Digital Twin Technology for Robot Grasping in Complex Environments

Hongfanfan Wang

{toby0612whff@163. com}

China Changsha Experimental High School, 410024, Changsha, China

Abstract. Robotic manipulation tasks involving complex and cluttered circumstances remain a challenge for existing autonomous robots. The digital twin (DT) platforms open new avenues for improving robustness of grasp by real-time model learning and prediction-guided correction, and virtual-to-physical adaptation. This paper reviews the DT technology for robotic grasping. First, the fundamental concepts and architectures of DTs for robotic grasping tasks was provided. Then, a comprehensive review of the current research status was conducted, including the static objects in simulated environment, adaptive grasping under varying conditions, and algorithms. Finally, a summary of the current key issues was made based on the current situation, and the direction of future development was discussed. This paper provides a systematic overview of this evolving field and offers development insights for enhancing the autonomy of robots operating under uncertain conditions.

Keywords: Digital twin, Robotic Grasping, Real-Time Simulation

1 Introduction

Robotics grasping is one of the most important research fields over the last four decades in robotics and automated industry. In many high-precision automatic lines as well as logistics chains, Robotic manipulation is an important part of current automation, which can be performed reliably by robots in many scenarios. Early methods for robotic grasping were mostly based on heuristic rules or physics-based modeling [1]. All of these methods were aimed at finding the best grasping points when an object has a well-defined shape [1]. Theoretically sound, these classical approaches were limited by strong requirements on the environment to be highly structured, as well as full information about objects' parameters being known a priori. These conditions are rarely met in practice, leading to serious limitations for practical applications. Robotic manipulators developed early on tended towards task specific solutions which did not account for object position or orientation error, or shape differences. With the appearance of computer vision technologies and the creation of neural networks algorithms. The past five years have been an era of breakthrough for robotic grasping methods. With the help of powerful CNN-based visual perception, robots could directly predict full-grasps from raw Red, Green, Blue+Depth (RGB-D) images, teaching robots how to manipulate new objects [2]. Novel

ICMEEA 2026, April 24-26, Singapore, Singapore

DOI 10.4108/eai.24-4-2026.2364998

techniques like generative residual nets demonstrated significant improvements for manipulating unknown objects with large scale training data. However, they also revealed some drawbacks. The methods were dependent on huge amount of labeled data for training and models trained on simulations often performed poorly when applied to reality due to the well-known sim-to-real gap [3]. Furthermore, rapidly changing situations where targets move, lighting changes, or there is a complicated background exposed the weaknesses of purely visual feedback loops [4].

In recent years, there has been growing interest in combining DT technology with robotic grasping systems. A DT is a virtual replica that mirrors a physical entity in real-time, enabling bidirectional data flow, simulation, and predictive optimization [5]. Unlike static simulations, DTs maintain synchronization with their physical counterparts through continuous sensor feedback, creating a closed-loop interaction between the virtual and real worlds as a result for its strong functions. Firstly, DTs are being harnessed to enhance perception and decision-making under uncertainty. Wang et al. provided a comprehensive survey of DT-empowered robotic arm manipulation, covering a gamut of tasks such as catching, pick-and-place operations, and assembly. They also explored reinforcement learning-based path planning methods implemented in simulators like Unity, MuJoCo, and PyBullet [6]. Secondly, DTs play a crucial role in facilitating sim-to-real transfer by minimizing the visual and physical discrepancies between the virtual and real domains. Matl proposed a DT-enabled approach for effectively transferring deep reinforcement learning algorithms to physical robots. This method uses virtual outputs to correct real grasping points and has shown improved success rates even under lighting variations and occlusion [3]. Thirdly, synthetic data generation within DT frameworks has emerged as a powerful paradigm. Jia et al. proposed a comprehensive framework for DT-based robotic grasping that encompasses synthetic data generation with domain randomization, hybrid training strategies combining synthetic and real data, and the creation of dynamic working circumstances for motion generation and collision avoidance [4]. Beyond these algorithmic advancements, researchers have delved into specialized domains for DT applications. Singh et al. developed a deep-learning-based approach for strawberry grasping in greenhouse DT environments. By leveraging YOLOv9-GLEAN with super-resolution capabilities, they achieved a precision of 0. 996 and a recall of 0. 991 [7]. Liu et al. proposed a DT framework for soft robotic grippers, integrating computer vision for driving pressure control and Unity 3D for kinematics simulation based on four-piece wise constant curvature modeling [8]. Ren et al. introduced a DT robotic system with a continuous learning strategy for 7-degrees of freedom grasp detection, enabling robots to acquire new grasping skills without catastrophic forgetting as novel industrial parts enter production lines [9]. These studies showcase the expanding scope of DT applications from rigid industrial manipulators to soft robotics and agricultural automation.

The purpose of this survey paper is to provide a comprehensive overview, comparison and analysis of the growing body of work on using DTs for robotic grasp synthesis, in particular, under poor conditions. Despite many studies showing that DT models can be used to increase grasp robustness, a comprehensive overview on how traditional analytic methods evolved into current dynamic twin approaches is still missing. The question is then, what about the DT works well at capturing key challenges such as sensor noise, sim-to-real gap and environment dynamics to guide future researches and practical applications. The remainder of the paper is structured as follows: Section 2 briefly describes what are DTs and introduces the main approaches that can be applied to robots. Sections 3 reviews current works on DT-supported grasping, divided in three different regions, namely robust grasp execution under supervision. Visual attention in multiple scenes, Adaptive grasping of objects in dynamic environments, and the tools that support them computationally. Section 5 discusses limitations of existing approaches, highlighting issues of scalability, efficiency, bottlenecks and possible remedies for them. Conclusion: It presented some interesting results regarding the utilization of DT to achieve robot autonomy in harsh circumstances.

2 Definition and fundamental approaches

DT is the process of generating and maintaining a digital replica of an actual entity in real time for its full life cycle. The existing methods related to DT are mainly aimed at constructing "replicas", while more advanced approaches such as parallel systems emphasize the smart management and control of complex systems via virtual-reality interaction. The DT is implemented with an iterative cycle consisting of three stages: real-time matching, bidirectional data flow, and holistic modeling [5]. In contrast to regular computer aided design (CAD) models or isolated simulations, DTs maintain a real-time link to the actual counterpart by streaming in sensor measurements . Such communication enables prediction modeling, scenario testing, and adaptive closed-loop control strategies [10]. In the context of robotic grasping applications, DT systems typically consist of four key components.

Firstly, a geometric model captures the physical dimensions and kinematic structure of the robot and its surrounding environment, often sourced from unified robot description format (URDF) files or CAD models [3]. Secondly, a behavioral model represents dynamic responses, encompassing joint limits, actuator characteristics, and contact interactions. Thirdly, a sensor model emulates perception inputs like RGB-D cameras, allowing for realistic image synthesis with controlled noise features [4]. Fourthly, a communication layer facilitates the two-way exchange of data between virtual and physical systems, commonly using protocols such as Robot Operating System (ROS), Message Queuing Telemetry Transport (MQTT), or Open Platform Communications Unified Architecture (OPCUA) [6]. Several fundamental techniques underlie DT-enabled grasping systems. Domain randomization modifies simulation parameters such as lighting, textures, object poses, and camera viewpoints. This generates diverse training data, enhancing the robustness of policies when implemented in real-world scenarios [3, 4]. Physics simulation engines, including MuJoCo, PyBullet, and Unity Physics, provide the computational basis for modelling rigid-body dynamics, contact forces, and gripper-object interactions [6]. Synthetic data generation capitalizes on these simulations to create large-scale labeled datasets for training perception models, alleviating the data scarcity issue that plagues purely real-world learning methods [4, 7]. There is an important difference between the concept of digital model, digital shadow and DT [10]. The first does not need any connection or synchronization with the real object automatically. The second has only one direction of information flow which from the physical world into the digital shadow. On the other hand, true DTs are two-way and allow for feedback from the simulation back to the actual system in the form of commands or interventions.

3 Current research of grasping based on DTs

3. 1 Static grasping in structured environments

The static grasp is defined as an object with stationary pose and environment during the whole grasping operation, where the role of the DT is mainly to support grasp planning with simulated evaluation, as well as creating synthetic data. Fig. 1 displays the architecture implementation for welding [11]. Also, the DTs are used for industrial robots [12]. In Matl, an approach of DT-aided grasp learning by deep reinforcement was presented in assembling scenarios and the DT mechanism adjusts real grasping locations with simulated feedbacks [3]. The twin model communicates with real robots bidirectionally, receives image information of a camera online, and outputs grasp point coordinates. The experiments demonstrated that the success rate of the proposed DT approach is higher than a classical vision-based approach, especially in harsh illumination conditions that are problematic to traditional approaches. There are many constraints for human observations. Jia et al introduced a novel DT framework combining synthetic data and domain randomization for training deep networks, which can generate massive labeled grasping data by varying numerous factors in simulation, object pose, lighting conditions and texture [4]. Blended Learning provides a combination of both synthetic data with real-world data that reduces the gap between simulation and reality to improve algorithms' performance for real world perception tasks. This approach can be applied to various robot platforms as shown by the framework such as motion planning, path planning, obstacle avoidance for moving obstacles. Choi developed an environment for sim-to-real transfer of deep reinforcement learning (DRL) techniques to manipulate deformable bodies which includes both simulation-based synthetic data generation and DRL policies [10]. They address challenges posed due to soft body manipulations through generating diverse samples during training phase via simulations, enabling adaptive control policies depending on the situation. In fact, the variations in object shape and stiffness are extensive, and here which can see how DTs help overcome the problems which lack of data scarcity.

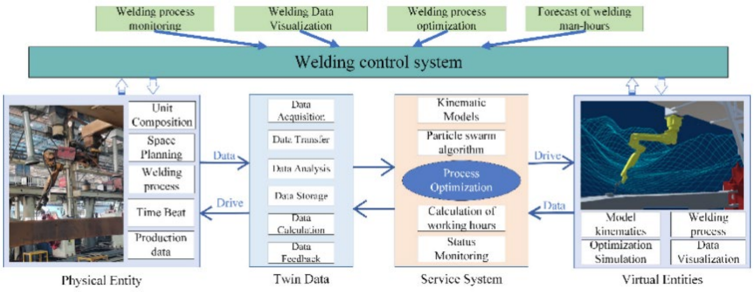

Fig. 1. The deployment of DT modeling system in architecture for welding [11].

3. 2 Dynamic grasping under environmental changes

Dynamic grasps involve moving objects, changing environments or adaptive manipulation tasks that evolve over time [6-8]. The use of a DT can significantly improve robot assembly performance, especially in complicated surgeries. To achieve the best possible digital reproduction, a good understanding of the system is required as well as real-time display, which poses significant implementation challenges. Fig. 2 illustrates how it can use a multi-source model driven DT methodology to achieve geometric, physical, and procedural rule representation [13]. Singh developed virtual machine model to simulate strawberry picking in greenhouses with soft, partially occluded objects under natural lighting conditions [7]. The following azebo-based virtual environment realistically replicates the strawberry greenhouse located in Abu Dhabi, for testing algorithms and avoiding a physical prototype. YOLOv9 trained on both the real world strawberry data sets and synthetic strawberry images achieves high accuracy and recall, which can avoid interrupting the current harvest operation. The motion controller is implemented using the system implementing visual servoing so that it can generate precise trajectories based on image information in a timely manner. DTs operations continuously improved all year long, avoiding the restrictions due to seasonality, environment, which are usually imposed in a physical test. To address the particularities of controlling a soft robotic gripper in order to obtain a dynamical grasp, Liu et al. presented a computer-vision-based DT which integrates both image processing techniques and kinematic models of the system [8]. These soft grippers have many benefits when grasping fragile objects, but their nonlinear deformability, which poses serious control problems. This novel approach of DT employs computer vision, which maps force exerted by finger in terms of relevant features related to the gripper such as radius or angle formed. This data is further fed back into the virtual representation developed through Unity 3d platform employing four stages constant curvature kinematics. They have shown that it is possible to achieve online reconfiguration where a deformable gripper changes its shape according to new goal location while grasping objects. In an excellent survey paper on how DTs can be used for controlling manipulators performing different tasks such as catching component placement, and component assembly.

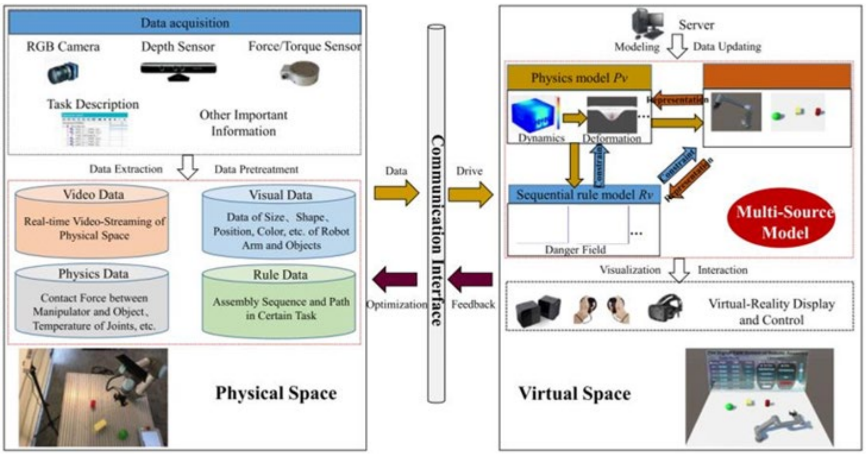

Fig. 2. Multi-source model-driven DT [13].

Reinforcement learning methods, for instances, Deep Q-Networks and Proximal Policy Optimization, are increasingly being implemented within DT frameworks. Simulation environments such as Unity, MuJoCo, and PyBullet provide safe training grounds before real-world deployment. The survey emphasizes that DT systems not only enable policy learning but also continuous performance monitoring and adaptation as environmental conditions change**.**

3. 3 Algorithmic breakthroughs

The progress on static and dynamic grasp planning largely depends on computational methods that leverage virtual twin functionalities. Summary of some representative grasping approaches with virtual twin functionality, which can be collected from the literature in Table 1.

Table 1 Comparison of representative DT-enabled grasping approaches

Research	Application domain	DT platform	Key algorithm	Reported performance	sim-to-real method
Matl [3]	Industrial assembly	V-REP	DQN	Improved success	Virtual output correction
Jia [4]	General manipulation	Custom framework	Hybrid training	Reduced sim-to-real gap	Synthetic data generation
Singh [7]	Agricultural harvesting	ROS-Gazebo	YOLOv9-GLEAN	Improved success	Synthetic training
Liu [8]	Soft robotic grasping	Unity 3D	Curvature kinematics	Satisfactory real-time control	Vision-based parameter mapping
Ren [9]	Industrial manipulation	Custom framework	Continuous learning	Adaptive skill acquisition	Incremental learning
Choi [10]	Deformable object grasping	Custom framework	RL	Generalization deformations	Domain randomization

Reinforcement learning remains a dominant paradigm, with deep Q-networks (DQN) and policy gradient methods being adapted for DT environments. Matl shows how DQN-based policies trained in simulation can be effectively transferred to physical robots via DT correction mechanisms [3]. This approach addresses the sample-efficiency challenge by using simulation for initial training and the twin to account for real-world gaps during deployment. Synthetic data generation has emerged as a vital enabler for perception-based grasping. Jia systematically varies simulation parameters to generate diverse training datasets, reducing the gap between synthetic and real images [4]. Domain randomization parameters include lighting intensity and direction, camera viewpoint perturbations, background textures, and object pose variations. The hybrid training strategy, combining synthetic data with limited real examples, achieves performance close to that of models trained on purely real data while requiring significantly less manual annotation. Continuous learning represents an emerging trend for DT-enabled grasping.

Ren proposed a DT robotic system that can continuously learn to grasp new objects as they appear on production lines [9]. The system retains a memory of previously learned skills and uses the DT to generate training data for new object categories, enabling adaptation without catastrophic forgetting. This approach is especially valuable in dynamic manufacturing environments where product variants change frequently. Simulation platforms themselves are also evolving. Wang compares popular simulators and highlights the different trade-offs between physical accuracy, rendering quality, and computational efficiency [6].

With the advancement in complex physical modeling, more advanced tasks are being performed using DT-based manipulation [8]. Astress prediction framework was proposed by Yang to achieve a robust grasping task in underwater environments, which combines the power of machine learning (ML) and physics-based simulations for overcoming challenges posed by harsh environments [9, 14]. The model predicts the expected pressure profile while manipulating and can take corrective action in advance by reducing grip force before dangerous conditions occur that may lead to slippage or object drop.

4 Constraints and future prospects

4. 1 Generalization presents a fundamental difficulty

DTs tend to be static and purpose-built, target object or other environment features. In fact, improving realism in simulation tasks as well as interoperability between different robot arm platforms is a promising area for further research work [6]. Grasping performance degrades significantly on unseen objects of different shapes and materials, on new tasks or in different settings from what they were trained on. Large scale, pretrained models showed promise by exposing themselves to a large amount of diverse data but their integration into the context of DTs is still in its early stages [15].

4. 2 Sub-optimal data utilization efficiency

Although synthetic data generation addresses annotation scarcity, current approaches often need large volumes of simulation data without ensuring coverage or relevance. The hybrid training strategy reduces but does not eliminate the need for real-world examples [4]. Active learning methods that can intelligently choose which synthetic scenes to generate or which real examples to collect could enhance efficiency. Additionally, leveraging unlabeled real data through self-supervised learning within DT loops remains largely unexplored.

4. 3 Inaccuracies in physical constraint modelling

While DT is intended to simulate physics in the virtual world, some errors of simulation vs reality cause failed grasping. The simulation of interaction with soft objects, friction variations, and non-ideal actuators response are major modelling issues. Liu et al. for learning compliant grasp patterns uses piece wise-approximated curvatures for modeling deformable objects, although still further detailed modelling is required [8]. A stress model designed for underwater grabbing tasks was proposed by Yang et al. and shows progress on specific physics-based simulations, but even broader deployment requires further optimization [14].

4. 4 The persistent sim-to-real gap

Even with the development of DTs techniques, there are still differences between a simulated model and its real counterpart. The adaptive error compensation method can improve robustness, which assumes always correct forecasts from its twin [3]. Degradations arise whenever perturbations in the environment exceed randomized training environments. This gap needs to be closed with a stronger integration of simulation models and the physical system itself in order to obtain a true DT instead of a DT shadow [10].

Transfer learning techniques that can adapt policies trained in simulation to real-world conditions with minimal fine-tuning hold great promise. Recent research on vision-language models and large pre-trained representations could enable zero-shot or few-shot adaptation across various grasping scenarios [15]. Moreover, closer coupling between DTs and physical systems through edge computing and low-latency communication will enable real-time optimization loops that continuously refine grasp strategies based on live sensor feedback.

4. 5 Standardization and accessibility

In some cases, high-fidelity DTs require large computational resources which may be prohibitive to small companies. Standardization regarding the format of information and communication between systems within a DT could help reduce costs and facilitate adoption [4]. As more autonomous operation of a DT is need, key issues are raised around oversight by humans and responsibility if things go wrong, and the possible effects on employment, among other things, that must be considered [10].

5 Conclusion

In conclusion, this paper provide a comprehensive review on the use of DTs for robot grasping in complex scenarios, combining results from different papers together, and DT addresses fundamental challenges for robotic manipulation through enabling synthetic data generation, bridging the gap between simulation and reality, providing continuous observation of a system and support for anticipatory improvement of systems. DTs address fundamental challenges in robotic manipulation by enabling synthetic data generation, sim-to-real transfer, real-time monitoring, and predictive optimization. Current research encompasses static grasping in structured settings, where DTs enhance planning through simulation-based evaluation and synthetic data generation; dynamic grasping in changing environments, where real-time adaptation is enabled through continuous sensor integration and closed-loop control; and algorithmic innovations spanning reinforcement learning, domain randomization, hybrid training strategies, and continuous learning. Current implementations of DTs in challenging fields such as agriculture, soft robotics, underwater manipulation, and dynamic manufacturing circumstances, which demonstrating improved grasping success rates, enhanced robustness to lighting and occlusion, and successful deployment in challenging domains such as agriculture, soft robotics, underwater manipulation, and dynamic manufacturing environments. However, limitations persist in terms of generalization across object categories, data utilization efficiency, physical constraint modelling accuracy, and the persistent sim-to-real gap. Future research directions include transfer learning, foundation model integration, closer virtual-physical coupling through true bidirectional twinning, continuous learning frameworks, and standardization efforts to lower adoption barriers. As DT technology continues to mature, its integration with robotic grasping systems is set to accelerate the deployment of autonomous manipulation in manufacturing, logistics, agriculture, and other fields, contributing to the broader vision of intelligent, adaptive automation in complex real-world circumstances.

References

[1] Alfaro-Viquez, D. et al.: A Comprehensive Review of AI-Based Digital Twin Applications in Manufacturing: Integration across Operator, Product, and Process Dimensions. Electronics, Vol. 14, pp. 646 (2025)

[2] Hu, W. et al.: A New Differentiable Architecture Search Method for Optimizing Convolutional Neural Networks in the Digital Twin of Intelligent Robotic Grasping. Journal of Intelligent Manufacturing, Vol. 34, pp. 2943–2961 (2023)

[3] Liu Y. et al.: A Digital Twin-Based Sim-to-Real Transfer for Deep Reinforcement Learning-Enabled Industrial Robot Grasping. Robotics and Computer-Integrated Manufacturing, Vol. 78, pp. 102365 (2022)

[4] Dong, J. et al.: A Digital Twin Framework for Robot Grasp and Motion Generation. 2024 IEEE Smart World Congress (SWC), Nadi, Fiji, 2024, pp. 2045-2052

[5] Iqbal, Z. et al.: Performance Characterization of the LEAP Hand: Control Interface and Digital Twin Integration. 2025 IEEE 30th International Conference on Emerging Technologies and Factory Automation (ETFA), Porto, Portugal, 2025, pp. 1-8

[6] Wang, Y. et al.: Digital Twin-Empowered Robotic Arm Manipulation with Reinforcement Learning: A Comprehensive Survey. Robotics and Computer-Integrated Manufacturing, Vol. 98, pp.103151 (2025)

[7] Singh, R., Seneviratne, L., Hussain, I.: A Deep Learning-Based Approach to Strawberry Grasping Using a Telescopic-Link Differential Drive Mobile Robot in ROS-Gazebo for Greenhouse Digital Twin Environments. IEEE Access, Vol. 13, pp. 361–381 (2025)

[8] Xiang, T. et al.: A Novel Approach to Grasping Control of Soft Robotic Grippers Based on Digital Twin. 2024 29th International Conference on Automation and Computing (ICAC), Sunderland, UK, 2024, pp. 1-6

[9] Ren, L. et al.: Digital Twin Robotic System with Continuous Learning for Grasp Detection in Variable Scenes. IEEE Transactions on Industrial Electronics, Vol. 71, pp. 7650–7660 (2024)

[10] Choi, M. et al.: Digital Twin System Framework for Gripping Deformable Objects with Synthetic Data and Reinforcement Learning. Flexible Automation and Intelligent Manufacturing: Manufacturing Innovation and Preparedness for the Changing World Order, pp. 98–106 (2024)

[11] Zhang, Q. et al.: Process Simulation and Optimization of Arc Welding Robot Workstation Based on Digital Twin. Machines, Vol. 11, pp. 53 (2023)

[12] Liu, X. et al.: Digital-Twin-Based Real-Time Optimization for a Fractional Order Controller for Industrial Robots. Fractal and Fractional, Vol. 7, pp. 167 (2023)

[13] Li, X. et al.: Multisource Model-Driven Digital Twin System of Robotic Assembly. IEEE Systems Journal, Vol. 15, pp. 114–123 (2021)

[14] Yang, X. et al.: Digital Twin-Based Stress Prediction for Autonomous Grasping of Underwater Robots with Reinforcement Learning. Expert Systems with Applications, Vol. 267, pp. 126164 (2025)

[15] Zheng, F. et al.: Digital Twin of Dynamics for Parallel Kinematic Machine with Distributed Force/Position Interaction. Journal of Manufacturing Systems, Vol. 80, pp. 70–88 (2025)