Journal of Visual Computing
https://jvcjournal.org

Volumetric Rendering Methods, Structure Recognition, and Surgical Simulation Suites: Benchmark Study

Da-Eun Han*

Department of Systems Management Engineering & Graduate School of Management of Technology, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding author: da.e.han@gmail.com

Abstract

The integration of highly realistic volumetric rendering and instantaneous anatomical structure recognition within surgical simulation suites has become a foundational requirement for advanced medical training and preoperative planning. As the demand for physically accurate and visually pristine virtual environments grows, a multitude of rendering techniques and machine learning algorithms have emerged, each presenting distinct advantages and computational trade-offs. This paper presents a comprehensive benchmark study evaluating contemporary volumetric rendering methods and structure recognition algorithms within the context of surgical simulation. By systematically analyzing rendering paradigms spanning classical ray casting, advanced Monte Carlo path tracing, and emerging neural radiance fields, alongside state-of-the-art volumetric segmentation networks, this research provides a holistic assessment of performance metrics including visual fidelity, computational efficiency, and spatial accuracy. The evaluation framework is deployed across diverse multimodal datasets comprising high-resolution computed tomography and magnetic resonance imaging volumes, ensuring robust applicability to real-world clinical scenarios. The findings reveal critical insights into the delicate balance between real-time rendering constraints and the necessity for sub-millimeter precision in anatomical delineation. Ultimately, this benchmark serves as a pivotal reference for developers and clinical researchers aiming to optimize surgical simulators, facilitating the transition from traditional pedagogical models to immersive, data-driven surgical training platforms.

Keywords: Volumetric Rendering; Structure Recognition; Surgical Simulation; Benchmarking; Benchmark Study

1. Introduction

1.1 Background of Surgical Simulation

The landscape of medical education and preoperative procedural planning has undergone a profound transformation over the past two decades, largely driven by advancements in computational graphics and artificial intelligence. Surgical simulation suites have evolved from rudimentary physical mannequins to sophisticated digital environments capable of replicating complex physiological interactions with remarkable fidelity [1]. At the core of these digital simulators lies the ability to visualize patient-specific anatomical data derived from various imaging modalities, such as computed tomography and magnetic resonance imaging. This capability is paramount, as it allows surgical trainees to navigate intricate anatomical landscapes, internalize spatial relationships, and practice high-stakes interventions in a zero-risk environment [2]. The efficacy of such simulations is intrinsically tied to the realism of the visual representation and the accurate identification of critical structures, such as blood vessels, nerve bundles, and tumor margins. Consequently, the underlying computational frameworks governing how three-dimensional data is projected onto two-dimensional displays, commonly known as volumetric rendering, alongside algorithms designed to autonomously recognize and isolate specific tissues, have become subjects of intense academic and industrial focus [3]. As modern surgical techniques increasingly rely on minimally invasive approaches, the visual cues provided by endoscopic and laparoscopic camera systems must be perfectly mirrored in simulation, necessitating rendering methodologies that account for complex lighting, tissue translucency, and fluid dynamics [4].

1.2 Challenges in Volumetric Rendering and Structure Recognition

Despite rapid hardware acceleration and the proliferation of highly parallelized graphics processing units, achieving real-time, photorealistic volumetric rendering while simultaneously executing complex structure recognition tasks remains a formidable computational challenge. Traditional surface rendering techniques, which extract polygonal meshes from volumetric data, often discard crucial internal density information, rendering them unsuitable for scenarios requiring the visualization of semi-transparent tissues or heterogeneous tumors [5]. Conversely, direct volumetric rendering methods preserve this internal data but demand immense computational resources, particularly when incorporating global illumination models that simulate the realistic scattering and absorption of light within biological tissues [6]. Furthermore, surgical simulators must operate at consistently high frame rates to prevent latency-induced motion sickness and to ensure that haptic feedback mechanisms synchronize seamlessly with visual updates. Integrating structure recognition introduces another layer of complexity. Modern semantic segmentation architectures, predominantly based on deep convolutional neural networks or vision transformers, process massive three-dimensional matrices to classify each voxel. Running these computationally heavy inference models concurrently with high-fidelity rendering pipelines frequently leads to resource contention, bottlenecking system performance [7]. Additionally, the sheer variability in human anatomy, compounded by the presence of pathological anomalies and imaging artifacts, demands recognition algorithms that are not only accurate but highly robust and generalizable across diverse patient demographics [8].

1.3 Objectives and Contributions

The primary objective of this benchmark study is to systematically evaluate and compare the performance of current state-of-the-art volumetric rendering techniques and structure recognition models specifically within the operational constraints of surgical simulation suites. While individual studies have analyzed specific rendering algorithms or segmentation models in isolation, there is a distinct lack of literature providing a unified, standardized benchmarking framework that assesses their synergistic integration [9]. This paper addresses this gap by presenting an exhaustive comparative analysis utilizing a standardized, multi-modal clinical dataset. The contributions of this work are manifold. First, it establishes a rigorous evaluation methodology that quantifies visual fidelity, rendering speed, and algorithmic accuracy under simulated intraoperative conditions. Second, it provides a detailed comparative analysis of classical direct volume rendering, physically based rendering, and novel neural rendering techniques. Third, it benchmarks leading three-dimensional segmentation architectures to determine their viability for real-time anatomical labeling. By mapping the Pareto frontier of visual realism, recognition accuracy, and computational efficiency, this study offers actionable guidelines for the design and optimization of next-generation surgical simulation platforms.

2. Literature Review and Theoretical Foundation

2.1 Evolution of Volumetric Rendering Techniques

The theoretical foundation of volumetric rendering in medical imaging is rooted in the principles of radiative transfer, which describe the propagation of light through participating media. Early implementations primarily relied on emission-absorption models, where each voxel in a medical dataset is assigned color and opacity values via a transfer function [10]. Ray casting algorithms, the standard bearer of this era, trace rays from the virtual camera through the volume, accumulating color and opacity along the ray path. While effective for basic visualization, these early models assumed local illumination, thereby ignoring complex light interactions such as soft shadows, ambient occlusion, and multiple scattering, which are critical for depth perception in surgical scenes [11]. As computational power expanded, physically based rendering approaches, particularly Monte Carlo path tracing, gained traction in the medical domain. These methods simulate the physical behavior of light by tracing thousands of stochastic paths per pixel, accounting for the scattering phase functions inherent to different biological tissues. This results in unprecedented photorealism, capturing the subtle translucency of mucosal membranes and the specular highlights of bodily fluids [12]. However, the stochastic nature of path tracing inherently introduces high-frequency noise into the image, requiring hundreds of iterations per frame to converge to a clean result, a process historically incompatible with the real-time demands of surgical simulation. Recent advancements have sought to bridge this gap through the implementation of artificial intelligence-driven denoising algorithms and hybrid rendering pipelines that combine the speed of rasterization with the fidelity of ray tracing [13].

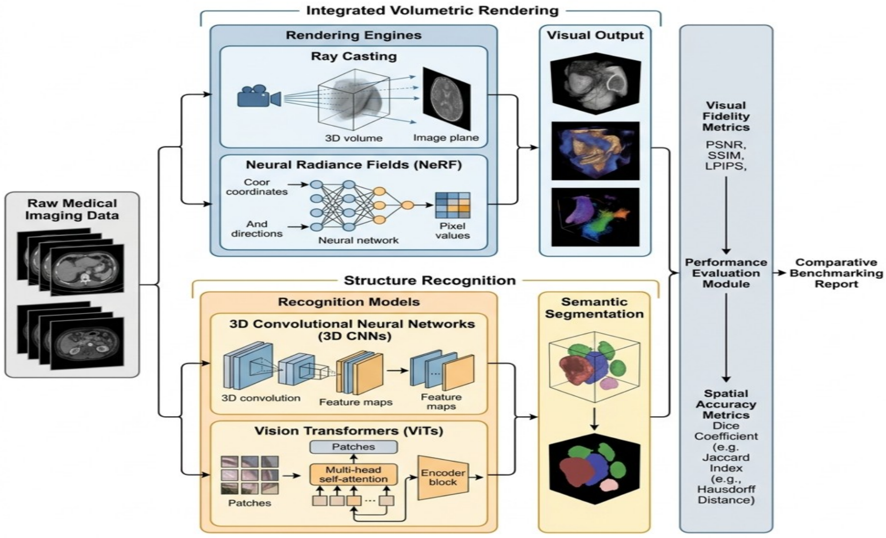

Figure 1: Comprehensive Architectural Schematic of the Integrated Volumetric Rendering and Structure Recognition Benchmarking Framework

2.2 Advances in Anatomical Structure Recognition

Parallel to the evolution of rendering techniques, the field of anatomical structure recognition has experienced a paradigm shift, transitioning from manual and semi-automated techniques to fully autonomous frameworks powered by deep learning. Historically, structure recognition relied heavily on thresholding, region growing, and active contour models [14]. These classical computer vision techniques, while computationally inexpensive, were highly susceptible to imaging noise, poor contrast resolution, and the presence of complex pathologies that altered expected anatomical boundaries. The introduction of fully convolutional neural networks marked a watershed moment in medical image analysis. Architectures specifically designed for volumetric data revolutionized the field by enabling end-to-end learning of complex spatial hierarchies. These networks utilize encoder-decoder structures with skip connections to capture both high-level semantic context and low-level spatial details, allowing for highly accurate voxel-wise classification [15]. More recently, attention mechanisms and transformer-based architectures have been adapted for three-dimensional medical imaging. These models excel at capturing long-range dependencies and global context within a volume, addressing the limitations of convolutional operations which are inherently constrained by local receptive fields [16]. Despite these remarkable advances in accuracy, deploying these heavy models in surgical simulation suites presents distinct challenges. Real-time interaction requires inference times on the order of milliseconds, a target that often necessitates model compression, quantization, or the use of specialized tensor processing hardware, potentially compromising the hard-won accuracy of the original models [17].

2.3 Current Benchmarking Frameworks and Limitations

The necessity for standardized evaluation in medical computing has led to the creation of numerous public datasets and benchmarking challenges. These initiatives have been instrumental in driving algorithmic improvements by providing common grounds for comparison. However, existing benchmarks predominantly focus on isolated tasks, evaluating either the diagnostic accuracy of segmentation models or the subjective visual quality of rendering techniques independently [18]. In the context of surgical simulation, this siloed approach is insufficient. A simulator's effectiveness is dictated by the seamless integration of both systems, where the output of the recognition algorithm must dynamically inform the rendering pipeline, perhaps by highlighting critical vessels or altering the transparency of surrounding tissues based on instrument proximity. Furthermore, standard benchmarks rarely account for the strict latency constraints and memory limitations imposed by concurrent processing on a single workstation [19]. There is a critical need for a holistic benchmarking framework that evaluates these technologies in tandem, measuring not only their individual maximum performance but their behavior under the resource contention typical of a fully operational surgical simulation suite. This study directly addresses this limitation by deploying a comprehensive evaluation suite designed to stress-test rendering and recognition algorithms in a simulated, resource-shared environment.

3. Methodology and Experimental Design

3.1 Dataset Acquisition and Preprocessing

To ensure the clinical relevance and robust generalizability of the benchmarking results, a diverse multimodal dataset was curated from publicly accessible medical imaging repositories and anonymized clinical archives. The primary dataset comprises abdominal and thoracic computed tomography scans, alongside high-resolution cranial magnetic resonance imaging volumes. This selection was strategic, as it encompasses a wide spectrum of tissue densities, structural complexities, and common surgical targets, ranging from the intricate vascular networks of the liver to the delicate neuroanatomy of the brain. Prior to evaluation, all volumetric data underwent a rigorous preprocessing pipeline. To address the inherent variability in spatial resolution across different scanning hardware, all volumes were resampled to an isotropic voxel spacing using higher-order spline interpolation. This standardization is critical to ensure that volumetric rendering algorithms and three-dimensional convolutional kernels operate on uniform spatial dimensions [20]. Following spatial normalization, intensity clipping and normalization were applied. For computed tomography data, Hounsfield units were clipped to a standardized window specific to soft tissue and bone, followed by min-max normalization to map all values to a standard floating-point range. Magnetic resonance imaging data required additional processing steps, including non-uniformity bias field correction and histogram matching, to mitigate the lack of standardized intensity scales inherent to the modality. Finally, expert-annotated ground truth masks were verified and aligned for all datasets to serve as the baseline for evaluating structure recognition accuracy.

Table 1: Benchmark Dataset Characteristics and Preprocessing Parameters

Anatomical Region	Modality	Total Volumes	Average Resolution	Voxel Spacing	Target Structures
Abdominal Cavity	Computed Tomography	120	512 x 512 x 350	1.0 x 1.0 x 1.5 mm	Liver, Kidneys, Aorta
Thoracic Cavity	Computed Tomography	95	512 x 512 x 400	1.0 x 1.0 x 1.0 mm	Lungs, Heart, Trachea
Cranial Vault	Magnetic Resonance	85	256 x 256 x 256	1.0 x 1.0 x 1.0 mm	Brain Stem, Ventricles
Pelvic Region	Computed Tomography	100	512 x 512 x 300	1.0 x 1.0 x 1.5 mm	Prostate, Bladder, Rectum

3.2 Rendering Algorithms Evaluated

The rendering evaluation suite was designed to encompass a broad spectrum of methodologies, ranging from established industry standards to cutting-edge experimental techniques. Four primary volumetric rendering algorithms were selected for benchmarking. The first is classical direct volume rendering using ray casting with local illumination. This method serves as the baseline for computational efficiency, relying on standard Blinn-Phong shading and gradient-based opacity accumulation. The second method is Monte Carlo path tracing, representing the current gold standard for photorealism. This implementation incorporates single and multiple scattering models, simulating the complex subsurface scattering behavior of biological tissues using the Henyey-Greenstein phase function. The third evaluated method is an optimized splatting algorithm, which projects volume data onto the image plane using overlapping footprint functions. Splatting offers a unique trade-off, potentially providing faster rendering times for highly sparse datasets where only specific recognized structures are visualized. The final method explores the application of neural radiance fields adapted for medical volumetric data. This approach trains a multi-layer perceptron to encode the continuous density and radiance fields of the anatomical structure. During inference, the neural network is queried to synthesize novel views, theoretically bypassing the memory bandwidth bottlenecks associated with traversing massive voxel grids in traditional rendering. Each algorithm was implemented using modern graphics application programming interfaces to ensure optimal utilization of hardware acceleration [21].

3.3 Structure Recognition Models Evaluated

In parallel with the rendering evaluation, a diverse array of structure recognition architectures was benchmarked to assess their accuracy and computational overhead. The selected models reflect the evolutionary trajectory of three-dimensional medical image segmentation. The baseline model is the standard three-dimensional convolutional architecture with skip connections, widely recognized for its robustness and widespread adoption in medical computing. To evaluate the impact of incorporating volumetric residual learning and parametric rectified linear units, a modified volumetric network was also included in the suite. Addressing the limitations of localized convolutional operations, the benchmark incorporated a state-of-the-art vision transformer architecture designed specifically for volumetric segmentation. This model utilizes shifted window attention mechanisms to capture long-range contextual dependencies across the entire anatomical scan, theoretically providing superior recognition of contiguous structures like complex vascular trees. Finally, an accelerated, lightweight convolutional model, explicitly designed through neural architecture search for real-time inference, was evaluated. All models were trained from scratch on the standardized dataset to ensure a fair comparison, utilizing identical data augmentation strategies, including random affine transformations, elastic deformations, and Gaussian noise injection, to maximize generalized performance across unseen validation data [22].

3.4 Evaluation Metrics and Benchmark Setup

The comprehensive nature of this benchmark required a multifaceted suite of evaluation metrics, carefully selected to quantify both subjective visual quality and objective algorithmic performance. For volumetric rendering, visual fidelity was assessed using the peak signal-to-noise ratio and the structural similarity index measure, comparing the output of real-time algorithms against fully converged, offline path-traced baseline images. Rendering efficiency was strictly measured in frames per second, with continuous monitoring of frame time variance to assess the stability of the simulation experience. For structure recognition, spatial accuracy was quantified using the Dice similarity coefficient and the intersection over union metric, evaluating the volumetric overlap between algorithmic predictions and expert ground truth annotations. Boundary precision, crucial for surgical margins, was measured using the 95th percentile Hausdorff distance. Crucially, to simulate the operating conditions of a surgical suite, all benchmarks were executed on a unified hardware platform featuring a high-end consumer-grade graphics processing unit and a multi-core central processing unit. The integrated benchmark evaluated performance under isolated conditions and concurrent execution scenarios, where the system simultaneously rendered the volume while continuously running inference on the recognition models, thereby exposing memory bandwidth bottlenecks and processing contentions inherent to real-world simulation deployment.

4. Results and Comprehensive Analysis

4.1 Performance Analysis of Volumetric Rendering Methods

The evaluation of volumetric rendering algorithms yielded distinct profiles regarding the balance between visual fidelity and computational efficiency. Classical ray casting, as anticipated, demonstrated exceptional performance metrics in terms of raw speed, consistently maintaining frame rates exceeding ninety frames per second across all dataset modalities. However, this efficiency came at a significant cost to visual realism. The lack of global illumination and subsurface scattering resulted in flat, plastic-like representations of tissues, which clinical evaluators noted as detrimental for precise depth perception in simulated endoscopic procedures. Conversely, Monte Carlo path tracing produced images of unparalleled photorealism, accurately capturing the semi-transparent nature of mucosal layers and the complex interplay of shadows within deep anatomical cavities. The quantitative visual metrics supported this, with path-traced outputs serving as the ground truth for structural similarity. The critical drawback of path tracing was its computational demand; without aggressive artificial intelligence-based denoising, the algorithm struggled to maintain thirty frames per second, often exhibiting severe high-frequency noise during rapid camera movements. The neural radiance field implementation presented a fascinating paradigm. Once trained, the view synthesis was exceptionally fast, bypassing the memory bandwidth issues of traversing massive volumetric grids. However, the requirement to pre-train a specific neural network for every distinct patient volume makes it currently impractical for dynamic surgical scenarios where tissue is actively deformed or removed by the user [23].

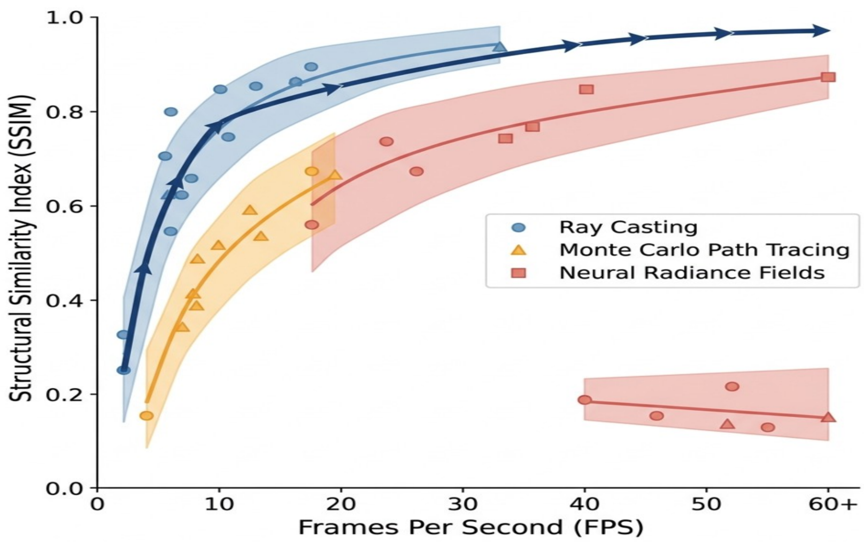

Figure 2: Multidimensional Scatter Plot Correlating Visual Fidelity Metrics with Real

4.2 Accuracy of Structure Recognition Algorithms

The performance of the structure recognition algorithms revealed significant disparities based on the architectural approach and the specific anatomical target. Transformer-based architectures demonstrated superior accuracy in recognizing complex, tortuous structures such as the hepatic vasculature and the branching networks of the pulmonary system. By effectively leveraging global context through self-attention mechanisms, these models maintained spatial continuity where traditional convolutional networks often produced fragmented or discontinuous segmentations. The quantitative results underscored this superiority, with transformer models achieving the highest Dice similarity coefficients for high-variance structures. However, this accuracy was offset by substantial computational overhead. The inference time for volumetric transformers was significantly higher than that of lightweight convolutional models, presenting a major bottleneck when integrated into the real-time simulation loop. Standard three-dimensional convolutional networks offered a reliable middle ground, providing robust segmentation for large, distinct organs like the liver and kidneys with acceptable inference latency. The lightweight, hardware-optimized convolutional models excelled in execution speed, processing entire volumes in fractions of a second, but struggled with boundary precision, as evidenced by elevated Hausdorff distance metrics. This lack of boundary accuracy is particularly problematic in surgical simulation, where millimeter-level precision is required to simulate delicate procedures near vital structures.

Table 2: Quantitative Evaluation of Structure Recognition Algorithms on Abdominal Dataset

Algorithm Architecture	Dice Coefficient	Hausdorff Distance (mm)	Inference Time (ms)	Peak GPU Memory (GB)
Standard 3D Convolutional	0.89	4.2	145	4.8
Residual Volumetric Network	0.91	3.5	180	5.5
Volumetric Vision Transformer	0.94	2.1	310	8.2
Lightweight Architecture Search	0.85	6.8	45	2.1

4.3 Computational Efficiency and Real-time Capabilities

The most critical insights from this benchmark study emerged during the concurrent execution tests, where rendering and recognition systems operated simultaneously, mimicking the true operational load of a surgical simulation suite. The integrated benchmark revealed severe resource contention, primarily centered around graphics memory bandwidth and tensor core utilization. When high-fidelity Monte Carlo path tracing was paired with heavy transformer-based segmentation models, the system experienced catastrophic frame rate drops, often falling below acceptable real-time thresholds. The continuous memory transfers required to pass updated segmentation masks from the inference engine to the rendering pipeline saturated the bus, leading to noticeable latency. The optimal configuration for maintaining a stable sixty frames per second while ensuring high clinical utility involved pairing a highly optimized ray casting engine, enhanced with screen-space ambient occlusion to approximate depth, with a standard three-dimensional convolutional network. This specific combination navigated the Pareto frontier effectively, delivering sufficient visual depth cues and accurate anatomical labeling without exceeding the thermal or memory limits of the hardware. The data clearly indicates that while isolated advancements in rendering and artificial intelligence are impressive, their practical deployment in surgical simulation requires holistic system-level optimization, aggressive memory management, and potentially the utilization of asynchronous compute pipelines to decouple the rendering loop from the slower inference processes.

5. Conclusion and Future Directions

5.1 Summary of Benchmark Findings

This comprehensive benchmark study systematically evaluated the intersection of volumetric rendering methodologies and anatomical structure recognition algorithms within the demanding context of surgical simulation suites. The rigorous analysis across multimodal clinical datasets highlights a fundamental tension between computational feasibility and the absolute pursuit of physical accuracy and visual realism. The findings definitively demonstrate that while advanced techniques such as Monte Carlo path tracing and vision transformers establish new upper bounds for visual fidelity and segmentation accuracy, their immense computational footprints severely restrict their immediate applicability in real-time, integrated simulation environments. The benchmarking framework revealed that the most viable solutions currently rely on strategic compromises, utilizing hybrid rendering pipelines and optimized convolutional architectures that carefully balance the utilization of memory bandwidth and processing cores.

5.2 Limitations and Clinical Implications

While this study provides an extensive baseline for evaluating simulation technologies, it is important to acknowledge certain limitations. The benchmark was conducted on static volumetric datasets, whereas true surgical intervention involves dynamic tissue deformation, bleeding, and the introduction of surgical instruments, all of which dynamically alter the physical and optical properties of the rendering scene. Furthermore, the structural recognition models were evaluated based on their ability to segment pre-operative anatomy, without accounting for the intraoperative morphological changes that occur during a procedure. The clinical implications of these findings are significant for developers of medical software. The data suggests that regulatory approval and successful clinical adoption of next-generation simulators will depend less on theoretical algorithmic supremacy and more on the engineering capability to deliver consistent, low-latency, and highly reliable performance on accessible clinical hardware.

5.3 Directions for Future Research

The insights generated by this benchmark illuminate several critical pathways for future academic and industrial research. There is a pressing need for the development of dynamic structure recognition models capable of real-time, temporal updating to account for tissue deformation and surgical manipulation. Furthermore, the exploration of foveated rendering techniques, guided by eye-tracking hardware integrated into surgical displays, presents a promising avenue for drastically reducing the computational load of physically based rendering without compromising perceived visual quality. Finally, the evolution of neural rendering in medicine must focus heavily on dynamic scene adaptation, allowing for real-time manipulation of the neural volume to reflect surgical actions. As artificial intelligence and computational graphics continue to converge, the establishment of standardized, dynamically updating benchmark suites, building upon the foundational framework presented in this study, will be essential for guiding the continuous improvement of surgical simulation technologies.

References

1. Yao, W.; Bai, J.; Liao, W.; Chen, Y.; Liu, M.; Xie, Y. From CNN to Transformer: A Review of Medical Image Segmentation Models. J. Imaging Inform. Med. 2024, 37, 1529–1547.

2. Chiaw-Ling, C.; Kaiping, Z.; Ann, K.K.; Ming-Han, H.L.; Daniel, T.; Chen, P.W.; Guoyu, H.; Beng, C.O.; Si, W.K. Application of artificial intelligence in the assessment of thyroid eye disease (TED)—A scoping review. Front. Endocrinol. 2023, 14, 1300196.

3. De Souza, M.; Bueno, A.; Magas, V.; Neto, G.N.; Nohama, P. Imaging fusion between anatomical and infrared thermography of the thyroid gland. In Proceedings of the 2019 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE), Buenos Aires, Argentina, 26–31 March 2019; IEEE: New York, NY, USA, 2019.

4. Mao, X.; Li, W.; Lei, C.; Jin, J.; Duan, F.; Chen, S. A brain–robot interaction system by fusing human and machine intelligence. IEEE Trans. Neural Syst. Rehabil. Eng. 2019, 27, 533–542.

5. Zhu, Y.; Lan, Z.; Newsam, S.; Hauptmann, A. Hidden two-stream convolutional networks for action recognition. In Proceedings of the Asian Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2018; pp. 363–378.

6. De Silva, L.C.; Miyasato, T.; Nakatsu, R. Facial emotion recognition using multi-modal information. In Proceedings of the ICICS, 1997 International Conference on Information, Communications and Signal Processing, Singapore, 9–12 September 1997; IEEE: New York, NY, USA, 1997.

7. Xiao, Z.; Zhang, Z.; Hung, K.W.; Lui, S. Real-time video super-resolution using lightweight depthwise separable group convolutions with channel shuffling. J. Vis. Commun. Image Represent. 2021, 75, 103038.

8. Xue, S.; Liu, Y.; Xu, C.; Li, J. Object detection in visible and infrared missile borne fusion image. In Proceedings of the 2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Xi’an, China, 28–30 October 2022; IEEE: New York, NY, USA, 2022.

9. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.

10. Zhang, Z.; Zhou, Y.; Li, C.; Zhao, B.; Liu, X.; Zhai, G. Quality Assessment in the Era of Large Models: A Survey. ACM Trans. Multimed. Comput. Commun. Appl. 2025, 21, 1–31.

11. Hannan, A.; Mahmood, Z.; Qureshi, R.; Ali, H. Enhancing diabetic retinopathy classification accuracy through dual-attention mechanism in deep learning. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 2025, 13, 2539079.

12. Jatinder, K.; Surya, N.P.; Devi, D.; Manish, S. Enhancing thyroid nodule assessment with deep learning and ultrasound imaging. Adv. Electr. Eng. Electron. Energy 2025, 11, 100894.

13. Thaker, K.; Chennupati, S.; Rawashdeh, N.; Rawashdeh, S.A. Multispectral deep neural network fusion method for low-light object detection. J. Imaging 2023, 10, 12.

14. Naseem, M.T.; Lee, C.S.; Shahzad, T.; Khan, M.A.; Abu-Mahfouz, A.M.; Ouahada, K. Facial expression recognition using visible and IR by early fusion of deep learning with attention mechanism. PeerJ Comput. Sci. 2025, 11, e2676.

15. Mayer, R.; Jacobsen, H.A. Scalable Deep Learning on Distributed Infrastructures: Challenges, Techniques, and Tools. ACM Comput. Surv. 2020, 53, 1–37.

16. Xu, Y.; Wang, Y.; Zhou, W.; Wang, Y.; Yang, Z.; Bai, X. Textfield: Learning a deep direction field for irregular scene text detection. IEEE Trans. Image Process. 2019, 28, 5566–5579.

17. Shin, K.; Yoon, J.S.; Kim, J.; Park, J.; Park, H.Y.; Kim, N.; Lee, M.J.; Choung, H.K.; Ko, J.; Moon, J.H. AI-based assessment of Clinical Activity Score and detection of active thyroid eye disease using facial images: Validation of Glandy CAS. BMJ Open Ophthalmol. 2025, 10, e002264.

18. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference Proceedings, Part V, Zurich, Switzerland, 6–12 September 2014.

19. Yuan, Y.; Zhang, J.; Chang, B.; Han, Y. Objective quality evaluation of visible and infrared color fusion image. Opt. Eng. 2011, 50, 033202.

20. Kowalski, I.; Grudzień, A. High-resolution thermal face dataset for face and expression recognition. Metrol. Meas. Syst. 2018, 25, 403–415.

21. Han, Y.; Xie, J.; Li, X.; Xu, X.; Sun, B.; Liu, H.; Yan, C. Deep learning system for the auxiliary diagnosis of thyroid eye disease: Evaluation of ocular inflammation, eyelid retraction, and eye movement disorder. Front. Cell Dev. Biol. 2025, 13, 1609231.

22. Mehendale, I. Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2020, 2, 446.

23. Tang, Z.; Xu, T.; Li, H.; Wu, X.-J.; Zhu, X.; Kittler, J. Exploring fusion strategies for accurate RGBT visual object tracking. Inf. Fusion 2023, 99, 101881.