DOI: 10.54254/2755-2721/2026.AS31463

Artificial Intelligence in Virtual Reality Surgical Simulation: From Scene Generation to Skill Assessment

Zixuan Yin

Yancheng No.1 Middle School, Yancheng, China
hellokitty961@qq.com

Abstract. Traditional surgical training is confronted with inherent bottlenecks including inadequate cadaver supply, exorbitant costs, and subjective performance evaluation. Although virtual reality (VR) technology enables the provision of an immersive training milieu, its fidelity, realism, and intelligent capabilities remain constrained. The integration of artificial intelligence (AI) is profoundly revolutionizing multiple dimensions of VR-based surgical simulation. This study systematically summarizes the core technologies and practical applications of AI in VR-assisted surgical training and simulation. We propose a classification framework based on technical modules, delineating the roles of AI into three core dimensions: (1) AI-driven virtual surgical scene generation; (2) AI-enhanced physics engine to achieve real-time and realistic tissue deformation and interaction simulation; and(3) AI-enabled skill assessment and feedback. This study further elaborates on the predominant challenges faced by current technologies, such as physical-visual inconsistency, small sample data issues, and insufficient clinical validation, while outlining prospective research orientations such as multimodal sensing, federated learning, and AI-empowered robotic surgery skill transfer.

Keywords: virtual reality, artificial intelligence, surgical training, surgical simulation, generative AI, skill assessment

1. Introduction

Surgical intervention is one of the core modalities of medical practice. Conventional surgical training paradigms primarily rely on animal experiments, cadaver dissections, and surgical observations. However, these approaches are confronted with non-negligible limitations: the shortage of cadaveric and animal specimens, prohibitive costs, ethical controversies, and a failure to deliver standardized and reproducible training scenarios [1].

In recent years, virtual reality (VR) technology has emerged as a novel approach to address the aforementioned challenges in surgical training. By constructing a highly immersive three-dimensional virtual environment, VR simulators allow trainees to conduct risk-free repetitive practice, facilitating proficiency in surgical workflows and anatomical structures. However, early-generation VR simulators exhibited prominent limitations: their virtual scenes were predominantly pre-generated fixed content, lacking dynamic variability and diversity; the simulation accuracy and real-time performance of the physics engine struggled to balance simulation accuracy and real-time

DOI: 10.54254/2755-2721/2026.AS31463

performance, leading to inadequate fidelity in effects such as tissue deformation and hemorrhage [2]; and the rapid advancement of artificial intelligence (AI), especially deep learning (Deep Learning) and generative artificial intelligence (Generative AI), has furnished robust technical tools to overcome the aforementioned bottlenecks [3,4].

This study endeavors to present a systematic review of the application of AI in VR surgical simulation. Unlike previous reviews that focused exclusively on a single technical aspect (such as discussing only rendering or only assessment), this article innovatively proposes a taxonomy framework from the perspective of a comprehensive training workflow, systematically classifying the roles of AI into three core modules: scene construction, physical simulation, and surgical skill assessment. We will elaborate on the core technologies, representative studies, and prevailing challenges associated with each module, and outline prospects for future research directions, with the hope of providing a clear research roadmap for researchers and clinicians in this interdisciplinary domain.

2. Technical module classification: the role of AI in VR surgical simulation

Based on the distinct technical challenges addressed by AI in VR surgical simulation workflows, this study classifies AI's functionalities into three core modules: scene generation, physics engine acceleration, and skill evaluation. The classification framework is illustrated in Figure 1.

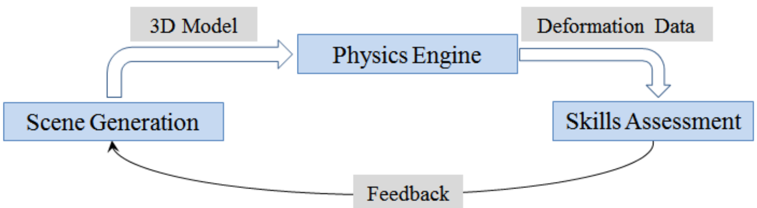

Figure 1. Three key modules of AI technology involved in surgical simulation

2.1. Module 1: AI-driven generation of virtual surgical scenarios

A high-fidelity, diversified virtual surgical milieu constitutes the cornerstone of effective surgical training. Traditional manual 3D modeling approaches are characterized by high time and labor costs, posing challenges to the rapid generation of large-scale anatomical models, especially for rare cases. AI technology, notably generative models, is revolutionizing this paradigm [3].

2.1.1. Generative 3D anatomical modeling

• Technical methods: Techniques based on Generative Adversarial Networks (GANs) and Diffusion Models enable the automatic segmentation and reconstruction of 3D anatomical structures from 2D medical imaging data (such as CT and MRI). For example, advanced segmentation networks like nnUNet can identify and extract the contours of organs (Dice $\geq$ 0.95), blood vessels (Dice $\geq$ 0.94), and diseased tissues (Dice $\geq$ 0.82) with high precision [3]. Furthermore, by utilizing Latent Diffusion Models, pathological models with different shapes, sizes, and textures (such as tumors at different stages) can be generated, thereby providing trainees with a rich, personalized repertoire of training cases.

• Case: A study published in MICCAI in 2024 successfully employed a spatially variant deconvolution GAN to realize high-quality real-time 3D ophthalmic ultrasound image enhancement,

DOI: 10.54254/2755-2721/2026.AS31463

with processing speeds far exceeding conventional methods, providing robust diagnostic support for clinical screening of anterior segment ophthalmic diseases [5].

2.1.2. Dynamic scenario and pathological simulation

Beyond static model generation, AI enables the simulation of pathological dynamic progression processes. By integrating physiological parameters (such as hemodynamics and pressure fluctuations) into virtual models, AI can drive the progression of lesions, such as simulating the growth of tumors or the expansion of aneurysms, rendering training scenarios more dynamically realistic and clinically challenging [4].

2.2. Module 2: AI-enhanced physics engine

2.2.1. Neural network surrogate model

• Technical methods: The study employs Graph Neural Networks (GNNs), Convolutional Neural Networks (CNNs), and other deep learning paradigms are leveraged to learn the input-output mapping relationships inherent in traditional physical simulations. For instance, MeshGraphNets can directly learn the mechanical properties and dynamic behaviors of mesh nodes, achieving real-time and high-fidelity simulations of soft tissue cutting, suturing, pulling, and other operations while maintaining high accuracy and simultaneously speeding up the simulation by several orders of magnitude [6].

2.2.2. Fluid and special effects simulation

Intraoperatively, fluid effects such as bleeding and smoke are critical for constructing a high-fidelity immersive environment. AI is also deployed to accelerate Computational Fluid Dynamics (CFD) simulations, or to synthesize realistic bleeding effects in real-time via generative methodologies (such as GANs). The flow rate and volume of hemorrhage are dynamically regulated based on the diameter and pressure parameters of transected blood vessels. A study in 2025 achieved accurate prediction of the risk of bleeding in the abdominal vascular network ($R^2 \geq 0.91$) by fusing a graph attention network with fluid dynamics equations, and successfully applied it to a VR surgical simulator to enable real-time generation of parameterized, high-fidelity hemorrhagic effects [1].

2.3. Module 3: AI-enabled skill assessment and feedback

Objective and precise skill assessment represents the core of the surgical training loop. AI, by analyzing multimodal training data, can deliver automated assessment and expert-level guidance [7,8].

2.3.1. Operational trajectory and motion analysis

• Technical methods: Spatio-temporal convolutional neural networks (3D-CNNs) and long short-term memory networks (LSTMs) are employed to analyze surgical instrument motion trajectories, velocity, acceleration, and tremor. These spatio-temporal features enable quantitative evaluation of surgical proficiency, operational efficiency, and procedural stability. Graph neural networks (GNNs) are employed to model complex interactions between surgical instruments and tissues, as well as inter-instrumental interactions [7].

DOI: 10.54254/2755-2721/2026.AS31463

• Evaluation Metrics: AI algorithms can compute objective indicators such as path length, idle time, and motion smoothness, and benchmark them against an expert-derived database to generate precise scoring. Research has shown that using AI to quantify the motion trajectory characteristics of surgical instruments (such as tremor intensity and path efficiency) enables the construction of an objective skill evaluation system, and the consistency (Dice coefficient) between its evaluation results and expert subjective evaluations can reach above 0.89 [7,8].

2.3.2. Multimodal fusion evaluation

• Eye-Tracking: Analyze the trainee's attention distribution and visual search strategies to verify their focus on critical anatomical regions.

• Haptic Feedback: Monitors operational force magnitudes and identifies tissue damage risks induced by excessive applied force (such as breaking tissue during suturing).

• Physiological signals: Monitor heart rate, skin conductance response, etc., to evaluate trainees' surgical stress levels.

Based on these analyses, the system can not only generate comprehensive evaluation reports but also dynamically adjust training difficulty through reinforcement learning (RL) strategies, or deliver optimal operational mode demonstrations for trainees via generative adversarial imitation learning (GAIL), thus realizing personalized and adaptive intelligent tutoring [8].

3. Analysis of key challenges

Despite the promising application prospects of AI+VR surgical simulation, its development, clinical translation, and widespread adoption still face a series of challenges.

3.1. Technical challenges

• Physical-visual inconsistency: A prominent technical bottleneck in current systems is the substantial discrepancy between the insufficient fidelity of haptic feedback and that of visual simulation. Specifically, it manifests as:

Limitations of force feedback accuracy: The force feedback resolution of existing commercial VR surgical instruments is typically below 100 millinewtons (0.1N), whereas in real surgery, the mechanical perception threshold of human soft tissues (such as intestines and blood vessels) can be as low as 10-50 millinewtons. This means that simulators cannot accurately reproduce the subtle force variations during membranous tissue dissection or suture tension adjustment [6].

Haptic rendering delay: High-quality visual rendering necessitates a frame rate of over 90Hz to avoid dizziness, but the update frequency for haptic rendering is required to be higher ($\geq$1 kHz) to ensure continuous and stable force. Current systems face difficulties in synchronizing high-fidelity graphics and haptic rendering, resulting in haptic delays often exceeding 30 milliseconds, causing perceptual asynchrony such as "cutting but not feeling" or "force feedback lagging behind visual operation" [6].

Distortion in tissue deformation simulation: When simulating tissue traction or instrument interaction, the visual deformation may be dependent on pre-computed animations, which may not be fully coupled with physically-based force feedback calculations. Research has shown that when the coupling error between visual deformation and force feedback exceeds 2 millimeters, users' immersive experience and operational confidence will significantly decrease. This inconsistency is

DOI: 10.54254/2755-2721/2026.AS31463

particularly prominent around delicate anatomical structures (such as nerves and blood vessels), increasing the risk of developing erroneous muscle memory during training [6].

• Small sample learning problem: It is extremely difficult to obtain a large number of accurately labeled surgical videos and operation data (particularly expert-derived data). How to develop AI models with small samples, weak supervision, or even unsupervised learning is the key to the large-scale advancement of this domain [3]. To address the bottleneck of scarce labeled data, research frontiers are transitioning from the traditional paradigm relying on massive annotations to new paradigms of small sample and weak supervision learning, represented by meta-learning and synthetic data generation [3,4].

• Computational complexity and real-time performance: Complex AI models, such as large diffusion models, demand considerable computational resources. Optimizing these models into lightweight architectures to be compatible with conventional VR hardware platforms is a practical issue [6].

3.2. Clinical and validation challenges

• Insufficient clinical validation: Most studies are still in the technical validation phase, lacking large-scale, multi-center randomized controlled trials (RCTs) to demonstrate that trainees trained using AI-VR achieve significantly superior performance in clinical surgical procedures compared to traditional methods. Many studies have small sample sizes (typically <50 cases), resulting in insufficient statistical power [2,8].

• Standardization and generalizability: There is an absence of unified evaluation metrics and benchmark datasets. AI models developed on different platforms and for various surgical types demonstrate limited generalizability, making them difficult to apply universally [8,9].

4. Outlook on future directions

In response to the aforementioned challenges, future research can make breakthroughs in the following directions:

Multimodal fusion and perception: Develop a comprehensive multimodal perception system that integrates visual, haptic, force, auditory, and physiological signals. The integration of high-fidelity haptic feedback devices and AI prediction models will be the next key research focus to address the issue of sensory inconsistency [5].

Federated Learning and Privacy Protection: By utilizing technologies such as Federated Learning, AI models can be jointly trained across multiple clinical institutions without sharing original patient data, alleviating the dilemmas of data silos and few-shot learning constraints [4].

Explainable AI (XAI): Enhancing the interpretability of AI-driven skill assessment models not only informs learners of their "shortcomings" but also explicitly identifies the root causes of inadequacies and provides targeted improvement strategies, thereby enhancing the system's credibility and educational efficacy [8].

Migration from simulation to reality: Exploring effective strategies for transferring surgical skills acquired via VR simulation to clinical practice. This entails seamless integration with robotic surgical systems (such as the da Vinci surgical robot), applying the parameter settings from virtual training directly to clinical surgical platforms [1].

Establish standards and open ecosystem: Promote the development of open surgical data standards, algorithm benchmarks, and shared datasets; encourage academic and industrial participation; and accelerate innovation [9,10].

5. Conclusion

The deep convergence of AI and VR is propelling surgical training and simulation technologies into a new era of intelligence. This study systematically delineates the three core functionalities of AI in VR-based surgical simulation: scene construction, physical simulation, and skill assessment. Specifically, generative AI is capable of creating highly realistic and diverse anatomical models and surgical scenarios; neural network surrogate models significantly expedite physical simulation, enabling real-time and realistic tissue interaction; and a multimodal data-driven AI skill assessment system delivers objective, precise, and insightful feedback on surgical proficiency. However, on its march toward maturity and widespread clinical translation, this interdisciplinary field still confronts core challenges such as physical-visual sensory inconsistency, scarcity of high-quality labeled data, and insufficient clinical validation. Overcoming these bottlenecks hinges not only on algorithmic innovations but also on the deep convergence of engineering technologies with clinical demands.

Looking ahead, we anticipate the following directions becoming the focus of research: The integration of multimodal perception and feedback technology will target the resolution of sensory consistency issues; cutting-edge paradigms such as federated learning and few-shot learning will effectively mitigate data silos; and advancements in explainable AI (XAI) and simulation-to-reality skill transfer will seek to bridge the "last mile" between virtual training and clinical practice. Ultimately, constructing an open, collaborative, and standardized interdisciplinary R&D ecosystem constitutes the fundamental prerequisite for accelerating the translation of AI-driven intelligent surgical simulation technology from the laboratory to the operating room, ultimately empowering surgeons and benefiting patients.

References

[1] Esposito S, Rebain D, et al. ROOM: Reality Optical Observation in Medicine - A High-Fidelity Simulator for Training Continuum Robots in Bronchoscopy. Robotics and Computer Science. 2025; 15(3): 112-125.
[2] Zuo Ming. The Development and Application of an Audiovisual-Tactile Interactive Metaverse Virtual Surgery System in Clinical Training. Chinese Journal of Medical Education. 2025; 44(2): 45-50.
[3] Frontier Editors (Kuang Ming). Application and Challenges of Generative Artificial Intelligence in Surgical Teaching. Journal of Modern Medical Education. 2025; 10(1): 78-85.
[4] Chen Yu, et al. A Personalized Online Learning Resource Recommendation Algorithm Based on LinUCB for Surgical Education. Journal of Medical Systems. 2024; 48(3): 45-52.
[5] Medicalholodeck Inc. Multi-User Collaborative Surgical Training Platform: Technical Specifications and Case Studies. Journal of Medical Internet Research. 2024; 26(1): e51023.
[6] NVIDIA Corporation. NVIDIA PhysX Engine: Real-Time Physics Simulation for Medical Virtual Environments. In: GPU Technology Conference Proceedings. 2023: 155-160.
[7] Scott D J, et al. Automated Video Analysis and Feedback in Surgical Training Using Convolutional Neural Networks. Journal of Surgical Education. 2024; 81(2): 300-310.
[8] Salvatore E, Daniel R. The ROOM Framework: An Open-Source Platform for Realistic Medical Robotics Simulation. Nature Communications. 2025; 16(1): 89-102.
[9] Frontier Editors. The Application of the Metaverse in Surgical Clinical Teaching: Transforming Medical Education Through Immersive Approaches. Frontiers in Oncology. 2025; 15(4): 233-245.
[10] BM Corporation. Watson for Oncology: A Clinical Decision Support System for Cancer Care. IBM Research Reports. 2023; 8: 77-84.