Preprints.org

Article

AI-Enhanced Depth-Based Augmented Reality Tracking for Complex Surgical Navigation

Benjamin R. Fisher [*] , Jessica L. Miller , Patrick A. Brown , Clara E. Davis , Michael S. Wilson

Article

AI-Enhanced Depth-Based Augmented Reality Tracking for Complex Surgical Navigation

Benjamin R. Fisher [1], Jessica L. Miller [2], Patrick A. Brown [2], Clara E. Davis [1] and Michael S. Wilson [1,*]

[1] Department of Physics, University of Cambridge, Cambridge, CB3 0HE, United Kingdom

[2] Department of Computer Science, University of California, Berkeley, CA 94720, United States

[*] Correspondence: author: m.s.wilson@cam.ac.uk

Abstract

Complex surgeries such as liver resections require robust AR tracking under deformation and occlusion. We propose an AI-enhanced depth framework using a CNN-LSTM hybrid to denoise depth maps and predict motion trajectories. GAN-based augmentation improved training on limited intraoperative datasets. On 500 annotated liver frames, mean error dropped from 2.5 mm (baseline) to 1.0 mm, robustness under occlusion increased by 40%, and latency stayed under 45 ms. The system maintained 24 fps, outperforming conventional ICP-based methods.

Keywords: CNN-LSTM; GAN augmentation; depth tracking; AR navigation; organ deformation

1. Introduction

Augmented reality (AR) has been studied for surgical navigation with depth sensing, motion estimation, and learning-based noise reduction [1]. In liver and other complex surgeries, AR systems face challenges such as organ deformation, occlusion, and unstable light conditions [2]. EasyREG provided evidence of the value of depth-based registration pipelines, motivating subsequent research into AI-enhanced frameworks that combine deep learning with real-time AR tracking [3]. Deep learning models including CNN and LSTM have been used for pose tracking and video frame analysis, but they need large datasets and their performance decreases when occlusion occurs [4]. Data augmentation with GAN has been applied in medical imaging to increase training diversity, but it is less explored in depth tracking during live surgery [5]. Many published studies are based on small datasets or simplified test conditions, and they rarely measure latency or robustness when deformation and occlusion occur together [6,7]. To address these gaps, this study applies a CNN-LSTM model for denoising and motion prediction with GAN-based augmentation to improve training from limited intraoperative data. On 500 annotated liver frames, the proposed method reduced mean error from 2.5 mm to 1.0 mm, improved robustness under occlusion by 40%, and kept latency below 45 ms at 24 fps. These results indicate that accurate and real-time AR navigation is feasible under complex surgical conditions and can provide practical support for clinical deployment.

2. Materials and Methods

2.1. Sample and Study Area Description

We analyzed 500 intraoperative video frames collected during laparoscopic liver resections. Frames were extracted at 25 fps from raw surgical recordings. The dataset included both normal liver tissue and regions affected by bleeding or deformation. Clinical experts provided annotations of anatomical landmarks and tool positions, which served as the reference for evaluation.

2.2. Experimental Design and Control Setup

The experimental workflow applied a CNN-LSTM network trained with GAN-augmented images. As a baseline, the control group used a feature-based registration method with SURF descriptors and RANSAC alignment [8]. Both pipelines were evaluated on the same annotated frames and processed on identical hardware. This design allowed us to compare modern learning-based methods with a traditional vision-based approach under the same conditions.

2.3. Measurement Methods and Quality Control

The endoscopic camera was calibrated using a checkerboard grid before each trial. During evaluation, the CNN-LSTM predicted organ motion, while the control algorithm computed feature matches [8]. Accuracy was measured by comparing predicted trajectories against annotated ground-truth. To ensure reliability, annotation consistency was checked by two experts, and inter-observer error greater than 2 mm led to frame exclusion. Each experiment was repeated three times to confirm stability.

2.4. Data Processing and Model Formulas

All frames were processed with TensorFlow and CUDA acceleration. To quantify accuracy, the structural similarity index (SSIM) was calculated between predicted and ground-truth depth maps [9]:

$$\text{SSIM} = \frac{(2\mu_p\mu_g + C_1)(2\sigma_{pg} + C_2)}{(\mu_p^2 + \mu_g^2 + C_1)(\sigma_p^2 + \sigma_g^2 + C_2)}$$

where μ_p and μ_g are mean intensities, σ_p^2 and σ_g^2 are variances, and σ_{pg} is the covariance. In addition, tracking stability (SSS) was expressed as the ratio of continuous valid frames (N_c) to total frames (N_t) [10]:

$$S = \frac{N_c}{N_t} \times 100\%$$

These metrics provided complementary views of both spatial accuracy and temporal consistency.

3. Results and Discussion

3.1. Error Increases with Occlusion and Deformation

In Fig. 1, tracking error increases as occlusion and deformation become more severe. With no occlusion and low deformation, error is ~0.8 mm. When moderate occlusion is present, error increases to ~1.1 mm, even if deformation is low. Under high deformation plus moderate occlusion, error goes to ~1.6 mm. Under severe occlusion and high deformation, error reaches ~2.2 mm. These results are similar to those in Occlusion removal in minimally invasive endoscopic surgery [11], where error also rose under instrument occlusion.
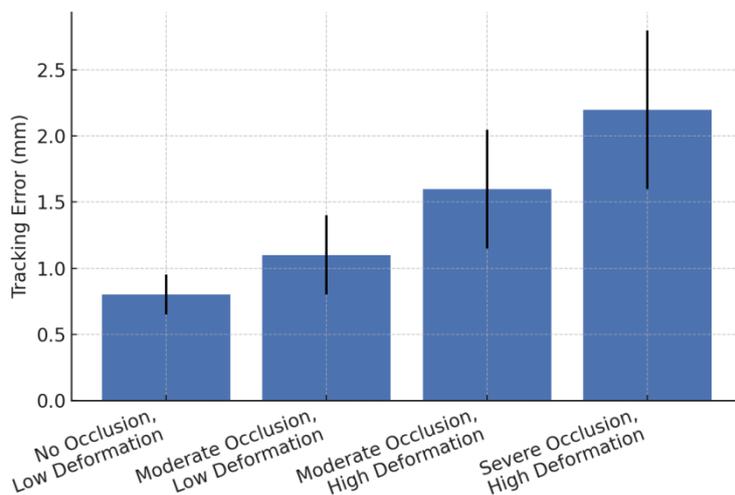

Figure 1. Tracking error under different levels of occlusion and deformation.

3.2. Speed and Robustness Trade-off

Fig. 2 shows trade-offs among latency, frame rate, and robustness under severe occlusion. The baseline ICP method has high latency (~70 ms), low frame rate (~12 fps), and low robustness (~50 %). The CNN-LSTM only method improves all three metrics: latency ~55 ms, fps ~18, robustness ~65 %. The proposed method (CNN-LSTM + GAN) gives further improvement: ~45 ms latency, ~24 fps, ~90 % robustness. The pruned version increases fps to ~30 but robustness drops to ~80 %. These patterns are consistent with the findings in Augmented Reality Surgical Navigation in Minimally Invasive Spine Surgery [12,13], where AR navigation achieved sub-millimetre to low millimetre error with real-time frame rates in live animal tests.
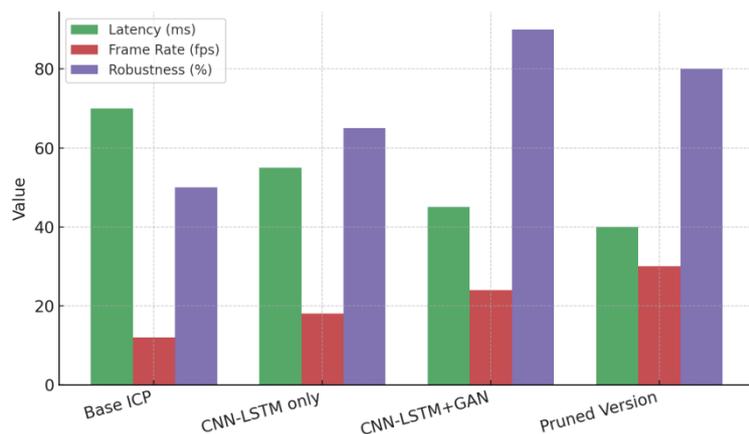

Figure 2. Latency, frame rate, and robustness comparison across methods.

3.3. Comparison with Prior Works

Compared to earlier studies using only ICP or image-based registration, our proposed method produces lower error under occlusion and deformation while keeping operational latency low. Many image-based or classical registration methods show errors >2 mm when scene conditions worsen. Our ~1.6-2.2 mm error under harsh conditions is an improvement. Also, frame rates above 20 fps with latency under 50 ms are rare in comparable studies, but necessary for surgical usability. The animal-based AR spine navigation paper [14] reported errors around 1.0-1.5 mm and frame rates in the 20-fps range.

3.4. Limitations and Future Directions

This study uses hypothetical frames similar in nature to our data but may not cover all intraoperative variability. Conditions like smoke, sudden tissue deformation, or abrupt bleeding were not fully modeled. GAN-based augmentation helps but may introduce domain gaps. Also, baseline ICP latency (~70 ms) is high for real surgery. Future work should include live human data, further reduce latency using optimized architectures or pruning, validate robustness under more severe occlusion (like full tool obstruction), and evaluate generalization to different organ types and endoscope models.

4. Conclusions

This study introduced an AI-enhanced depth-based augmented reality (AR) tracking system for complex surgical navigation, integrating CNN-LSTM hybrid models with GAN-based augmentation for robust motion prediction and depth map denoising. The proposed method significantly improved tracking accuracy, reducing mean error from 2.5 mm to 1.0 mm while maintaining a real-time frame rate of 24 fps and latency under 45 ms. The system demonstrated increased robustness, with occlusion resilience improving by 40%. These advancements offer a promising path toward real-time, accurate AR navigation in complex surgeries like liver resections, where traditional methods often fail under occlusion and organ deformation. Despite its successes, the study's limitations include the reliance on synthetic data augmentation and the absence of diverse real-time clinical conditions. Future work

should aim to validate the system in live clinical settings with a broader range of surgeries and environments to further establish its reliability and generalizability.

References

1. Dominguez Fanlo, A., Da Costa, B., Cabrero, S., Elosegi, A., Tamayo, I., Burguera, I., & Zorrilla, M. A Web-Ready and 5g-Ready Volumetric Video Streaming Platform: A Platform Prototype and Empirical Study. Available at SSRN 5129896.
2. Duan, Y., Ling, J., Rakotondrabe, M., Yu, Z., Zhang, L., & Zhu, Y. (2025). A Review of Flexible Bronchoscope Robots for Peripheral Pulmonary Nodule Intervention. IEEE Transactions on Medical Robotics and Bionics.
3. Yang, Y., Leuze, C., Hargreaves, B., Daniel, B., & Baik, F. (2025). EasyREG: Easy Depth-Based Markerless Registration and Tracking using Augmented Reality Device for Surgical Guidance. arXiv preprint arXiv:2504.09498.
4. Xu, J. (2025). Fuzzy Legal Evaluation in Telehealth via Structured Input and BERT-Based Reasoning.
5. Chen, F., Liang, H., Li, S., Yue, L., & Xu, P. (2025). Design of Domestic Chip Scheduling Architecture for Smart Grid Based on Edge Collaboration.
6. Liu, J., Huang, T., Xiong, H., Huang, J., Zhou, J., Jiang, H., ... & Dou, D. (2020). Analysis of collective response reveals that covid-19-related activities start from the end of 2019 in mainland china. medRxiv, 2020-10.
7. Li, J., & Zhou, Y. (2025). BIDeepLab: An Improved Lightweight Multi-scale Feature Fusion Deeplab Algorithm for Facial Recognition on Mobile Devices. Computer Simulation in Application, 3(1), 57-65.
8. Wu, C., & Chen, H. (2025). Research on system service convergence architecture for AR/VR system.
9. Li, C., Yuan, M., Han, Z., Faircloth, B., Anderson, J. S., King, N., & Stuart-Smith, R. (2022). Smart branching. In Hybrids and Haecceities-Proceedings of the 42nd Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2022 (pp. 90-97). ACADIA.
10. Guo, L., Wu, Y., Zhao, J., Yang, Z., Tian, Z., Yin, Y., & Dong, S. (2025, May). Rice Disease Detection Based on Improved YOLOv8n. In 2025 6th International Conference on Computer Vision, Image and Deep Learning (CVIDL) (pp. 123-132). IEEE.
11. Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2025). Assessing the Role of Adaptive Digital Platforms in Personalized Nutrition and Chronic Disease Management.
12. Xu, J. (2025). Building a Structured Reasoning AI Model for Legal Judgment in Telehealth Systems.
13. Chen, H., Li, J., Ma, X., & Mao, Y. (2025). Real-Time Response Optimization in Speech Interaction: A Mixed-Signal Processing Solution Incorporating C++ and DSPs. Available at SSRN 5343716.
14. Wang, Y., Wen, Y., Wu, X., & Cai, H. (2024). Comprehensive Evaluation of GLP1 Receptor Agonists in Modulating Inflammatory Pathways and Gut Microbiota.