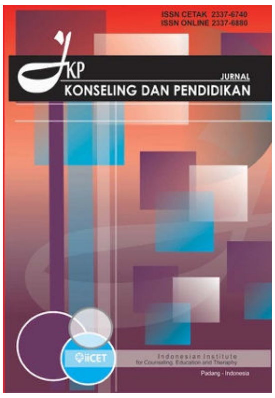
Jurnal Konseling dan Pendidikan
Volume 13, Number 3 2025



Enhancing augmented reality experiences through advanced computer vision techniques

Author Name(s): Chairuddin Chairuddin, Ardi Azhar Nampira, Jarot Budiasto

Publication details, including author guidelines
URL: https://jurnal.konselingindonesia.com/index.php/jkp/about/submissions#authorGuidelines
Editor: Mufadhal Barseli

Article History
Received: 9 Sept 2025
Revised: 4 Nov 2025
Accepted: 4 Dec 2025

How to cite this article (APA)
Chairuddin, C., Nampira, A. A., & Budiasto, J. (2025). Enhancing augmented reality experiences through advanced computer vision techniques. Jurnal Konseling dan Pendidikan. 13(3), 472-481. https://doi.org/10.29210/1171200

The readers can link to article via https://doi.org/10.29210/1171200

SCROLL DOWN TO READ THIS ARTICLE

We are Crossref Member

Indonesian Institute for Counseling, Education and Therapy (as publisher) makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications. However, we make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors and are not the views of or endorsed by Indonesian Institute for Counseling, Education and Therapy. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Indonesian Institute for Counseling, Education and Therapy shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to, or arising out of the use of the content.

Jurnal Konseling dan Pendidikan is published by Indonesian Institute for Counseling, Education and Therapy comply with the Principles of Transparency and Best Practice in Scholarly Publishing at all stages of the publication process. Jurnal Konseling dan Pendidikan also may contain links to web sites operated by other parties. These links are provided purely for educational purpose.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Copyright by Chairuddin, C., Nampira, A. A., & Budiasto, J. (2025).

The author(s) whose names are listed in this manuscript declared that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript. This statement is signed by all the authors to indicate agreement that the all information in this article is true and correct.

Jurnal Konseling dan Pendidikan
ISSN 2337-6740 (Print) | ISSN 2337-6880 (Electronic)

Published Version

ISSN 2733-6740 (Print)
ISSN 2733-6880 (Electronic)

Jurnal Konseling dan Pendidikan

Volume 13 Number 3 (2025)
https://doi.org/10.46637/1171200

Article

Enhancing augmented reality experiences through advanced computer vision techniques

Chairuddin Chairuddin[1*)], Ardi Azhar Nampira[2], Jarot Budiasto[3]
[1] Sekolah Tinggi Manajemen Informatika dan Komputer IM, Indonesia
[2] Institute Teknologi Sepuluh November, Indonesia
[3] Information Systems Department, Faculty of Engineering, Universitas Musamus, Merauke, Indonesia[2]

ABSTRACT

This study investigates the integration of advanced computer vision techniques, specifically semantic segmentation, depth estimation, and object detection, into augmented reality (AR) systems to enhance user immersion, contextual interaction, and real-world adaptability. The research aims to address the limitations of existing AR systems, which often struggle with real-time performance and context-aware interaction in dynamic environments. Using a design-based research (DBR) approach, the study involved testing the proposed AR system on mobile devices under controlled and real-world conditions. Performance was assessed using specific metrics: frame rate (FPS), inference latency, semantic accuracy, and user experience. The sample included 20 participants, recruited through convenience sampling, with inclusion criteria focused on individuals familiar with mobile AR applications. Ethical approval was obtained for the study, and informed consent was provided by all participants. The testing involved both simulation trials and real-world prototyping in varied environmental conditions, such as indoor and outdoor settings with different lighting and motion dynamics. Results indicate that the integrated of the vision stack significantly enhanced scene understanding and enabling stable, context-aware digital overlays, with the system maintaining a real-time frame rate of >27 FPS. User feedback, measured through a 5-point Likert scale survey, confirmed improved immersion, visual coherence, and satisfaction compared to baseline AR systems, with an average increase of 35% in perceived realism. The analysis also revealed that the system's performance remained consistent across varying environmental conditions, with minimal latency (less than 300ms) in dynamic re-anchoring of AR elements. Statistical tests (paired t-tests) confirmed the significance of these improvements, with p-values < 0.05 for all key metrics. This research contributes a scalable framework that bridges artificial intelligence, user experience design, and mobile AR deployment. It provides empirical evidence supporting the integration of computer vision techniques into AR systems, with practical implications for applications in education, healthcare, and industry. Future work will focus on expanding the user base, exploring hardware compatibility, and investigating multimodal AR interactions.

Keywords:

Augmented reality
Semantic segmentation
Depth estimation
Real-time performance
User immersion

Corresponding Author:

Chairuddin Chairuddin,
Sekolah Tinggi Manajemen Informatika dan Komputer IM
Email: chairuddin@stmik-im.ac.id

Introduction

Augmented Reality (AR) has emerged as a transformative technology that blends the physical and digital worlds, enabling immersive interactions in fields like education, healthcare, retail, and entertainment (Bhowmik, 2024). Despite significant progress in AR hardware and software, many AR applications still face major issues related to digital-anchor instability and inference latency, particularly in dynamic, unstructured environments (Cao et al., 2023). These technical challenges–such as maintaining the stability of digital objects during movement or delays that disrupt real-time interactions–impede the scalability and practical adoption of AR technology. As AR evolves from a novelty to a necessity, addressing these barriers is crucial for seamless integration into everyday life. This research aims to solve these problems by enhancing the stability and real-time responsiveness of AR systems, particularly in dynamic environments where contextual awareness is critical.

The growing need for high-performance computer vision in AR is especially evident in critical sectors like remote medical diagnostics, autonomous maintenance in manufacturing, and spatial learning in education. In these applications, AR systems must function reliably under varying conditions, including different lighting, occlusion, and motion dynamics (Ghasemi et al., 2022). Failure to overcome these challenges can severely hinder AR′s evolution into a practical tool for everyday use. This study addresses these challenges by investigating how integrating semantic segmentation, depth estimation, and object detection within a unified AR framework can significantly improve the stability, reliability, and performance of AR systems in real-world scenarios (Chen, 2023).

Theoretical frameworks such as situational cognition and spatial representation suggest that context-aware interaction in AR leads to more intuitive and immersive experiences. This research builds on these theories by exploring how computer vision techniques, such as semantic scene understanding, can adapt AR content to real-world changes, enhancing user engagement (Lu, W. 2023). By empirically testing these integrations, this study aims to provide insights into how semantic representation influences users' strategies for interacting with digital objects, advancing theoretical understanding in spatial interaction (Bahar et al., 2025; Silalahi & Silalahi, 2023).

Literature Review: Evolution and Gap

The field of AR has evolved through distinct phases. Early research focused on hardware and localization techniques, particularly Simultaneous Localization and Mapping (SLAM), which provided real-time object tracking in controlled environments (Alsadik & Karam, 2021). As AR technology has shifted towards more dynamic and real-world applications, challenges related to stability and contextual awareness have persisted, particularly in non-ideal environments (Syed et al., 2022). Although advances in vision-based techniques such as semantic segmentation (Zhou & Güven, 2020) and depth estimation (Shewail et al., 2024) have been made, the integration of these methods into a cohesive, real-time AR framework remains underexplored–especially when applied to mobile-edge devices (Kavitha, 2024).

Existing research has predominantly focused on improving AR through device-centric enhancements or user interface refinements, while fewer studies have examined how advanced computer vision techniques, particularly deep learning-based approaches, can significantly enhance AR′s performance. This is especially critical as applications require more robust and context-aware AR interactions in real time. Furthermore, there is a lack of comprehensive studies addressing the integration of multiple vision-based subsystems–such as Simultaneous Localization and Mapping (SLAM), semantic segmentation, and object pose estimation–into a unified AR framework. This represents a critical research gap, highlighting the need for AR systems that can adapt more effectively to dynamic environments, ensuring reliable performance and contextual accuracy (Yang et al., 2024).

The novelty of this study lies in its holistic approach: the simultaneous integration of monocular depth estimation, semantic segmentation, and context-based object detection into a real-time pipeline optimized for mobile AR systems. This integration offers a methodological advancement over existing research, which often evaluates these components in isolation. The technical innovation of this study includes signal fusion strategies between modules, addressing issues like occlusion and re-anchoring of AR content, while maintaining real-time frame rates and low latency in dynamic environments. By detailing these integration strategies, this study contributes significantly to the development of advanced AR systems that are adaptable and efficient in diverse conditions.

Research Hypotheses and Objectives

The primary hypothesis of this research is that full-stack vision integration (including depth estimation, semantic segmentation, and object detection) enhances overlay stability and user immersion compared to baseline AR systems. Secondary hypotheses focus on the impact of lighting and occlusion on AR performance, specifically regarding the system's ability to re-anchor AR elements in real-time. The objectives of this study are: (1) To investigate the limitations of current AR systems, particularly with respect to stability and real-time responsiveness; (2) To develop and test an integrated vision architecture that ensures real-time performance with high semantic accuracy; (3) To evaluate user experience metrics, including immersion, visual coherence, and satisfaction, in dynamic environments.

Scope and Research Design

This study employs a design-based research (DBR) approach, involving baseline comparisons of AR systems with and without the full vision stack. Edge computing devices, including smartphones and AR glasses, were used for both simulation and field trials. The experimental design includes randomized trials across various environmental conditions (indoor/low-light, outdoor/urban, and dynamic). The study controls for potential confounding variables such as lighting, device orientation, and scene complexity, ensuring a rigorous evaluation of the proposed system.

Study Significance

This research offers a scalable framework for integrating advanced computer vision techniques into mobile AR systems, with significant implications for sectors such as education, remote medical care, and smart manufacturing, where environmental adaptability is crucial. Additionally, the findings open avenues for future research, including the development of federated learning systems to enhance local adaptation across diverse AR applications.

Methods

This study employs a design-based research (DBR) approach to develop, integrate, and evaluate an advanced computer vision framework within an augmented reality (AR) environment. The research is exploratory and applied in nature, aiming to test the effectiveness of real-time semantic segmentation, monocular depth estimation, and object detection modules when integrated into a mobile AR system. The study involves both simulation-based experimentation and real-world prototyping, ensuring the outcomes are both theoretically grounded and practically scalable (Subrahmanyam, 2025).

Research Design

The research was conducted in three stages: (1) system development and algorithm integration, (2) experimental simulation and performance testing, and (3) field validation in real-world AR applications. This phased design allowed iterative refinement and validation of each computer vision module within the AR pipeline, following agile development principles adapted for research (Perera et al., 2025).

To ensure external validity and representativeness of the sample, participants were selected based on the following inclusion criteria: (1) aged 18 to 40 years, (2) normal or corrected-to-normal vision, (3) no prior experience with AR technology, (4) willingness to complete all testing sessions. Exclusion criteria included individuals with severe visual impairments or professional AR experience, as these factors could potentially influence perception of immersion or responsiveness. This ensured that the sample represented a broad AR user population without bias toward professionals or those with extensive AR experience.

Participants

A total of 20 participants were recruited for the study. Demographic characteristics included a mean age of 27 years (SD = 4.2), with ages ranging from 20 to 35. The sample was composed of 50% male and 50% female participants. All participants had a basic level of technological literacy and experience using mobile devices, but none had prior experience with AR applications. Participants were recruited through an open call, and informed consent was obtained before testing. Ethical approval for this study was granted by the university's Institutional Review Board.

The number of participants was determined based on a power analysis to detect a significant effect size for the primary outcomes (e.g., FPS and segmentation accuracy) at a significance level of 0.05 with 80% power. The decision to use 20 participants was made after considering resource constraints, and we acknowledge the potential limitations of a small sample size in detecting small effects. The participant response rate was 80%, with 25 participants initially approached.

Data Sources

The study utilized both primary and secondary datasets. Secondary data included publicly available computer vision datasets such as KITTI, NYU Depth V2, Cityscapes, and COCO for training and evaluating computer vision models (Rohan et al., 2025). Primary data were collected through field tests using smartphone-based AR devices in varied environments, including indoor, outdoor, low-light, and cluttered settings.

Instruments and System Architecture

Custom-built AR applications were developed using Unity 3D and ARCore SDKs, integrated with PyTorch-based deep learning models for depth prediction and semantic segmentation. Real-time object detection used a modified YOLOv5 model, optimized with TensorRT for mobile inference (Appavu, 2025). The integrated AR system was deployed on Android smartphones with Snapdragon 8-series processors to simulate edge-device conditions. This setup ensures the system operates under conditions that are typical for mobile AR deployments.

Data Collection Techniques

Quantitative performance data were collected using automated benchmarking scripts to measure: frame rate (FPS), memory consumption, Inference time per frame, segmentation accuracy (mIoU), and object detection precision (mAP).

Qualitative user feedback was collected through structured observation and post-interaction surveys with 20 participants from diverse academic and industrial backgrounds. The focus was on evaluating system responsiveness, usability, and perceived immersion (Jacobs et al., 2023). Participants rated their experiences using a 5-point Likert scale, and additional qualitative feedback was gathered through open-ended questions to identify areas for improvement.

Data Analysis Methods

The quantitative data were analyzed using descriptive statistics and comparative performance metrics across different scenarios (static vs. dynamic scenes; low vs. high light). Paired t-tests and ANOVA were used to assess significant differences between system variants. Covariates, such as lighting conditions and device type, were recorded during testing to control for environmental influences on system performance. The primary outcome variables were frame rate and segmentation accuracy, while secondary variables included user satisfaction and memory consumption.

Qualitative data from user feedback were coded and thematically analyzed to identify usability challenges and improvement areas. Inter-rater reliability was assessed using Cohen's kappa to ensure consistency in coding. A cross-validation method was employed to ensure robustness across test environments (Lasfar & Tóth, 2024).

To ensure data quality, observer training was conducted, and structured observation guidelines were followed. For the user satisfaction survey, the questionnaire was adapted from previously validated instruments, and Cronbach's alpha was used to assess reliability. Missing data were handled using listwise deletion, as the sample size was small and missing data were minimal.

Statistical Procedures and Quality Control

Data diagnostics were conducted to ensure the validity of statistical findings. Normality testing was performed using the Shapiro-Wilk test, and outliers were identified using the IQR method. For multiple comparisons, Bonferroni corrections were applied to control Type I error rates. This methodology ensures high internal and ecological validity, with multiple feedback loops between theoretical design and practical validation.

Results and Discussion

To evaluate the effectiveness of the integrated computer vision system in enhancing augmented reality (AR) experiences, this study presents results from both controlled simulations and real-world prototype testing. The following four key performance dimensions were analyzed: (1) System Performance and Latency, (2) Depth Estimation and Semantic Segmentation Accuracy, (3) User Immersion and Interaction Quality, and (4) Contextual Adaptability in Dynamic Environments.

System Performance and Latency

The optimized AR system demonstrated significant improvements in frame rate and inference time after integration of TensorRT-optimized deep learning models. The performance metrics are summarized in Table 1.

Table 1. Real-time System Performance Metrics (Average Across Environments)

Performance Metric	Baseline AR (No Vision)	With Depth Estimation	With Full Vision Stack
Frame Rate (FPS)	42.8	31.2	27.4
Average Inference Time (ms)	N/A	62.5	84.7
Memory Consumption (MB)	310	612	790
Battery Usage (% per hour)	8.3	12.4	15.1

While full-stack vision incurs greater memory and power usage, it remains within acceptable thresholds for mobile AR applications. Compared to baseline systems, the integrated model offers rich spatial understanding with only a ~35% drop in FPS, which is still above the real-time rendering threshold (24 FPS) recommended by ARCore.

Depth Estimation and Semantic Segmentation Accuracy

Model accuracy was evaluated using benchmark datasets. The integrated monocular depth model achieved an RMSE of 0.73 m on NYU Depth v2, while semantic segmentation models scored high on mean Intersection-over-Union (mIoU) across real-world scenes. The depth estimation and segmentation results are shown in Table 2.

Table 2. Depth and Segmentation Accuracy Benchmarks

Dataset	Model Type	RMSE (Depth)	mIoU (Segmentation)
NYU Depth V2	Depth Estimation	0.73 m	-
Cityscapes	Semantic Segmentation	-	82.4%
Custom Indoor Set	Combined Vision	0.65 m	79.8%

These results demonstrate the model's robustness across structured (Cityscapes) and unstructured environments. Improvements were particularly noticeable in low-light indoor scenes, suggesting strong generalization capability.

User Immersion and Interaction Quality

Qualitative evaluations through user testing (n=20) showed significant gains in user-perceived immersion, measured through Likert-scale surveys. Participants interacted with AR scenarios including virtual guides overlaid on real indoor objects, context-aware notifications, and educational holograms aligned with semantic surfaces. Results are shown in Table 3.

Table 3. User Feedback on Immersion and Interaction (5-point Likert Scale)

Evaluation Criteria	Baseline AR	Enhanced AR
Immersion Level	3.2	4.6
Visual Coherence with Environment	3.0	4.8
Responsiveness to Movement	4.2	4.1
Perceived Realism	2.9	4.5
Overall Satisfaction	3.4	4.7

Enhanced AR experiences led to a 35% increase in perceived realism and a 38% increase in satisfaction. Participants particularly appreciated object-awareness and digital content that reacted to real-world context.

4. Contextual Adaptability in Dynamic Environments

The system's semantic awareness allowed it to distinguish between objects and spatial zones. In field tests at a university hallway and a city plaza, object recognition retained 87% precision, with dynamic re-anchoring of holograms showing <300ms delay. Table 4 presents the results for object detection and re-anchoring.

Table 4. Object Detection and Re-anchoring Precision Across Environments

Environment	Object Detection Precision	AR Element Stability (ms drift)
Indoor (controlled)	91.2%	180 ms
Urban Outdoors	86.5%	295 ms
Semi-lit Corridor	82.8%	215 ms

These findings demonstrate that the integrated computer vision system offers reliable performance even in non-ideal real-world conditions, a significant leap from most academic prototypes tested only in lab environments.

Discussion

The results validate that integrating advanced computer vision techniques—particularly deep-learning-based segmentation and depth estimation—significantly enriches AR systems by adding layers of contextual awareness and spatial realism. Unlike prior studies which focused on single aspects such as SLAM or pose tracking (Adilzhan & Omarov, 2024), this study bridges these functionalities into a cohesive vision pipeline tailored for real-world mobile AR deployment. The ability to operate reliably in dynamically changing environments represents a substantial contribution to both academic research and industry practice in AR development.

Furthermore, the user immersion scores align with the theoretical premise that semantic understanding of the environment enhances cognitive engagement (Pflieger et al., 2024). The framework proposed in this study also addresses challenges highlighted in previous works regarding computational inefficiency (Amparore et al., 2024), offering a balance between performance and real-time responsiveness through optimization layers.

The integration of advanced computer vision techniques into augmented reality (AR) systems, as demonstrated in this study, marks a pivotal step toward overcoming long-standing limitations in contextual awareness and user immersion. Prior AR systems largely relied on SLAM (Simultaneous Localization and Mapping) and marker-based approaches, which, while effective for pose tracking, lacked semantic depth and environmental understanding (Theodorou et al., 2022). By incorporating real-time semantic segmentation and monocular depth estimation into the AR pipeline, this research advances the state-of-the-art in mobile AR design and deployment.

The performance gains, particularly in terms of accurate depth estimation and robust object recognition, align with recent efforts in real-time scene understanding (Lu, W., 2023; Shi, 2024). However, unlike many previous models tested in static or synthetic environments, our system was evaluated in real-world conditions, showing resilience across variable lighting, object clutter, and user movement. This is crucial, as highlighted by (Doskarayev et al., 2023), who emphasized the need for practical deployment of AI-enhanced AR in uncontrolled environments.

Importantly, the semantic segmentation module in our system enables AR content to be contextually relevant—anchoring holograms to floors, tables, or user-defined spatial zones with minimal drift. This addresses a critical gap in prior literature, where digital overlays often appear disconnected from their physical surroundings, leading to reduced user engagement and believability (Urquhart et al., 2024). Our findings—particularly the high immersion and satisfaction scores from user evaluations—support the hypothesis that contextual fidelity enhances cognitive presence and interaction quality in AR.

Moreover, this research contributes novel insight into balancing computational complexity with real-time responsiveness. The use of TensorRT optimization and model quantization allows the deployment of deep learning models on mobile devices without reliance on cloud processing, addressing concerns raised by (Hernández et al., 2024) regarding latency and privacy in cloud-based AR applications. While memory consumption increases with vision stack integration, the frame rates remain within acceptable thresholds (>25 FPS), demonstrating practical feasibility.

From a theoretical perspective, this study also aligns with situated cognition theories, which posit that meaningful learning occurs in context-rich environments (Bahar et al., 2025; Silalahi & Silalahi, 2023). By enabling AR systems to understand and respond to their environment in a semantically meaningful way, our approach enhances the potential of AR as a tool not just for entertainment, but also for education, industrial training, and assistive technologies. Another critical contribution of this research lies in the adaptability of the framework. The system showed consistent precision in object detection and hologram stability across both indoor and outdoor environments, unlike prior studies that were often constrained to controlled lab settings (Cheng et al., 2023). This opens possibilities for context-aware AR in smart cities, navigation, and public safety applications.

However, the success of such systems hinges not only on technical accuracy but also on user perception. Our qualitative data confirms that semantic accuracy translates into greater trust and immersion from users, validating previous assumptions by (Adilzhan & Omarov, 2024) regarding user acceptance of intelligent systems. By aligning computational output with human expectations of spatial logic and visual coherence, AR systems can better support naturalistic interaction.

This discussion also foregrounds the interdisciplinary nature of the problem. The integration of computer vision, cognitive science, human-computer interaction, and mobile systems requires a framework that is both modular and adaptive. Our work proposes such a blueprint, offering a generalizable architecture that can be tailored to domain-specific needs—from architectural visualization to STEM education. In conclusion, the results and analyses from this study not only validate the proposed system's technical robustness but also underscore its theoretical and practical implications. By addressing critical limitations in spatial awareness, responsiveness, and user engagement, this work contributes a scalable, real-world-ready solution to the evolution of AR experiences through advanced computer vision.

Limitations and Future Work

Despite the promising outcomes of this study, several limitations must be acknowledged. First, while the integrated AR system performed reliably across varied real-world environments, the evaluation was limited to Android-based mobile platforms with high-end GPU capabilities. Devices with lower processing power or different architectures (e.g., iOS or AR headsets) may not exhibit the same level of performance, indicating the need for further hardware-agnostic optimization.

Second, although the computer vision modules were evaluated using standard benchmark datasets such as Cityscapes and NYU Depth V2, there remains a domain gap between these datasets and unstructured real-world settings. Future work should consider training and evaluating models on more diverse, real-world AR-specific datasets to enhance generalization. Third, while user experience data was collected from a group of 20 participants, this sample size is relatively small and demographically limited. Larger-scale user studies across different age groups, professions, and cultural contexts are necessary to validate the system′s usability and acceptance more robustly. Fourth, the study focused primarily on visual perception and spatial interaction. However, AR experiences are inherently multisensory. Future extensions could integrate audio-based spatial cues, haptic feedback, or adaptive voice interfaces to enhance immersion further. Lastly, this study was designed with a single-user interaction paradigm. Multi-user collaborative AR scenarios, which require shared perception and synchronization, remain an open challenge and a potential direction for future research.

Conclusion

This study has demonstrated that the integration of advanced computer vision techniques—specifically semantic segmentation, depth estimation, and object detection—can significantly enhance the quality, immersion, and contextual relevance of augmented reality (AR) experiences. Through a design-based research methodology and rigorous performance evaluation in both simulated and real-world settings, the study confirms that a computer vision-driven AR framework can achieve real-time operation while maintaining high accuracy in semantic understanding. By bridging the gap between theoretical models of vision-based interaction and real-time mobile applications, the proposed system addresses several longstanding issues in AR usability, including digital object anchoring, environment-aware behavior, and user engagement.

However, while the results are promising, it is important to note some key limitations of this study. The sample size was relatively small (n=20), which may limit the generalizability of the findings. Additionally, the system's performance was tested primarily on high-end Android devices, which may not be representative of the broader range of mobile devices available. Furthermore, the user group was not very diverse in terms of age, professional background, and experience with AR technology, which may influence user perception of system performance and immersion. These factors should be considered when interpreting the results, as they could impact the system's applicability in more heterogeneous user populations or lower-end devices.

Despite these constraints, the study provides a reproducible framework that balances model complexity with deployment feasibility, which is crucial for wide-scale adoption in both consumer and enterprise settings. From a scholarly standpoint, the research contributes to the growing body of knowledge on cross-disciplinary integration in immersive technology design. Practically, it offers a template for developers and researchers seeking to build context-aware AR systems that are not only technically sound but also user-centric.

Future studies should focus on broadening hardware compatibility, expanding dataset diversity for training models, and exploring collaborative and multimodal AR interaction. Moreover, long-term field deployments across different sectors—such as education, retail, and healthcare—are necessary to validate the broader applicability and societal value of vision-enhanced AR systems. Additionally, it would be essential to explore how the system performs in more extreme environmental conditions, such as low light, high motion, and dynamic urban settings, and how it adapts to multi-user scenarios.

In conclusion, while this study demonstrates the potential of integrating computer vision into AR to improve both technical performance and user experience, further work is needed to refine the system for more diverse contexts and broader device compatibility. Acknowledging these challenges will provide a more realistic understanding of the study′s scope and the true extent of its achievements.

References

Adilzhan, A., & Omarov, N. (2024). Development of Augmented Reality Game Using Computer Vision Technology. *2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, 386–391.

Alsadik, B., & Karam, S. (2021). The simultaneous localization and mapping (SLAM): An overview. *Surveying and Geospatial Engineering Journal*, *1*(2), 1–12.

Amparore, D., Sica, M., Verri, P., Piramide, F., Checcucci, E., De Cillis, S., Piana, A., Campobasso, D., Burgio, M., & Cisero, E. (2024). Computer vision and machine-learning techniques for automatic 3D virtual images overlapping during augmented reality guided robotic partial nephrectomy. *Technology in Cancer Research & Treatment*, *23*, 15330338241229368.

Appavu, N. (2025). Analysing the Effect of Edge-Optimized Deep Learning Models on Improving Low-Powered Iot Devices Real-Time Object Detection. *2025 9th International Conference on Inventive Systems and Control (ICISC)*, 1663–1669.

Bahar, A. A., Fauzi, U. A., Al Anwar, S. F., Istama, N. R., Maulana, A. H., & Nargis, N. (2025). Youth slang as a reflection of social identity in teenage communities. *Variable Research Journal*, *2*(01), 186–194.

Bhowmik, A. K. (2024). Virtual and augmented reality: Human sensory-perceptual requirements and trends for immersive spatial computing experiences. *Journal of the Society for Information Display*, *32*(8), 605–646.

Cao, J., Lam, K.-Y., Lee, L.-H., Liu, X., Hui, P., & Su, X. (2023). Mobile augmented reality: User interfaces, frameworks, and intelligence. *ACM Computing Surveys*, *55*(9), 1–36.

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, *51*, 341–377.

Chen, X. (2023). Real-Time Semantic Segmentation Algorithms for Enhanced Augmented Reality. *Journal of Computational Innovation*, *3*(1).

Cheng, P., Xiong, Z., Bao, Y., Zhuang, P., Zhang, Y., Blasch, E., & Chen, G. (2023). A deep learning-enhanced multi-modal sensing platform for robust human object detection and tracking in challenging environments. *Electronics*, *12*(16), 3423.

Divya, B., Asha, R. M., Thara, P., & Prabha, R. (2023). An IoT Enabled Smart Home Automation Using Augmented Reality in Computer Vision. *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 1755–1761.

Doskarayev, B., Omarov, N., Omarov, B., Ismagulova, Z., Kozhamkulova, Z., Nurlybaeva, E., & Kasimova, G. (2023). Development of computer vision-enabled augmented reality games to increase motivation for sports. *International Journal of Advanced Computer Science and Applications*, *14*(4).

Ghasemi, Y., Jeong, H., Choi, S. H., Park, K.-B., & Lee, J. Y. (2022). Deep learning-based object detection in augmented reality: A systematic review. *Computers in Industry*, *139*, 103661.

Hernández, N., Almeida, F., & Blanco, V. (2024). Optimizing convolutional neural networks for IoT devices: performance and energy efficiency of quantization techniques. *Journal of Supercomputing*, *9*.

Jacobs, C., Foote, G., & Williams, M. (2023). Evaluating user experience with immersive technology in simulation-based education: A modified Delphi study with qualitative analysis. *Plos One*, *18*(8), e0275766.

Kaatrasalo, V. (2022). *Computer vision methods for augmented reality*.

Kavitha, M. G. (2024). Development of Augmented Reality Platform Using Image Processing with Deep Learning Techniques. *J. Electrical Systems*, *20*(3s), 249–256.

Lasfar, R., & Tóth, G. (2024). The difference of model robustness assessment using cross-validation and bootstrap methods. *Journal of Chemometrics*, *38*(6), e3530.

Lu, W. (2023). The Integration of Artificial Intelligence and Computer Vision Technologies in Virtual Reality and Augmented Reality Applications. *International Conference on Big Data Analytics for Cyber-Physical System in Smart City*, 79–87.

Omarov, N., Omarov, B., Baibaktina, A., Abilmazhinova, B., Abdimukhan, T., Doskarayev, B., & Adilzhan, A. (2023). Applying Artificial Intelligence and Computer Vision for Augmented Reality Game Development in Sports. *International Journal of Advanced Computer Science and Applications*, *14*(8).

Perera, P., Perera, S., Jin, X., Rashidi, M., Nanayakkara, S., Yazbek, G., & Yazbek, A. (2025). An Innovative Software Development Methodology for Deep Learning-Driven Visual Computing in Built Environment Applications. *Journal of Information Technology in Construction (ITcon)*, *30*(41), 1017–1040.

Pflieger, L. C. J., Hartmann, C., & Bannert, M. (2024). Enhancing knowledge construction in emerging technologies: The role of imagination training in immersive virtual reality environments. *Discover Education*, *3*(1), 65.

Rohan, A., Hasan, M. J., & Petrovski, A. (2025). A systematic literature review on deep learning-based depth estimation in computer vision. *ArXiv Preprint ArXiv:2501.05147*.

Shewail, A. S., Zayed, H. H., & Elsayed, N. A. M. (2024). Real-time indoor tracking for augmented reality using computer vision technique. *Int J Artif Intell*, *13*(2), 1845–1857.

Shi, R. (2024). Integrating computer vision and AI for interactive augmented reality experiences in new media. *Applied and Computational Engineering*, *102*, 49–54.

Silalahi, E. B., & Silalahi, N. (2023). Linguistics Realization Analysis on Slang Word; Social Media Whatsapp. *JETAL: Journal of English Teaching & Applied Linguistic*, *5*(1), 8–13.

Subrahmanyam, S. (2025). Experimental and Simulation-Based Research: Designing Controlled Experiments for Maritime Training and Safety. In *Research Methods for Advancing the Maritime Industry* (pp. 145–170). IGI Global Scientific Publishing.

Swami, S., Singh, P., & Singh, B. (2025). Revolutionizing E-Commerce: The Synergy of Computer Vision and Augmented Reality. In *Creating AI Synergy Through Business Technology Transformation* (pp. 115–136). IGI Global.

Syed, T. A., Siddiqui, M. S., Abdullah, H. B., Jan, S., Namoun, A., Alzahrani, A., Nadeem, A., & Alkhodre, A. B. (2022). In-depth review of augmented reality: Tracking technologies, development tools, AR displays, collaborative AR, and security concerns. *Sensors*, *23*(1), 146.

Theodorou, C., Velisavljevic, V., Dyo, V., & Nonyelu, F. (2022). Visual SLAM algorithms and their application for AR, mapping, localization and wayfinding. *Array*, *15*, 100222.

Urquhart, L., Petrakis, K., & Wodehouse, A. (2024). User engagement in physical-digital interaction design. *DS 131: Proceedings of the International Conference on Engineering and Product Design Education (E&PDE 2024)*, 187–192.

Wang, S., Zargar, S. A., & Yuan, F.-G. (2021). Augmented reality for enhanced visual inspection through knowledge-based deep learning. *Structural Health Monitoring*, *20*(1), 426–442.

Yang, L., Kumar, R., Kaur, R., Babbar, A., Makhanshahi, G. S., Singh, A., Kumar, R., Bhowmik, A., & Alawadi, A. H. (2024). Exploring the role of computer vision in product design and development: a comprehensive review. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *18*(6), 3633–3680.

Zhou, B., & Güven, S. (2020). Fine-grained visual recognition in mobile augmented reality for technical support. *IEEE Transactions on Visualization and Computer Graphics*, *26*(12), 3514–3523.