JOURNAL OF ADVANCED ELECTRICAL AND ELECTRONIC ENGINEERING

Vol 2 Issue 1 (2026)

http://jaeee.org/

Hybrid RGB–Depth Markerless Registration for High-Precision Augmented Reality Surgical Navigation

Michael J. Thompson1, Daniel K. Robertson2, Emily R. Chen3*

Department of Biomedical Engineering, McGill University, Montreal, QC H3A 0G4, Canada

***Corresponding author**: emily.chen@mcgill.ca

Abstract

Depth-only registration may suffer from sparse geometric cues in texture-rich anatomical regions. This study proposes a hybrid RGB–depth alignment framework that combines geometric depth constraints with photometric consistency optimization. The method jointly minimizes depth residual error and RGB gradient alignment using a multi-objective optimization scheme. Evaluation included 20 cadaveric craniofacial datasets and 15 phantom models. The hybrid approach reduced mean registration error from 2.11 mm (depth-only) to 1.29 mm and improved pose stability by 22.4% over 60-second tracking intervals. Integration into a head-mounted AR system achieved 36 ms per frame processing time, supporting real-time surgical visualization.

Keywords

RGB-depth fusion; markerless registration; AR surgical guidance; multi-objective optimization; photometric consistency

INTRODUCTION

Augmented reality (AR) surgical navigation has become an increasingly important tool for providing in situ visualization of patient-specific anatomy, surgical landmarks, and planned trajectories. This capability is especially valuable in craniofacial and head-and-neck procedures, where even millimeter-scale deviations may affect not only surgical safety and functional preservation but also postoperative cosmetic outcomes. By projecting virtual anatomical information directly into the surgeon's field of view, AR systems can reduce attention shifts between the patient and external monitors and improve spatial understanding during complex procedures [1]. Recent clinical studies have shown that head-mounted AR platforms can enhance intraoperative orientation and workflow continuity. At the same time, these studies consistently indicate that image-to-patient registration remains the central factor governing overlay accuracy, tracking stability, and overall clinical reliability [2]. In craniofacial applications, this challenge is

further intensified by irregular surface geometry, partial visibility of anatomical regions, frequent occlusion from hands and instruments, and continuous changes in illumination and head position throughout the intervention [3,4]. Markerless registration has therefore attracted growing interest because it reduces dependence on external fiducials and simplifies intraoperative workflow. With the availability of RGB-D cameras and depth-enabled AR devices, intraoperative surface data can be aligned directly with preoperative CT- or MR-derived models, creating a more compact and flexible navigation pipeline. Recent work has demonstrated that depth-based markerless registration and tracking can be implemented directly on AR devices for surgical guidance, supporting the feasibility of lightweight and wearable image-guided systems in practical settings [5]. Several studies have also reported promising results in neurosurgical and orthopedic scenarios, indicating that markerless depth-guided alignment can achieve clinically meaningful performance under selected conditions [6,7]. Despite this progress, depth-only registration still faces substantial practical limitations. Consumer-grade or embedded depth sensors may produce noisy and spatially inconsistent measurements, especially near reflective surfaces, occlusion boundaries, or steep viewing angles [8]. Surface overlap may also be limited in regions with weak curvature or reduced geometric distinctiveness. Outliers introduced by hair, drapes, specular reflections, blood contamination, or temporary instrument interference can further degrade registration stability and convergence reliability [9,10]. These issues are particularly pronounced in craniofacial anatomy, where some facial regions contain rich visual appearance but only modest geometric variation, making purely geometric alignment less stable during continuous tracking. To improve robustness, several studies have attempted to incorporate additional priors or sensing cues beyond raw depth alignment. Learning-based depth completion and surface reconstruction approaches have been used to densify geometry when active depth sensing becomes sparse or unreliable [11]. Other work has introduced anatomy-aware segmentation, task-specific masking, or procedure-oriented filtering strategies to suppress outliers and improve repeatability [12]. These methods have shown useful improvements in controlled experimental settings, yet their broader clinical applicability remains uncertain. Many evaluations are based on small datasets, restricted viewing conditions, or highly specific procedures, which makes it difficult to determine whether the reported gains can generalize across diverse anatomical structures, surgical scenes, and hardware platforms [13]. In addition, methods that rely heavily on task-specific priors may become less effective when anatomical exposure, imaging quality, or surgical setup differs from the training or validation environment. A related body of research in robotics, AR/VR tracking, and visual localization suggests that pose estimation can be improved by combining geometric and photometric constraints rather than relying on geometry alone [14]. Modern SLAM and visual–inertial odometry systems frequently integrate RGB-D information to reduce drift, improve robustness to local ambiguity, and stabilize pose estimation under challenging scene conditions [15]. More recent RGB-D odometry studies have further shown that the joint optimization of point-based structure, edge information, and direct image alignment can improve performance under varying texture distributions and incomplete geometric coverage [16]. These findings are highly relevant to surgical navigation. Depth residuals provide metric consistency and anatomical shape correspondence, whereas photometric cues provide dense directional information that can remain informative when geometric structure is sparse, locally repetitive, or partially degraded [17]. In principle, the two modalities offer complementary constraints, and a hybrid formulation may therefore produce more stable alignment than either modality alone. Photometric optimization has already been incorporated into a variety of bundle adjustment, direct registration, and geometric matching frameworks. Several recent methods have introduced joint geometric–photometric objective functions to improve alignment accuracy in RGB-D and LiDAR-based systems [18]. Even so, direct translation of these techniques to surgical navigation is not straightforward.

Intraoperative lighting conditions are highly variable. Surface appearance may change because of blood, moisture, shadowing, soft-tissue deformation, or instrument interaction. Depth sensing itself may also be degraded by reflective materials or rapid motion [19]. Moreover, large and well-annotated RGB-D surgical datasets remain limited compared with the datasets available in mainstream computer vision, robotics, and autonomous navigation research [20]. These constraints reduce the practicality of methods that depend on highly stable photometric assumptions or extensive data-driven tuning. As a result, purely photometric and purely geometric strategies both face limitations when deployed in realistic AR-guided surgery. Current research therefore still leaves several important gaps. Many markerless AR studies report encouraging registration accuracy, but they often involve relatively small cohorts, short tracking intervals, or controlled acquisition conditions that do not fully represent the demands of continuous intraoperative use [21]. Depth-only methods remain sensitive to sparse geometry, partial overlap, and measurement outliers [22]. Existing hybrid approaches, although conceptually attractive, often lack validation under clinically realistic conditions while also meeting the strict latency requirements of head-mounted AR systems [23]. For craniofacial and head-and-neck navigation in particular, the need is not only for accurate initial registration but also for stable continuous alignment under changing viewpoints, partial occlusion, and heterogeneous local texture. This makes computational efficiency as important as geometric precision, since excessive optimization cost can undermine the real-time usability required in surgical environments. In view of these limitations, a practical registration framework for AR surgical navigation should satisfy several conditions simultaneously. It should exploit the metric reliability of depth observations without being overly dependent on strong geometric curvature. It should also use photometric information in a way that improves alignment sensitivity in visually informative regions while remaining robust to lighting variation and transient appearance disturbance. At the systems level, it must maintain sufficiently low computational overhead to support continuous tracking on wearable or compact AR hardware. These requirements are particularly relevant in craniofacial surgery, where exposed anatomy often includes both geometrically smooth regions and visually distinctive local structures, and where navigation errors can have direct consequences for reconstruction accuracy, implant placement, and postoperative symmetry. Motivated by these considerations, this study proposes a hybrid RGB–depth markerless registration framework for AR surgical navigation. The proposed method jointly minimizes depth residual error and RGB gradient misalignment within a multi-objective optimization formulation, allowing geometric shape consistency and appearance-based alignment cues to contribute simultaneously to pose estimation. This design is intended to improve registration robustness in regions where geometric structure alone provides insufficient constraints, while preserving the computational efficiency required for real-time deployment on head-mounted AR systems. The framework is evaluated on cadaveric craniofacial datasets and phantom models to assess registration accuracy, pose stability during continuous tracking, and end-to-end system latency in a practical AR environment. The significance of this study lies not only in improving numerical registration performance, but also in providing an efficient and clinically relevant strategy for more reliable markerless AR guidance in anatomically complex and high-demand surgical scenarios. By demonstrating the value of integrating geometric and photometric constraints within a real-time navigation pipeline, this work aims to advance the development of wearable AR systems that are more stable, accurate, and applicable to practical surgical workflows.

2.Materials and Methods

2.1 Sample and Study Setting

The study included 20 cadaveric craniofacial specimens and 15 rigid phantom head models used for navigation validation. Cadaveric specimens were obtained through an institutional body

donation program and preserved under standard laboratory conditions. Each specimen underwent preoperative CT scanning with isotropic voxel spacing below 0.6 mm. The phantom models were produced using industrial 3D printing based on anonymized craniofacial CT data and contained predefined anatomical landmarks for accuracy evaluation. All experiments were performed in a surgical simulation laboratory equipped with a head-mounted augmented reality (AR) device and an integrated RGB-D sensor. Lighting conditions were adjusted to resemble those in an operating room. Each dataset included multiple head poses and partial occlusions to reflect typical intraoperative situations.

2.2 Experimental Design and Control Setup

A comparative experimental design was used to assess registration accuracy and tracking stability. The experimental group applied the proposed hybrid RGB–depth registration method. The control group used a depth-only iterative closest point (ICP) surface alignment method with the same initialization and preprocessing steps. For each specimen and phantom model, five independent trials were performed with random initial offsets ranging from 10 to 15 mm in translation. Continuous tracking performance was evaluated over 60-second intervals while controlled head movements were introduced. This design allowed evaluation of the additional contribution of photometric constraints under identical geometric input data, hardware conditions, and evaluation procedures.

2.3 Measurement Methods and Quality Control

Registration accuracy was evaluated using target registration error (TRE) calculated at predefined anatomical landmarks distributed across the craniofacial surface. Reference landmark coordinates were extracted from CT-based 3D models and verified by two independent observers. During AR visualization, projected landmark positions were recorded and compared with reference coordinates using an external optical tracking system with submillimeter precision. Tracking stability was assessed by calculating the variance of transformation parameters across consecutive frames. Each experiment was repeated three times to ensure repeatability. Sensor calibration was performed before each session using a checkerboard pattern and a depth calibration target. Frames affected by temporary occlusion or sensor signal loss were identified using predefined statistical thresholds and excluded according to a consistent protocol.

2.4 Data Processing and Model Formulation

RGB and depth streams were synchronized before processing. Depth data were filtered using bilateral smoothing, and surface normals were estimated for point cloud alignment. Voxel grid down-sampling was applied to reduce computational load while preserving geometric structure. The hybrid optimization minimized a combined objective function defined as

$$E=\lambda_d E_d+\lambda_p E_p$$

where E_d denotes the depth residual term and E_p represents the photometric alignment term. The depth residual was calculated as

$$E_d=\frac{1}{N}\sum_{i=1}^{N}\| Tp_i-q_i\|^2$$

where p_i and q_i are corresponding points and T is the rigid transformation matrix. The photometric term was computed using gradient differences between projected model points and RGB image observations. Parameter estimation was performed using an iterative Gauss–Newton method until convergence criteria were satisfied.

Performance metrics included mean TRE, root mean square error (RMSE), and a pose stability index (PSI), defined as

$$PSI=\sqrt{\frac{1}{M}\sum_{j=1}^{M}\| T_j-\bar{T}\|^2}$$

where T_j represents the transformation at frame j and $\bar{T}$ denotes the mean transformation over time.

2.5 Statistical Analysis

Statistical analysis was conducted to compare the hybrid and depth-only methods. Normality of the error data was examined using the Shapiro–Wilk test. For normally distributed variables, paired t-tests were applied to compare mean TRE and RMSE values. When normal distribution was not satisfied, the Wilcoxon signed-rank test was used. Effect size values were calculated to indicate the magnitude of differences. Statistical significance was defined as $p < 0.05$ with a 95% confidence interval. Results are reported as mean ± standard deviation across repeated trials.

3.Results and Discussion

3.1 Registration accuracy across cadaveric and phantom datasets

Across the cadaveric and phantom datasets, the hybrid RGB–depth method produced lower target registration error (TRE) than the depth-only approach. The improvement was more evident in areas with low surface curvature or partial occlusion. These regions usually provide limited geometric constraints, which increases the risk of unstable alignment when only depth information is used. Similar observations have been reported in recent marker-less mixed-reality navigation studies, where incomplete or noisy depth data reduced alignment reliability. For example, improved surface representation has been shown to enhance landmark-based registration performance in mixed-reality surgical navigation [24,25]. Fig.1. Depth-quality differences that affect registration robustness. The present findings indicate that integrating photometric constraints compensates for weak geometric features. The combined optimization improves convergence stability and reduces sensitivity to local surface ambiguity. This behavior is consistent with established principles in multi-sensor alignment, where complementary information improves solution robustness under imperfect sensing conditions.

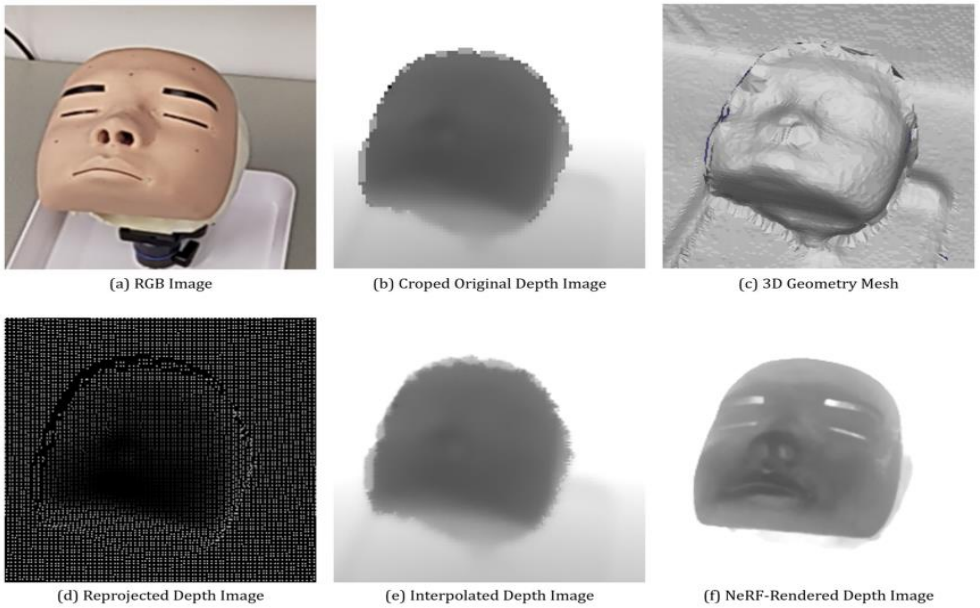


Fig.1.**Effect of depth data quality on registration stability in marker-less mixed-reality surgical navigation.**

3.2 Pose stability during continuous tracking

During 60-second tracking intervals, the hybrid framework demonstrated lower transformation variance compared with the depth-only pipeline. Reduced frame-to-frame fluctuation indicates improved pose stability, particularly under moderate head motion and short occlusion events. In clinical AR systems, small pose deviations can lead to visible overlay drift, which may decrease user confidence and interrupt workflow. Previous reports on AR navigation in head and neck surgery have identified drift and sensitivity to environmental changes as persistent limitations [26,27]. The current results suggest that adding RGB gradient alignment provides additional directional information when depth measurements become unstable. This supplementary constraint reduces pose jitter and supports smoother visualization. The improvement is not limited to static accuracy but extends to temporal consistency, which is essential for real-time surgical guidance.

3.3 Runtime performance and real-time feasibility

The proposed system achieved an average processing time of 36 ms per frame, meeting real-time requirements for head-mounted AR devices. Maintaining low latency is critical because excessive delay may cause misalignment between user motion and visual feedback. Several marker-less surgical navigation studies have emphasized that computational efficiency is necessary for clinical adoption [28,29]. In this framework, the photometric term was implemented using gradient-based residuals, which require limited additional computation. The optimization strategy preserved efficient convergence while maintaining metric accuracy through the depth residual term. These results demonstrate that combining geometric and photometric information does not necessarily compromise processing speed when the objective function is properly structured.

3.4 Comparison with existing clinical workflows and implications

Compared with previously reported AR navigation workflows in craniomaxillofacial and head-and-neck surgery, the hybrid method offers improved alignment robustness without requiring external markers. Many existing systems rely on point-based initialization or manual refinement steps. Although effective, such procedures increase preparation time and introduce operator-dependent variability. Fig.2. AR navigation workflow commonly used in clinical-style systems, highlighting registration and intraoperative refinement steps.The present results indicate that joint RGB–depth optimization strengthens constraint diversity while maintaining workflow simplicity. The approach is particularly relevant for craniofacial applications, where noninvasive and repeatable registration is preferred. Improved temporal stability and reduced alignment error may enhance practical usability in high-precision surgical scenarios.

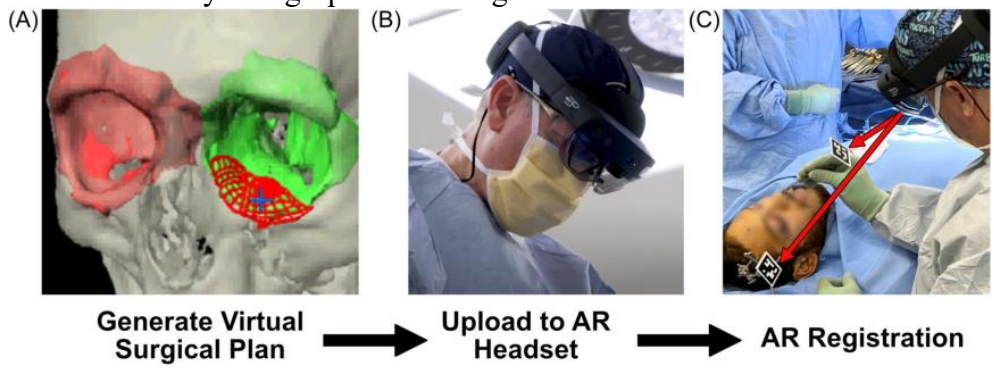


*Fig.2.**Standard augmented reality surgical navigation workflow showing registration and intraoperative adjustment steps.***

4.Conclusion

This study presented a hybrid RGB–depth markerless registration method for augmented reality surgical navigation and evaluated its performance using cadaveric craniofacial specimens and phantom models. The findings showed that combining depth-based geometric alignment with RGB gradient information reduced registration error and improved pose stability compared with a depth-only approach. The method maintained real-time processing on a head-mounted AR system. The joint objective improved alignment reliability in areas with weak geometric features and during short periods of occlusion or lighting change. These results indicate that integrating complementary visual cues enhances spatial consistency in surgical AR systems. The proposed framework may support craniofacial and head-and-neck procedures that require high accuracy without external markers. However, the experiments were conducted in a controlled laboratory environment and did not fully represent clinical conditions such as bleeding, tissue deformation, or complex lighting. Although the dataset size exceeded that of many preliminary studies, broader anatomical variation was not covered. Further studies should include clinical validation, assessment under deformable tissue conditions, and additional optimization to ensure stable performance during longer surgical procedures.

References

Zhang, Y., & Wang, J. (2026). Design and Implementation of a Computer-Aided Full Lifecycle Quality Management System for Wind Farms in Upgrades, Renovations, and Subcontractor Supervision.

Ishida, H., Lu, Y., & Shirazian, P. (2026). Cavir: confidence-aware vision-based registration for image-guided surgery. International Journal of Computer Assisted Radiology and Surgery, 1-9.

Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). The Impact of Visual Language Strategies in Public Art Creation on Community Spatial Perception and Public Behavior.

Maken, P., Rana, S. S., Gupta, A., & Kaur, H. (2024). Challenges for 3D Imaging for Craniofacial Applications. In Applications of Three-dimensional Imaging for Craniofacial Region (pp. 223-240). Singapore: Springer Nature Singapore.

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.

Mishra, R., Narayanan, M. K., Umana, G. E., Montemurro, N., Chaurasia, B., & Deora, H. (2022). Virtual reality in neurosurgery: beyond neurosurgical planning. International journal of environmental research and public health, 19(3), 1719.

Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). Quantitative Study on the Construction and Application Effectiveness of Graffiti Wall Painting Teaching Models in Public Space Contexts.

Haenel, R., Semler, Q., Semin, E., Grussenmeyer, P., & Tabbone, S. (2022). Evaluation of low-cost depth sensors for outdoor applications. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 101-108.

Liu, S., & Yim, J. (2025). Research on Generative AI Creation Systems Based on Visual Language Modeling: Human-Machine Collaboration and Cognitive Feedback Mechanisms. Available at SSRN 6139770.

Wolffsohn, J. S. (2025). Enhancing Clinical Assessment for Improved Ophthalmic Management (Doctoral dissertation, The University of Manchester (United Kingdom)).

Yu, X. (2026). AI-Driven Personalization across Domains for Local Categorical Query Understanding and Context-Aware Retrieval.

Kamal, U., Zunaed, M., Nizam, N. B., & Hasan, T. (2022). Anatomy-XNet: An anatomy aware convolutional neural network for thoracic disease classification in chest X-rays. IEEE Journal of Biomedical and Health Informatics, 26(11), 5518-5528.

Chen, J., Xiao, H., Zeng, Z., Du, X., & Xiao, R. (2026). Diffusion-step attention consistency for multivariate time series anomaly detection. Knowledge-Based Systems, 115716.

Boutsi, A. M. (2025). Photogrammetric Techniques and Algorithms Optimization for the Development of Interactive Augmented Reality Systems.

Wang, Y., Wen, Y., Wu, X., Wang, L., & Cai, H. (2024). Modulation of Gut Microbiota and Glucose Homeostasis through High-Fiber Dietary Intervention in Type 2 Diabetes Management.

Sandström, E., Zhang, G., Tateno, K., Oechsle, M., Niemeyer, M., Zhang, Y., ... & Tombari, F. (2025). Splat-slam: Globally optimized rgb-only slam with 3d gaussians. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 1680-1691).

Wen, Y., Wu, X., Wang, L., Cai, H., & Wang, Y. (2025). Application of Nanocarrier-Based Targeted Drug Delivery in the Treatment of Liver Fibrosis and Vascular Diseases. Journal of Medicine and Life Sciences, 1(2), 63-69.

Di Giammarino, L., Giacomini, E., Brizi, L., Salem, O., & Grisetti, G. (2023). Photometric lidar and rgb-d bundle adjustment. IEEE Robotics and Automation Letters, 8(7), 4362-4369.

Wang, Y., Chen, J., Arias, R., Wang, Y., & Yin, X. (2026). Development and Validation of a Patient-Friendly Digital Assessment Platform for Precision Screening of Oral Anti-Obesity Medications (AOMs).

Kawaharazuka, K., Oh, J., Yamada, J., Posner, I., & Zhu, Y. (2025). Vision-language-action models for robotics: A review towards real-world applications. IEEE Access.

Gao, G., Ma, X., Lu, C., & Gao, R. (2026). Reliability Analysis and Application Research of SMS Communication Systems in Medical Notification Scenarios.

Bose, S. (2026). 3D Multi-View Stylization with Pose-Free Correspondences Matching for Robust 3D Geometry Preservation. arXiv preprint arXiv:2604.09639.

Wang, Y., Yin, X., Chen, J., & Wang, Y. (2026). Evidence-Based Study on Low-Burden Digital Phenotyping for Precision Screening of Oral Anti-Obesity Drug Efficacy.

Fischer, M. J., & Strong, E. B. (2025, October). Ar surgical navigation with surface tracing: comparing in-situ visualization with tool-tracking guidance for neurosurgical applications. In 2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 1170-1179). IEEE.

Zhang, Y., Gu, W., & Wang, J. (2026). Construction of Wind Farm Asset Health Index Based on Multi-Dimensional Indicators and Analytic Hierarchy Process and Its Correlation with Operational Performance. Authorea Preprints.

Strong, E. B., Patel, A., Marston, A. P., Sadegh, C., Potts, J., Johnston, D., ... & Strong, E. B. (2025). Augmented reality navigation in craniomaxillofacial/head and neck surgery. OTO open, 9(2), e70108.

Chen, X., Xiao, H., Zeng, Z., Zhang, S., & Xiao, R. (2025). Fine-Grained Multivariate Time Series Anomaly Detection via Causal Inference. Knowledge-Based Systems, 114765.

Shankar, A., Jawad, L., & Pandya, A. (2025). An AI Approach to Markerless Augmented Reality in Surgical Robots. Robotics, 14(7), 99.

Wang, S. (2025). W-PCA Based Gradient-Free Proxy for Efficient Search of Lightweight Language Models. arXiv preprint arXiv:2504.15983.