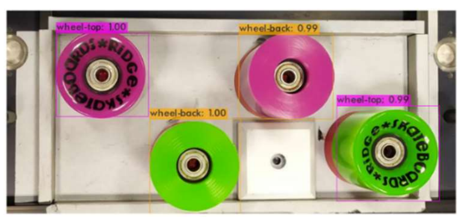

Fig. 4. An example of real-time detection on the real pallet.

Inside the VE, a GUI panel displayed near the cobot shows at each frame both the acquired snapshot and the detection result, whereas the user can restart the detection by a button (e.g., after correcting the pose of a misplaced wheel). The ONNX module is able to run YOLO and give back the results in about 2 seconds, which is a reasonable delay for testing the behavior of the cobot inside the DT.

4 System Evaluation

Due to the Covid-19 Pandemic, it was not possible to statistically evaluate the DT system by end user tests. However, the DT has been tried by both researchers and technicians from the CIM 4.0 competence center and has been assessed through interviews as a proper and complete digitalization of the physical production line.

Moreover, the YOLO detection system allowed to test an enhancement of the physical line, overcoming a limitation of the real model. In regards of the YOLO object detection system, the accuracy has been estimated by calculating the mAP during the training phase. The mAP compares the ground-truth bounding box to the detected box and returns a score. The higher the score, the more accurate the model is in its detection. For this purpose, a validation dataset of 55 images (labelled using YOLO mark) taken from the real world has been defined. Every 4 epochs of the training phase the AP is being calculated using images from validation dataset. Fig. 5 shows the loss function in blue and the mAP trend in red. The loss function is very high at the beginning, but it rapidly decreases as the model learns how to detect the wheels. The mAP increases during the epochs, and it reaches its maximum value nearby the 2400 epoch. The validation dataset and the corresponding trend of mAP are useful to select meaningful parameters for the detection system configuration. In this case, the mAP indicates that the best weights are around the 3000th step of the training phase. Testing the model in the real environment, with the weights obtained in the dataset validation step and the input from an industrial camera, proved that the system can achieve an accuracy of 85%.

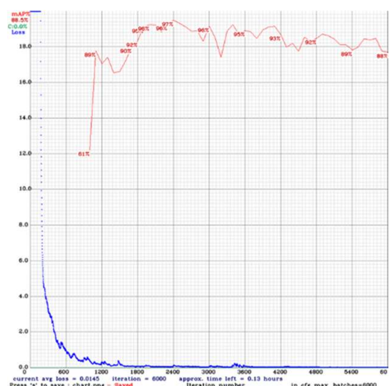

Fig. 5. Average Precision (red line) and loss function (blue line).

5 Conclusions and Future Works

The proposed research describes the development and deployment of a Digital Twin of a production line for assembling skateboards, which combines different recent technologies and image-processing algorithms, especially to assist a cobot in a pick and place task. Since this step of the procedure relies on a detection approach prone to errors, the DT was exploited to simulate and evaluate a possible upgrade for the production line. To overcome this limitation, the DT integrates YOLO, a state-of-the-art, object detection system based on a CNN trained on synthetic data, which provides real-time performances for real-time immersive VR simulations. Overall, all the technologies provided in the physical line have been successfully digitalized in the DT, resulting in a compelling and valuable tool that will be used to visualize, navigate and inspect the production line through immersive VR and to train technicians. The proposed upgrade to the image-processing system clearly enhances the system reliability and will lead to a physical update of the real-world system. Future works will be aimed at further testing and evaluating upgrades to the physical line in the VE prior to deploying them in the real world.

References

1. Tao, F., Zhang, H., Liu, A. and Nee, A.Y.: Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics, 15(4), pp. 2405-2415 (2018)
2. Janiesch, C., Zschech, P., & Heinrich, K.: Machine learning and deep learning. Electronic Markets, pp. 1-11 (2021)
3. Alexopoulos, K., Nikolakis, N., & Chryssolouris, G.: Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. International Journal of Computer Integrated Manufacturing, 33(5), pp. 429-439 (2020)
4. Tao, F., & Qi, Q.: New IT driven service-oriented smart manufacturing: framework and characteristics. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(1), 81-91 (2017)
5. Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H., & Vasilakos, A. V.: A manufacturing big data solution for active preventive maintenance. IEEE Transactions on Industrial Informatics, 13(4), pp. 2039-2047, (2017)
6. Nikolakis, N., Alexopoulos, K., Xanthakis, E., & Chryssolouris, G.: The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor. International Journal of Computer Integrated Manufacturing, 32(1), pp. 1-12, (2019)
7. Farhat, M. H., Chiementin, X., Chaari, F., Bolaers, F., & Haddar, M.: Digital twin-driven machine learning: ball bearings fault severity classification. Measurement Science and Technology, 32(4), (2021)
8. Srikonda, R., Rastogi, A., & Oestensen, H.: Increasing Facility Uptime Using Machine Learning and Physics-Based Hybrid Analytics in a Dynamic Digital Twin. In Offshore Technology Conference, (2020)
9. Ibrahim, M. S., Fan, J., Yung, W. K., Prisacaru, A., van Driel, W., Fan, X., & Zhang, G.: Machine Learning and Digital Twin Driven Diagnostics and Prognostics of Light-Emitting Diodes. Laser & Photonics Reviews, 14(12), (2020).
10. Rendall, R., Castillo, I., Lu, B., Colegrove, B., Broadway, M., Chiang, L. H., & Reis, M. S.: Image-based manufacturing analytics: Improving the accuracy of an industrial pellet classification system using deep neural networks. Chemometrics and Intelligent Laboratory Systems, 180, pp. 26-35, (2018)
11. Aivaliotis, P., Zampetis, A., Michalos, G., & Makris, S.: A machine learning approach for visual recognition of complex parts in robotic manipulation. Procedia Manufacturing, 11, pp. 423-430, (2017)
12. Gaidon, A., Wang, Q., Cabon, Y., & Vig, E.: Virtual worlds as proxy for multi-object tracking analysis. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4340-4349, (2016)
13. Dekhtiar, J., Durupt, A., Bricogne, M., Eynard, B., Rowson, H., & Kiritsis, D.: Deep learning for big data applications in CAD and PLM–Research review, opportunities and case study. Computers in Industry, 100, pp. 227-243, (2018)
14. Dahmen, T., Trampert, P., Boughorbel, F., Sprenger, J., Klusch, M., Fischer, K., Kübel, C. and Slusallek, P.: Digital reality: a model-based approach to supervised learning from synthetic data. AI Perspectives, 1(1), pp.1-12, (2019)
15. Redmon, J., and Farhadi, A.: YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018).