Leveraging Generative AI for Synthetic Data Generation: Improving 6-DOF Pose Estimation in Assembly Systems

Christos Konstantinou, Nikos Kampouroglou, Nikos Theodoris, Fotis Basamakis, Christos Gkournelos, and Sotiris Makris(✉)

Laboratory for Manufacturing Systems and Automation, University of Patras, Rion Patras 26504, Greece
makris@lms.mech.upatras.gr

Abstract. In recent years, accurate 6-DOF (six degrees of freedom) pose estimation has emerged as a pivotal technology in manufacturing, enabling the precise localization and manipulation of objects in complex environments. The effectiveness of 6-DOF pose estimation algorithms critically depends on the availability of diverse, well-annotated datasets. However, obtaining and annotating such datasets present significant challenges due to their scarcity and the intensive labor required for accurate labeling. To address these issues, we propose an innovative approach that employs synthetic data generation, powered by generative artificial intelligence (AI) techniques specifically tailored for industrial applications. Our method enhances the synthetic data generation process by utilizing generative adversarial networks (GANs), which infuse the data with contextual details relevant to manufacturing environments. This process is further augmented by advanced rendering techniques and simulations that create realistic industrial scenes, complete with accurately annotated ground truth for 6-DOF poses. We validate the effectiveness and robustness of our proposed solution through its application in a real-world industrial use case, demonstrating its potential to substantially improve 6-DOF pose estimation in a manufacturing case, used for robotic picking of electronic parts.

Keywords: Synthetic Dataset Generation · Generative Adversarial Networks · Machine Learning · 6D Pose Estimation

1 Introduction

Manufacturing environments present unique challenges [4] to pose estimation algorithms, particularly in cluttered scenes characterized by disorganized backgrounds, occlusions between objects, and changes in lighting conditions. The development of robust 6-DOF pose estimation models comes with significant challenges, primarily due to the reliance on extensive, accurately annotated datasets [8]. Within manufacturing environments, products and industrial parts

K. Alexopoulos et al. (Eds.): ESAIM 2024, LNME, pp. 71–79, 2025.
https://doi.org/10.1007/978-3-031-86489-6_8

diverge from the commonplace objects typically found in large open datasets. Industrial objects often exhibit unique characteristics like complex geometries and uncommon material textures. Thus, they cannot be used seamlessly in manufacturing applications, creating the need for a systematic framework for data generation.

As highlighted in recent studies, expanding existing datasets with synthetic data, has proved to be a promising strategy to overcome the limitations raised by the absence of physical industrial data. Based on latest research in large models, Generative adversarial networks (GANs) can be a promising approach to overcome such constrains, providing a virtually limitless pool of annotated data by adding contextual details relevant to manufacturing environments. GANs can be used to create synthetic datasets that are tailored to replicate complex, real-world scenarios with remarkable accuracy, leading to pose estimation models that are both flexible and robust after appropriate training[11]. However, the transition from synthetic to real-world application presents its own set of challenges. The "reality gap" [12], a term denoting the discrepancy between model performance in synthetic versus real environments-remains a significant obstacle [2].

Recent literature underscores a range of approaches, each addressing unique challenges within the domain as it can be seen summarized in Table 1. In general, estimating 6D poses from RGB images presents a number of challenges. Perspective ambiguities, wherein objects exhibit similar appearances from varying viewpoints, hinder effective learning, especially in cluttered scenarios [2][3]. Moreover, environmental factors such as lighting variations and complex backgrounds further complicate the algorithmic performance.

Table 1. Comparison of Various 6D Pose Estimation Approaches

Reference	Real-Time	Uses RGB	Uses Depth	Cluttered	Refinement	Real DT	Synth. DT	Implementation	Testing Dataset
PoseCNN [14]		✓		✓		✓	✓	TensorFlow	YCB-V, LineMOD
DOPE [12]	✓	✓					✓	PyTorch	YCB-V
G2L-Net [3]	✓	✓	✓		✓			PyTorch	YCB-V, LineMOD
Megapose6D [6]		✓	✓	✓	✓		✓	PyTorch	BOP datasets , ModelNet
SAM6D [7]		✓	✓	✓	✓		✓	PyTorch	BOP datasets
FoundationPose [13]	✓	✓	✓	✓	✓		✓	PyTorch	BOP datasets , YCBInEOAT

Despite the innovative characteristics of all these mentioned approaches, notable gaps persist within the current landscape of 6D pose estimation

approaches. The integration of synthetic datasets, while beneficial for training due to many factors [1], such as time efficiency in data generation and collection, they introduce challenges related to domain adaptation and real-world generalization. Furthermore, achieving real-time performance without compromising accuracy and robustness, in terms of pose estimation, remains an ongoing challenge, particularly in cluttered manufacturing environments.

In summary, there exists a need for further research to address the inherent challenges and bridge the gap between synthetic training environments and real-world application scenarios. In context, this paper proposes a framework that can identify 6-DOF poses of novel objects, based solely on their Computer Aided Design (CAD) files, and in a textual description of their external visual characteristics. For further enhancing the detection precision, an automated way of generating the bounding boxes of these novel objects was implemented based upon a synthetic generated dataset. This research paper is organized as follows: In Sect. 2 the overall approach structure is defined, which is subsequently addressed in Sect. 3 where the implementation details are presented. The application of the proposed synthetic data generation for 6D pose estimation is evaluated in the Sect. 4 using a real industrial use case. Finally, in Sect. 5 the conclusions and future work are reported.

2 Approach

An approach combining GANs, CAD models and advanced simulation techniques has been developed in an attempt to construct synthetic datasets dedicated to 6-DOF pose estimation in manufacturing environments.This approach created detailed and varied training data that enhances the model's ability to accurately estimate poses of industrial-oriented objects.

As depicted in Fig. 1, the initial stage of the process involves the utilization of a pretrained GAN model [11] dedicated to the background image generation.

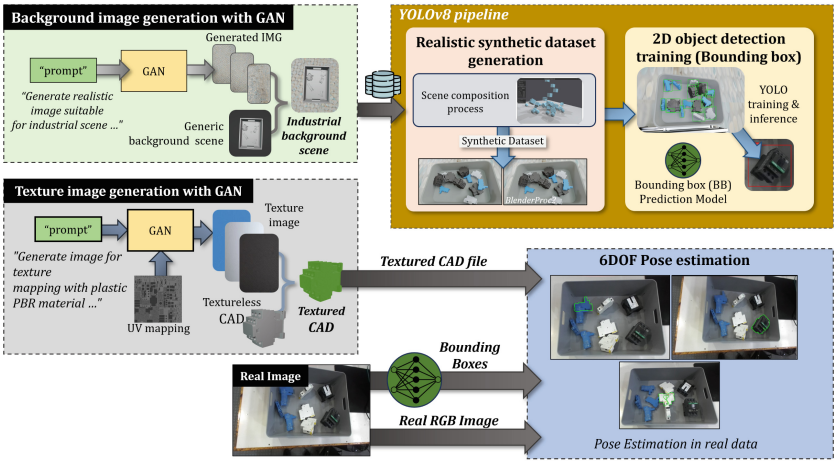

Fig. 1. Overview of the proposed framework

Since these type of models are trained on a vast amount of diverse environments, they can generate background images such as industrial floors, assembly lines and warehouses. This network uses both text and image prompts to accurately render the requested environment and its intricate patterns. The text prompt specified the desired outcome, such as "Produce high-fidelity background images of industrial rug featuring light-blue hues and multiple circular patterns."

Following the background generation process, another instance of a GAN was set up, aimed specifically for object texture generation. This model uses the object's exported UV mapping as a reference image, and a textual description outlining the desired texture characteristics. Both inputs fed into the network for it to generate an accurate texture that can wrap the object's geometry. This technique achieves a high degree of visual fidelity within the synthetic scenes. In this GAN, a text prompt guides the network to generate textures that mimic the desired material industrial characteristics. Using prompts such as "Generate high-quality image suitable for texture mapping focusing on plastic PBR material", it was possible to generate several accurate texture images, as can be seen in Fig. 1, leading to the synthetic dataset containing the objects of interest.

The integration of background images and detailed CAD models, in combination with advanced simulation techniques such as physics-based modeling, varying lighting conditions, and dynamic camera movements, generates synthetic scenes annotated with 6-DOF poses. These scenes exhibit a high degree of variety and realism, closely mirroring actual manufacturing environments. To automate this process and minimize human intervention, a YOLO based object detection system was developed and trained on the generated synthetic dataset. It successfully produced 2D bounding boxes for the physical components, which were subsequently used as fine-tuning inputs for the pose estimation model.

3 Implementation

In order to add photorealism to our data, Stable Diffusion XL model [9] was selected as the GAN architecture for the proposed work as illustrated in Fig. 2. Stable Diffusion XL model requires a text prompt as an input to render images based on the provided text. Diffusion models are designed to refine images by adding noise to them and then removing it, effectively diffusing the noise across the image space. The synthesis and generation of a synthetic dataset covering different scenarios, was achieved using BlenderProc2 [5]. This is a Blender pipeline capable of rendering realistic images after randomly placing the objects in a simulation environment.

All the physics-based calculations are applied to the simulation environment via a Python API. In this study, the synthetic data generation pipeline was executed on an Nvidia RTX 3060 GPU, in Ubuntu 22 environment, resulting in the creation of a dataset consisting of 50,000 synthetic images accompanied by the ground truth 6D poses of the objects.

The YOLO (v8) (You Only Look Once) object detection framework, as described in [10], was trained on the generated synthetic dataset with the 50,000

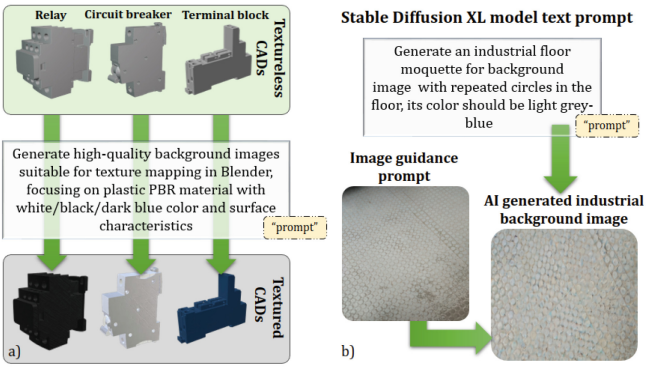

Fig. 2. a)Text-to-image texture generation using GANs b) Background image generation using GANs

images to provide the 2D bounding boxes of the industrial objects given an unseen image. The training procedure included 100 epochs on the synthetic dataset with a ratio of 70-15-15% for training, validation and test images respectively. The extracted bounding box was essential for the subsequent stages of the pose estimation process, thus enhancing the accuracy and automation of the system.

Regarding the 6D pose estimation, the MegaPose model [6], was utilized for estimating the translation and orientation of the objects. This model plays a benchmark role in order to demonstrate the reliability of the proposed method. As depicted in Fig. 1, the system receives as an input a real (not previously known) image of the region of interest, a CAD file with no specific texture of the part of interest, and two textual descriptions. One for the external visual characteristics of the industrial part, and the second for the surrounding environment. As it will be described in the following sections the evaluation of this method was performed both in the synthetic and the real domain.

4 Case Study

The presented work for the generation of an estimated 6D pose, has been deployed and tested into two use cases, that involve the detection and the handling of electrical parts placed randomly in a bin. These parts include terminal blocks, relays and circuit breakers as can be seen in Fig. 3. The proposed application for incorporating synthetic datasets for 6-DOF pose estimation is crucial for manufacturing, as it significantly augments the accuracy of object localization and manipulation in complex industrial environments. This field faces significant challenges due to the occlusions, varying lighting conditions, and diverse geometries of industrial parts, which complicate accurate pose estimation. The first case study involves comparing the impact of the GAN-textured CAD objects to texture-less CAD on the results of pose estimation algorithm.

Furthermore, the synthetic dataset of electrical parts was employed to train the YOLO model, enabling it to generate 2D bounding boxes for the detected parts. These bounding boxes were then used to compare the performance of the Megapose6D method on synthetic images, both with and without the fine-tuning provided by the bounding box input.

Following the proposed approach, a simulation was performed through Blenderproc resulting in a synthetic dataset comprising 50 scenes and each scene containing 100 frames. In order to evaluate the impact of GAN-generated textures on the produced synthetic data for 6-DOF pose estimation, metrics such as the angular difference of quaternions and Euclidean distance between translations are used to quantify the enhancement in the detected parts' poses estimation. Results were categorized based on whether the synthetic data included GAN-generated textures or not, and whether real data was used with or without GAN textures. The findings revealed a notable improvement in performance with the application of GAN textures. When the CAD models were overlayed by the GAN generated textures, a noticeable increase in the pose estimation accuracy was observed. On the other hand, the accuracy remained almost the same. Moreover, the impact of GAN textures was even more noticeable when analyzing real-world data. These results, as it can be observed at Table 2, demonstrate that GAN-generated textures enhance the accuracy and robustness of pose estimation models, especially in scenarios involving complex and varied textures.

Table 2. Comparison of pose estimation accuracy with and without GAN textures

	With GAN texture	Without GAN texture	Difference
Synthetic Data			
Blue terminal block	85.92%	85.90%	+0.02%
White circuit breaker	83.06%	78.12%	+4.94%
Black relay	81.12%	80.07%	+1.05%
Real Data			
Blue terminal block	92.43%	74.50%	+17.93%
White circuit breaker	95.55%	74.30%	+21.25%
Black relay	80.44%	80.43%	+0.01%

Similar pose estimation tests were carried out with real images of electrical parts. The results of the pose estimation followed the previous logic, with lower pose scores presented in occluded conditions or when the bounding box was not sufficiently precise to indicate the exact boundaries of the object. These scores were lower than those for synthetic objects, ranging from 65% to 80%. Realistic object textures from GAN were added to the synthetic data, along with an extracted bounding box from the YOLO trained model. The predicted bounding box of the electric part, as a result from the YOLO training on the dataset, achieved markedly higher pose scores, reaching 95% to 98%, even when

the object was not easily disguised from the environment, as illustrated in the Fig. 3.

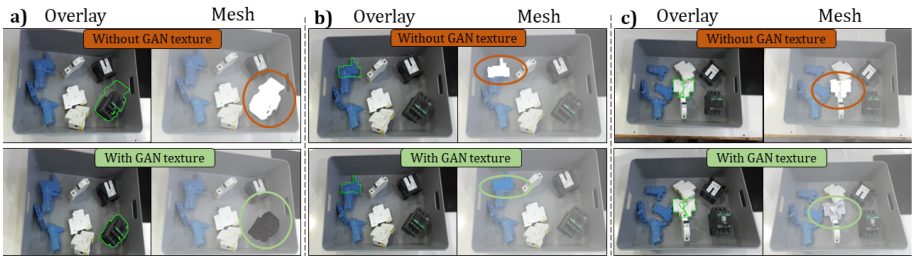

Fig. 3. Pose estimation results on real sample

5 Conclusion

In this paper, a synthetic dataset enhanced by GAN-generated textures is presented to improve pose estimation. The distinguish feature compared with novel 6-DOF pose estimation models, is the utilization of GAN generated content that augmented key characteristics of the approach, mimicing real industrial parts, and environments. These characteristics consist of the textures of the CAD models, and the industrial background spaces. For further enhancing the estimation results, a YOLO based object detection network was implemented, trained upon the GAN generated images and scenes. The proposed method was tested and evaluated in a real industrial use case, consisting of a set of 3 different components: a term block, a circuit breaker and a relay switch. Using the CAD files of the parts, and real images taken from a top-view camera, this approach was able to correctly identify the poses of all the components and even successfully handling the random occlusions. Comparisons with and without assist from the GAN image generation were performed, demonstrating significant improvements in the pose estimation accuracy, indicating an almost 18% increase. In conclusion, this framework highlights the effectiveness of GAN content generation, and proves its usability in industrial environments and complex use cases. However, despite achieving high evaluation scores, a number of challenges arouse. Future work aims to create an ecosystem of tools that finalize the application of an autonomous bin-picking, achieving an end-end interaction from the robotic environment. Finally, smart 3D scanner technologies, reconstructing the CAD models of industrial parts, can be utilized, further simplifying the overall flow.

Acknowledgments. This research has been supported by the EC funded projects "MASTERLY: Nimble Artificial Intelligence driven robotic solutions for efficient and self-determined handling and assembly operations" (Grant Agreement: 101091800) and "RENÉE: Flexible remanufacturing using AI and advanced Robotics for circular value chains in EU industry (Grant Agreement: 101138415)"

References

1. Abufadda, M., Mansour, K.: A survey of synthetic data generation for machine learning. In: 2021 22nd International Arab Conference on Information Technology (ACIT), pp. 1–7. IEEE (2021)
2. Cao, H., Dirnberger, L., Bernardini, D., Piazza, C., Caccamo, M.: 6IMPOSE: bridging the reality gap in 6d pose estimation for robotic grasping (Mar 2023). http://arxiv.org/abs/2208.14288, arXiv:2208.14288
3. Chen, W., Jia, X., Chang, H.J., Duan, J., Leonardis, A.: G2L-Net: Global to Local Network for Real-time 6D Pose Estimation with Embedding Vector Features (2020). https://arxiv.org/abs/2003.11089
4. Chryssolouris, G.: Manufacturing systems: theory and practice. Mechanical engineering series, Springer, New York, 2nd ed edn. (2006), oCLC: ocm61253973
5. Denninger, M., Winkelbauer, D., Sundermeyer, M., Strobl, K.H., Humt, M., Triebel, R.: BlenderProc2: A procedural pipeline for photorealisticrendering. J. Open Source Softw. **8**(82), 4901 (2023). https://doi.org/10.21105/joss.04901
6. Labbé, Y., et al.: MegaPose: 6D Pose Estimation of Novel Objects via Render & Compare (Dec 2022). http://arxiv.org/abs/2212.06870, arXiv:2212.06870 [cs]
7. Lin, J., Liu, L., Lu, D., Jia, K.: SAM-6D: Segment Anything Model Meets Zero-Shot 6D Object Pose Estimation (Mar 2024). http://arxiv.org/abs/2311.15707, arXiv:2311.15707 [cs]
8. Makris, S.: Cooperating Robots for Flexible Manufacturing. Springer International Publishing, Cham (2021)
9. Peebles, W., Xie, S.: Scalable Diffusion Models with Transformers (Mar 2023). arXiv:2212.09748 [cs]
10. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection (May 2016). arXiv:1506.02640 [cs]
11. Rojtberg, P., Pollabauer, T., Kuijper, A.: Style-transfer GANs for bridging the domain gap in synthetic pose estimator training. In: 2020 IEEE International Conference on AIVR, pp. 188–195. IEEE, Utrecht, Netherlands (Dec 2020). https://doi.org/10.1109/AIVR50618.2020.00039
12. Tremblay, J., To, T., Sundaralingam, B., Xiang, Y., Fox, D., Birchfield, S.: Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects (Sep 2018). arXiv:1809.10790 [cs]
13. Wen, B., Yang, W., Kautz, J., Birchfield, S.: FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects (2023). https://doi.org/10.48550/ARXIV.2312.08344, https://arxiv.org/abs/2312.08344
14. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes (May 2018). arXiv:1711.00199 [cs]