ESAIM 2024 – 2nd European Symposium on Artificial Intelligence in Manufacturing

A Tool for Generating and Labelling Domain Randomised Synthetic Images for Object Recognition in Manufacturing

Giovanna Martínez-Arellano[1,*] and Michael G. Buck[1]

[1]Institute for Advanced Manufacturing, University of Nottingham, Nottingham, UK.

[*] Corresponding author. Tel.: +44-115 748 6822;. E-mail: giovanna.martinezarellano@nottingham.ac.uk

Abstract. Reconfigurable manufacturing systems are becoming the only viable option to respond to changing product volumes and product specification, which are currently major challenges for the manufacturing industry. Part of this adaptation requires vision systems to be quickly updated to handle new unseen products. For deep learning-based vision systems, this means re-training on images that might not be available. Although there is some existing work on synthetic image generation in manufacturing contexts using a variety of domain randomisation techniques, there is a lack of understanding of which domains are critical in the effectiveness of the resulting trained model. There are currently no open tools to systematically conduct such ablation studies. This paper presents a tool based on Blender and CAD models to enable the study of domain randomisation in the generation of synthetic-only datasets that can yield accurate object recognition models. Preliminary results to validate the implemented domain randomisation techniques and the ability to generate the synthetic images are presented. Once generated, synthetic data sets are used to train a YOLOv8 model for object detection as a second tool validation step. Future work will look at performing ablation studies and expanding the range of domain randomisation methods to further study the capabilities of synthetic images.

Keywords: synthetic data · cad model · domain randomisation

1 Introduction

Reconfigurable manufacturing systems (RMS) are starting to get more attention as a viable option to improve responsiveness and resilience of current manufacturing systems [1]. With the current advancements of object detection and segmentation using Machine Learning (ML) [2], it is possible for these systems to flexibly perform different tasks such as pose estimation for object pick and place [3], quality inspection [4], among others. To train such ML models, a large number of labelled images need to be available, with enough variability (noise, background, rotation, obstruction) to achieve generalisation. With a changing manufacturing environment and new product specifications, it is challenging to have large amounts of real labelled images available. Although pre-trained models can be leveraged [5], some real images of the new object are needed. One way to address this challenge is to use synthetic images. Recent works have shown that domain-randomised, synthetic training images can yield object detection accuracy equivalent to real training images. However, there is a lack of understanding of which domains are critical for generating a fully synthetic data set that can yield such results. To understand this, more exhaustive and systematic ablation studies need to be performed across multiple manufacturing scenarios. There is currently no open implementation tool that can support such studies. Domain randomisation methods can be implemented in various ways and so open implementations need to be available to ensure comparability of ablation studies. With this context, this paper presents an open tool for the automatic generation of synthetic images and conduction of ablation studies. The tool is developed in a modular way to easily incorporate additional domain randomisation methods. To validate the implementation of the tool, tests of image generation and initial training of an object detection model using YOLOv8 have been performed. The rest of the paper is organised as follows. Section 2 provides a state-of the art on current developments in domain randomisation for synthetic images. Section 3 introduces the methodology proposed for implementing different domain randomisation techniques using Blender to manipulate the virtual scene. Preliminary results on the implementation validation of the tool are presented in Section 4 and Section 5 presents conclusions and future work.

2 Related Work

Automated pipelines for developing manufacturing-relevant synthetic images is an area of research that has got recent attention. An emerging way to generate synthetic data is through the use of generative models. Jain et al, for example, use Generative Adversarial Models to generate new images from existing real images of hot-rolled steel strips for surface defect detection [6]. Another way is through computer-aided design (CAD) models. Synthetic object data for nearly all manufactured parts is available in the form of a CAD file. Although this opens the door to automatic synthetic image generation, there is a reality gap; a model trained using synthetic images only will learn to recognise the synthetic object and not its real-life counterpart. One way to overcome this is to make the synthetic images as realistic as possible using object textures, colours, and scene lighting that matched the real object and scene [7]. Alexopoulos et al. present an automated pipeline for synthetic data generation using digital twins [8], introducing details of the real manufacturing environment to make the resulting images more realistic from a context/background point of view. These approaches assume some of the working environment is known, which may still be able to yield general deep learning models provided such models do not pick up on features of the environment itself. A contrary approach is using synthetic images that utilise the full range of visual variation that can be achieved synthetically, this is referred as domain randomisation. Conceptually, a model trained under such high synthetic variation will see the real world as just another one of these environments [9]. Dekhtiar et al. propose a methodology based on the use of CAD models and several domain variations such as random rotation, background, saturation, contrast, brightness and blurring [10]. The authors successfully leveraged pre-trained models to classify objects from the synthetic images. The authors implement some of the known domain randomisation techniques. These works, however, do not focus on the study of the randomisation techniques and their influence in object detection accuracy. Manettas et al. propose a synthetic image generation pipeline focusing on only top and bottom views of the object and focusing on varying the rotation achieving very good accuracy from only those view points [11]. The studies presented by Tobin et al, Hinsterstoisser et al. and Trembaly et al. [9, 12, 13], present a more in depth study on the influence of different domain randomisation techniques in object detection accuracy. Through a series of ablation experiments, all these three studies conclude that the resultant models outperform their real image-trained counterparts. In each methodology, one or more of the domain randomisation parameters are excluded/weighted differently for each training data set and results are compared. Hinterstoisser et al. found that blurring and light colour are the most influential factors in detection accuracy. By varying the weight of randomisation types, Toby et al. found that the object detection accuracy was reliant on all domains except for noise. Finally, Tremblay et al. excluded randomisation types one at a time and found lighting position and textures to have the greatest effect on object detection accuracy. Overall, the three studies combined do not agree on a clear answer to the importance of each domain randomisation type. It is consistent, however, that lighting randomisation has a substantial effect on accuracy relative to other domains.

3 Methodology

In this work, a methodology for generating and labelling synthetic images using CAD models is introduced. As shown in Figure 1, there are for image generation steps, followed by a ML model development step to enables the ablation study. The automation pipeline was developed in the 3D physics simulator Blender [14]. Within Blender, a CAD part and a virtual camera can be manipulated in a 3D environment to capture images (Step 1). Here, object, background, and lighting are fully customisable, making this software a good option for implementing modularised domain randomisation methods. This is important for allowing the user full customisation of each domain during ablation studies (source available in [15]).

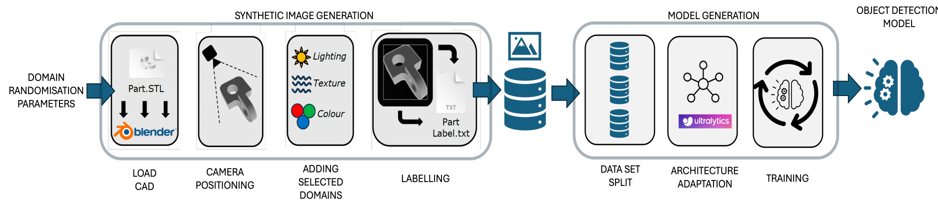

Fig. 1. Synthetic image generation pipeline used for the proposed tool using domain randomisation (DR) techniques.

3.1 Camera Positioning

For positioning the camera, the various angles from which to take the images must be decided. One potential approach is to use the 12 vertices of an icosahedron as the camera positions [12]. More vertices can be created by repeatedly splitting each face of the icosahedron to create additional vertices [16]. Although this method provides equal coverage of angles, it limits the number of angles. A more randomised approach is used by Tobin et al. [9]. Here, the position and orientation of the camera are randomised within the 3D space for each image. However, images are only captured from above the objects. In this study, a method similar to the one proposed by Dekhtiar et al. is implemented [10]. Here, the camera moves around the object on a primary axis. Each time this axis intercepts a rotation axis, the camera then follows the rotation axis, taking a set number of images of the part along the way. By allowing the user to define the number of rotation axes and image points, a unique angle can be used for every image regardless of the number of images. In addition, this approach, compared to having a fixed camera and a rotating object, avoid restraining the variation in the lighting between images.

3.2 Lighting

With the camera in position, the remaining domain randomisation features need to be set before the image is captured (Figure 1, phase 3). Lighting is the most influential domain randomisation type for object detection accuracy according to the literature. Tobin et al. varies three light domains: number of lights, colour and position [9]. The same study also restricts the lighting conditions to those that are offered by the lights within the software. Lighting in the real world is far more complex than that generated by spotlights. In this study, a more diverse lighting method was used: high dynamic range images (HDRIs). HDRIs are a type of 360-degree image containing complex lighting [17] that can be wrapped around the scene in Blender. When used as a background, they impart the full range of lighting conditions from that image onto the object. For the tool, a random HDRI background is automatically loaded into the scene for every image taken. Each HDRI imparts lighting of varied position, colour, and intensity on the object, having complex bright and shaded areas. Thus, all three lighting domain types are randomised using one HDRI loading function. With this, and all remaining domain randomisation types added (i.e. random colour, background, texture, position and distance), the image is captured and ready for labelling.

3.3 Labelling

Accompanying all images used for training object detection models must be an image label (Figure 1, phase 4). In this study, image label data was retrieved using the Python library OpenCV. To avoid background interference, a duplicate image is taken with the object in an identical position, but against a black background. This duplicate image is then processed using OpenCV to extract labelling information. After phase 2 (Figure 1), the position of the object in the camera frame is unchanged. Hence, the label extracted from the duplicate image can be directly used as the label for the final image.

4 Tool Validation and Verification

To validate and verify the implemented domain randomisation techniques and the tool as whole, a test to generate a set of synthetic images given a CAD model of a real part was performed followed by the development of a deep learning model for detecting the part in a real environment. The testing of the image generation process starts with inputting first the parameters for each of the domains to be randomised. This is done via a simple Python-built user interface, where domains to randomise in the images can be selected by entering either 1 (select) or 0 (deselect), and introducing the number of images to generate. Finally, a set of background images, HDRI lighting images, and part(s) to be used by the tool are uploaded to the working directory. Once this is set up, images are generated by the tool. The following parameters were selected for verifying the correctness of the domain methods implementation:

- The number of axes for positioning the camera was 45, with 45 images being taken on each axes (45*45 = 2025 images)
- Distance was randomly selected using a set of 4 different camera focal length values. These range from the focal length where the object fills the image, to one 40mm less (-10mm, - 20mm, -30mm, -40mm).
- The object roughness, how reflective it is, and colour were randomised.
- Backgrounds that depict different examples of tables available on this github project [18] were used and selected randomly for each generated image. It is worth noting that there is no intention to replicate the real environment in the synthetic data.
- The lighting is randomly selected from a set of 3 HDRI images: a studio lit room, an indoor lit house, and a lit town at night.
- The object position is randomised in both X and Y directions.

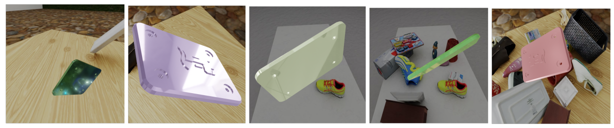

Fig. 2. Examples of images generated randomising background, lighting, position, distance and texture.

Once generated, the images were visually inspected to verify each domain was rendering the expected results according to their implementation. Figure 2 shows some examples of the generated images. As it can be observed, changes in rotation, texture, distance are present. Texture does not seems to be particularly noticeable, but reflective properties of the object can be observed.

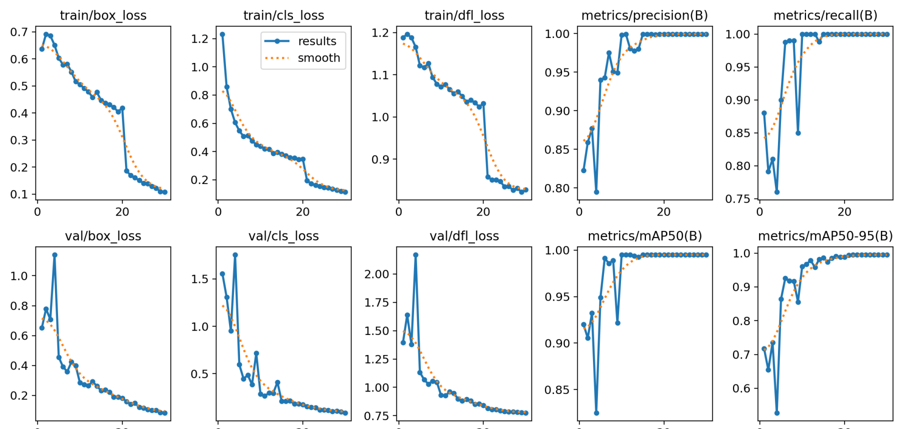

Fig. 3. Box, classification and distribution focal loss and precision during training on the training (top) and validation (bottom) sets. Validation shows mean average precision at different intersection over union thresholds.

After this, the images were used to train a YOLOv8 model (Ultralytics Python Library). This model was chosen for being a widely used and efficient model when working in real time [19]. The complete set of 2025 images were used for training and further 100 images were generated and used as a validation set. For testing, 20 real images taken within a robotics cell at University of Nottingham Robotics Lab where used, which contain white and black 3D printed parts corresponding to the CAD model. For using YOLOv8, the last layer was modified to introduce the class "Sensor Lid". The pre-trained model was then trained for 30 epochs as it was observed on the training/validation curves (Figure 3) that accuracy results in the validation set were already reaching the highest precision.

The resulting model was tested on the real images (some of them shown in Figure 4). The model was able to detect 47% of the white lids but failed to detect any of the black lids. Although the colour is highly varied in the training images and is not expected to play a factor in detection, it is evident that the model struggles with this particular colour. This may be related to the YOLOv8 model itself using the colour and contrast of the object. Also, some particular angles seem to be difficult to detect. It is worth noting that no particular pose strategy was used in this initial test, which according to Hintertoisser et al. can highly increase the accuracy of the model. Despite the low accuracy, it was possible to successfully validate the implementation which then will allow thorough ablation studies to be carried out. The results highlight why it is indeed important to understand the importance of domains and suggest that some domains are more useful to be randomised and others to be more strategically used.

Fig. 4. Examples of tested real images with their corresponding bounding boxes.

5 Conclusions and Future Work

Advancing the object detection accuracy that can be achieved by using domain randomisation is the next step in facilitating object detection, and enhanced flexibility in manufacturing. In this work, a tool for testing domain randomisation for the creation of synthetic images in manufacturing is presented. The lack of consistency between the domains randomised in recent studies, and the limited industrial testing of synthetically trained deep learning models demand for a novel range of domain randomisation types, combining all of those previously tested. Preliminary results have provided an initial validation of the implementation. Future work will look at performing different ablation studies as well as to implement distractor objects.

References

1. Brunoe, T. D., Soerensen, D. G. , Nielsen, K. Modular design method for reconfigurable manufacturing systems. 54th CIRP CMS 2021 - Towards Digitalized Manufacturing 4.0, Procedia CIRP, vol. 104,pp.1275–1279 (2021)
2. Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu,: Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212-3232 (2019)

3. C. Rennie, R. Shome, K. E. Bekris and A. F. De Souza, "A Dataset for Improved RGBD-Based Object Detection and Pose Estimation for Warehouse Pick-and-Place," in IEEE Robotics and Automation Letters, vol. 1, no. 2, pp. 1179-1185, (2016)
4. Zheng, X., Zheng, S., Kong, Y.,Chen, J.: Recent advances in surface defect inspection of industrial products using deep learning techniques. The International Journal of Advanced Manufacturing Technology, 113, pp. 35-58, (2021)
5. J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang :Transfer learning using computational intelligence: A survey, Knowledge-Based Systems, vol. 80, pp. 14-23, (2015)
6. Jain, S., Seth, G., Paruthi, A. et al. Synthetic data augmentation for surface defect detection and classification using deep learning. J Intell Manuf 33, 1007–1020 (2022)
7. Sampaio, I. G. B., Viterbo, J., Guerin, J.: Improving robustness of industrial object detection by automatic generation of synthetic images from CAD models, Computational Intelligence, Article vol. 39, no. 3, pp. 415-432, (2023)
8. 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), 429-439, (2020)
9. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P.: Domain randomization for transferring deep neural networks from simulation to the real world. In 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS) pp. 23-30, IEEE, (2017)
10. 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, 227-243, (2018)
11. Manettas, C., Nikolakis, N., Alexopoulos, K. Synthetic datasets for Deep Learning in computer-vision assisted tasks in manufacturing, Procedia CIRP, 103, 237-242, (2021)
12. Hinterstoisser, S., Pauly, O., Heibel, H., Martina, M. and Bokeloh, M.: An annotation saved is an annotation earned: Using fully synthetic training for object detection. In Proceedings of the IEEE/CVF international conference on computer vision workshops, pp. 0-0, (2019)
13. Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., Birchfield, S.: Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 969-977, (2018)
14. Blender: The Freedom to Create , https://www.blender.org/about/ . Last accessed 30/05/2024.
15. Synthetic Image Generation for YOLO, https://github.com/michaelgbuck/Synthetic-Image-Generation-for-YOLO/blob/main/README.md. Last accessed 30/05/2024.
16. Hinterstoisser, S., Lepetit, V., Wohlhart, P.,Konolige, K.: On pre-trained image features and synthetic images for deep learning. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops pp. 0-0, (2018)
17. Reinhard, E., Ward, G., Pattanaik, S., Debevec, P.: High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics). Morgan Kaufmann Publishers Inc., S.F., CA, USA, (2004)
18. Syn Table, https://github.com/ngzhili/SynTable/. Last accessed 30/05/2024.
19. Hussain, M., YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection, Machines, vol. 11, no. 7, p. 677, (2023)