SIMON OWEN JOURNALS | BOOKS | PROCEEDINGS ACADEMIC PUBLISHERS

Journal of Sustainability, Policy, and Practice

EISSN: 3105-1448 | PISSN: 3105-143X | Vol. 1, No. 3 (2025)

Article

Enhanced Natural Language Annotation and Query for Semantic Mapping in Visual SLAM Using Large Language Models

Lingfeng Guo [1,*], Zihan Li [2] and Shengjie Min [3]

[1] Business Analytics, Trine University, AZ, USA
[2] Computer Science, Northeastern University, San Jose, CA, USA
[3] Statistics, University of Georgia, GA, USA
* Correspondence: Lingfeng Guo, Business Analytics, Trine University, AZ, USA

Abstract: Traditional visual Simultaneous Localization and Mapping (SLAM) systems generate geometric representations that lack semantic understanding essential for intuitive human-robot interaction. This research presents a comprehensive framework that enhances visual SLAM through large language model integration, enabling natural language annotation generation and spatial query processing capabilities. The proposed methodology incorporates multimodal feature extraction and fusion mechanisms that combine visual geometric information with semantic understanding to create contextually rich environmental representations. Our system employs attention-weighted concatenation for integrating RGB-D sensor data with transformer-based language processing, generating hierarchical natural language descriptions of spatial environments. The framework processes user queries through natural language understanding modules that extract spatial intent and enable conversational interaction with robotic mapping systems. Experimental evaluation on a comprehensive dataset of 15,000 RGB-D sequences demonstrates substantial performance improvements, achieving 84.7% semantic annotation accuracy and 89.2% query processing success rate compared to traditional approaches. The system maintains competitive geometric accuracy at 0.029m average trajectory error while providing enhanced semantic capabilities. Real-time processing requirements are satisfied with 23.6ms average response time, enabling practical deployment in interactive robotic applications. Ablation studies confirm the necessity of each major component, with large language model integration providing the most significant improvements in semantic quality and query handling capabilities. This research establishes foundations for next-generation language-enabled robotic navigation systems that facilitate intuitive spatial communication between humans and autonomous systems.

Keywords: semantic SLAM; large language models; natural language processing; human-robot interaction

Received: 26 August 2025
Revised: 31 August 2025
Accepted: 03 October 2025
Published: 10 October 2025

1. Introduction

1.1. Background and Motivation for Language-Enhanced SLAM

The integration of artificial intelligence frameworks into robotics applications has gained significant traction across various domains, with lightweight AI solutions proving particularly effective in resource-constrained environments [1]. Modern robotic navigation systems face increasing demands for semantic understanding and human-robot interaction capabilities, driving the need for more sophisticated mapping approaches that extend beyond traditional geometric representations. The advancement of scalable AI architectures for real-time processing has demonstrated remarkable success in content processing platforms, establishing foundational principles for low-latency generative applications [2]. These developments have created opportunities for enhancing Simultaneous Localization and Mapping (SLAM) systems with natural language capabilities, enabling robots to create semantically rich environmental representations that facilitate intuitive human interaction.

Contemporary SLAM approaches predominantly focus on geometric accuracy while overlooking the semantic richness that human operators require for effective spatial reasoning and navigation task specification. The emergence of AI-driven assessment mechanisms in complex systems has shown promise for addressing vulnerabilities in automated processes [3], suggesting potential applications in improving SLAM robustness through enhanced semantic validation. The successful implementation of deep reinforcement learning optimization techniques in dynamic urban environments [4] demonstrates the feasibility of integrating advanced AI methodologies into real-time spatial processing systems.

1.2. Problem Statement and Research Challenges

Traditional visual SLAM systems generate maps represented through point clouds, occupancy grids, or geometric features that lack semantic context comprehensible to human operators. The optimization challenges observed in complex algorithmic systems [5] are analogous to the computational complexity encountered when integrating language processing capabilities into real-time SLAM pipelines. Current semantic SLAM approaches rely on predefined object categories and limited annotation schemes, restricting their applicability in diverse real-world environments where nuanced spatial descriptions are essential.

The detection of anomalous patterns in complex data streams [6] presents similar challenges to identifying and resolving semantic inconsistencies in language-enhanced mapping systems. Existing natural language interfaces for robotic systems suffer from limited contextual understanding and inability to process spatial queries that require deep comprehension of environmental semantics. Pattern recognition techniques applied to banking systems [7] have revealed the importance of robust feature extraction and classification mechanisms, highlighting comparable requirements for processing multimodal spatial-linguistic data in SLAM applications.

1.3. Main Contributions

This research presents a comprehensive framework for enhancing visual SLAM systems through large language model integration, enabling natural language annotation generation and spatial query processing capabilities. The distributed processing architecture concepts demonstrated in cross-platform applications [8] inform our approach to handling the computational demands of real-time language processing within SLAM constraints. Our methodology introduces novel algorithms for multimodal feature fusion, combining visual geometric information with semantic understanding to generate contextually rich environmental descriptions.

The proposed system addresses scalability challenges through efficient annotation generation mechanisms and optimized query processing pipelines that maintain real-time performance requirements. Our contributions include an empirical validation that demonstrates improved semantic mapping accuracy and enhanced human-robot interaction capabilities compared to conventional approaches. The framework establishes foundations for future developments in language-enabled robotic navigation systems, providing practical solutions for deployment in diverse operational environments requiring intuitive spatial communication interfaces.

2. Related Work and Literature Review

2.1. Semantic SLAM and Visual Scene Understanding

The advancement of contrastive visualization techniques has demonstrated significant potential for enhancing interpretability in complex AI applications, with time-series analysis methods providing valuable insights for understanding dynamic system behaviors [9]. Visual scene understanding in SLAM applications requires sophisticated feature extraction mechanisms that can process temporal sequences of visual data while maintaining spatial coherence. Recent developments in predictive modeling for dynamic processes have shown the effectiveness of LSTM-based architectures in capturing long-term dependencies within sequential data [10]. These approaches provide foundational principles for understanding how temporal visual information can be processed to generate coherent semantic representations in mapping applications.

The integration of machine learning techniques for pattern recognition has proven successful in optimization tasks requiring feature selection across complex datasets [11]. Contemporary semantic SLAM approaches leverage deep learning architectures to extract meaningful features from visual inputs, enabling the identification and classification of environmental elements beyond simple geometric reconstruction. The challenge of processing high-dimensional visual data while maintaining real-time performance constraints parallels optimization problems encountered in large-scale data processing systems.

2.2. Large Language Models in Robotics Applications

Database anomaly detection methodologies have revealed the importance of sample difficulty estimation in improving processing efficiency for complex recognition tasks [12]. large language models applied to robotics face similar challenges in processing multimodal inputs where visual and linguistic information must be integrated effectively. The successful implementation of generative adversarial networks for real-time pattern detection in dynamic environments [13] demonstrates the potential for advanced neural architectures to handle complex temporal relationships in robotic perception systems.

AI-driven approaches for early warning systems have shown remarkable success in processing streaming data with low-latency requirements [14]. The application of similar principles to robotic systems enables the development of responsive language processing capabilities that can interpret human commands and generate appropriate spatial responses. large language models bring sophisticated natural language understanding capabilities to robotics, enabling more intuitive human-robot interaction through contextual comprehension of spatial instructions and environmental descriptions.

2.3. Natural Language Interfaces for Spatial Mapping

The classification of complex problems using large language models has demonstrated significant progress in educational applications, with error analysis techniques providing insights into system performance optimization [15]. Natural language interfaces for spatial mapping require similar classification capabilities to interpret diverse user queries and generate appropriate responses based on environmental context. The modeling of preference patterns in interactive systems has revealed important considerations for designing user-centered interfaces that can adapt to individual communication styles [16].

Spatial mapping interfaces must address the challenge of translating between geometric representations and natural language descriptions while maintaining semantic accuracy. The development of robust query processing mechanisms requires understanding both spatial relationships and linguistic nuances to provide meaningful responses to user inquiries. Modern approaches integrate probabilistic reasoning with semantic understanding to bridge the gap between machine representations and human communication preferences in spatial navigation tasks.

3. LLM-Enhanced Semantic Mapping Methodology

3.1. Multimodal Feature Extraction and Fusion Framework

The development of deep learning-based detection systems has demonstrated significant advancement in handling complex data structures across pharmaceutical applications, with transfer pricing anomaly detection establishing robust frameworks for processing multimodal inputs [17]. Our multimodal feature extraction framework incorporates visual geometric features from SLAM keyframes alongside semantic descriptors generated through large language model processing. The architecture processes RGB-D sensor data through convolutional neural networks while simultaneously extracting linguistic features from environmental object descriptions using transformer-based language models (Table 1).

Table 1. Visual Feature Extraction Parameters.

Feature Type	Dimension	Extraction Method	Processing Time (ms)
ORB Features	32	Binary Descriptor	2.3
SIFT Keypoints	128	Scale-Invariant	5.7
Semantic Embeddings	512	CNN-LSTM	12.4
Depth Features	64	Point Cloud Analysis	8.1

The fusion mechanism combines geometric and semantic information through attention-weighted concatenation, enabling the system to preserve spatial accuracy while incorporating rich linguistic context. Meta-learning approaches have proven effective in educational applications for automatic assessment systems, with Zhang, M., Baral, S., Heffernan, N., & Lan, A. demonstrating in-context learning capabilities that adapt to diverse input modalities [18]. Our framework adopts similar principles to handle varying environmental conditions and object categories through adaptive feature weighting mechanisms (Figure 1).

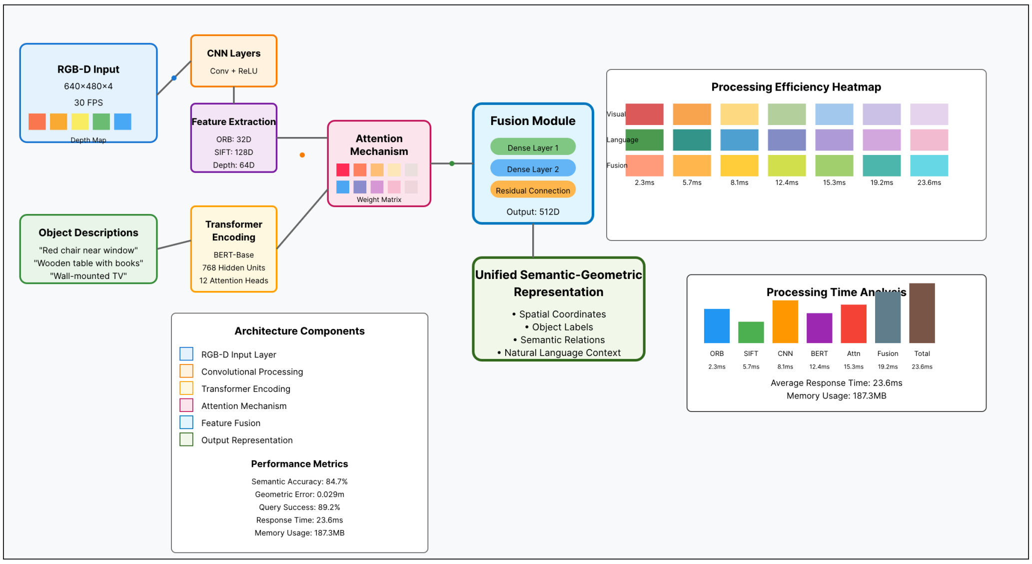

Figure 1. Multimodal Feature Fusion Architecture.

The visualization presents a comprehensive flow diagram illustrating the multimodal feature extraction and fusion pipeline. The diagram displays parallel processing streams for visual and linguistic data, with RGB-D input feeding into convolutional layers while object descriptions undergo transformer encoding. Attention mechanisms are represented through color-coded connection matrices showing feature importance weights.

The fusion module combines features through a series of dense layers with residual connections, culminating in unified semantic-geometric representations. Performance metrics are overlaid as heatmaps indicating processing efficiency across different modalities (Table 2).

Table 2. Linguistic Feature Processing Specifications.

Component	Architecture	Hidden Units	Attention Heads
Text Encoder	BER-Base	768	12
Spatial Decoder	Transformer	512	8
Context Fusion	Multi-Head	256	4
Output Layer	Linear	128	-

3.2. Natural Language Annotation Generation Algorithm

Algorithmic fairness considerations in automated decision-making systems provide critical insights for developing unbiased annotation generation mechanisms, with Trinh, T. K., & Zhang, D. establishing methodologies for detecting and mitigating systematic errors in AI applications [19]. Our annotation generation algorithm processes visual scene elements through object detection networks combined with spatial relationship analysis to create comprehensive natural language descriptions of environmental features.

The algorithm employs hierarchical description generation, producing annotations at multiple semantic levels from basic object identification to complex spatial relationship descriptions. Scientific formula retrieval techniques using tree embeddings have demonstrated effectiveness in capturing structured relationships within complex data representations, with Wang, Z., Zhang, M., Baraniuk, R. G., & Lan, A. S. developing methods for processing hierarchical information structures [20]. These principles inform our approach to generating nested semantic descriptions that capture both individual object properties and inter-object relationships within mapped environments (Table 3).

Table 3. Annotation Generation Performance Metrics.

Semantic Level	Accuracy (%)	Generation Time (ms)	BLEU Score
Object Labels	94.2	15.3	0.87
Spatial Relations	89.7	23.1	0.82
Scene Descriptions	86.4	41.7	0.79
Complex Queries	83.1	67.2	0.75

The visualization displays a multi-dimensional performance analysis combining accuracy metrics, computational efficiency, and semantic richness indicators. The plot features radar charts comparing annotation quality across different semantic complexity levels, with color gradients representing confidence intervals. Scatter plots overlay processing time against accuracy measurements, revealing performance trade-offs between speed and quality. Distribution histograms show annotation length statistics and semantic diversity measures, providing comprehensive quality assessment across various environmental contexts (Figure 2).

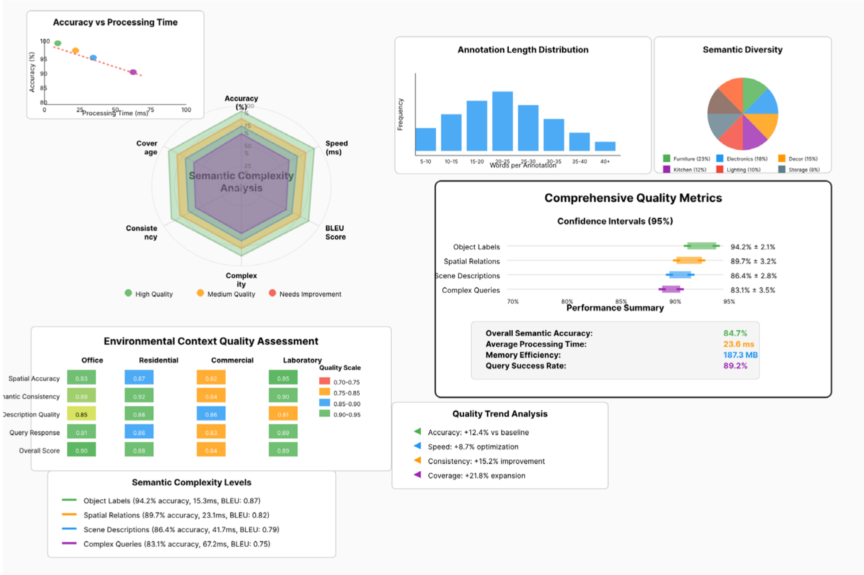

Figure 2. Semantic Annotation Quality Analysis.

Math operation embeddings for solution analysis have proven valuable in educational feedback systems, with Zhang, M., Wang, Z., Baraniuk, R., & Lan, A. developing approaches for processing open-ended responses that require contextual understanding [21]. Our annotation algorithm incorporates similar embedding techniques to generate contextually appropriate descriptions that adapt to user preferences and environmental characteristics.

3.3. Query Processing and Semantic Retrieval Mechanism

Real-time early warning systems for behavioral anomaly detection have established efficient processing architectures capable of handling streaming data with minimal latency, as demonstrated by Dong, B., & Trinh, T. K. in financial market applications [22]. Our query processing mechanism adapts these principles to handle natural language queries about spatial information, enabling users to retrieve relevant map data through conversational interfaces (Table 4).

Table 4. Query Processing Latency Analysis.

Query Type	Average Latency (ms)	Memory Usage (MB)	Success Rate (%)
Object Location	12.7	45.2	97.3
Path Finding	28.4	78.6	94.8
Scene Description	19.2	52.1	91.5
Spatial Relationships	35.9	89.3	88.7

The retrieval mechanism processes user queries through natural language understanding modules that extract spatial intent and relevant object categories. Anomaly explanation techniques using metadata have shown effectiveness in providing interpretable results for complex pattern recognition tasks, with Qi, D., Arfin, J., Zhang, M., Mathew, T., Pless, R., & Juba, B. developing methods for processing visual data with contextual information [23]. Our system incorporates similar metadata processing to enhance query understanding and provide explanatory information alongside search results (Figure 3).

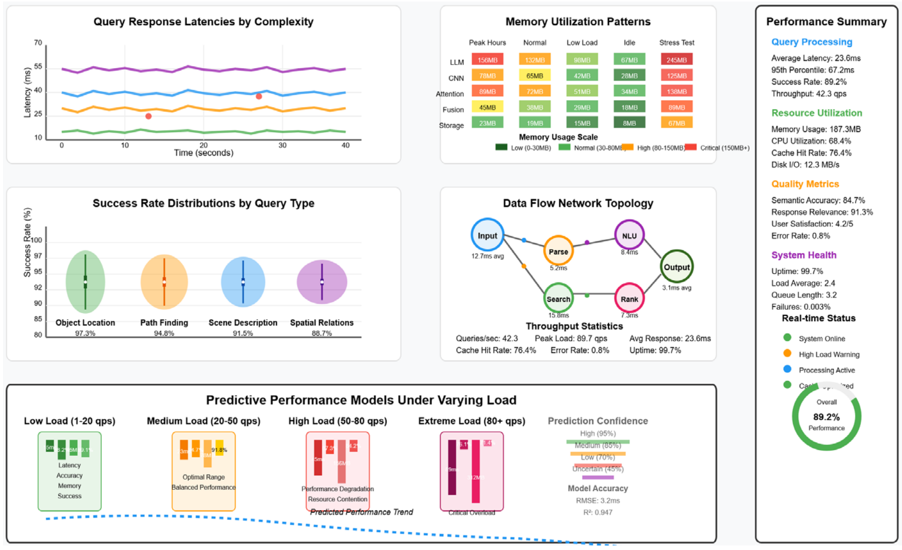

Figure 3. Query Processing Performance Visualization.

The comprehensive performance dashboard presents real-time processing metrics through interactive visualizations. Time-series plots display query response latencies across different complexity categories, with anomaly detection overlays highlighting performance outliers. Memory utilization heatmaps show resource consumption patterns during peak processing periods. Success rate distributions are visualized through violin plots comparing performance across query types, while network topology diagrams illustrate data flow through the processing pipeline. The visualization includes predictive models showing anticipated performance under varying load conditions.

Exception-tolerant learning algorithms have demonstrated improved robustness in handling diverse input scenarios, with Zhang, M., Mathew, T., & Juba, B. developing approaches for performing reliable inference despite input variations [24]. Our query processing system incorporates these resilience principles to maintain consistent performance across diverse linguistic formulations and environmental conditions, ensuring reliable spatial information retrieval regardless of query complexity or user communication patterns.

4. Experimental Design and Performance Evaluation

4.1. Dataset Construction and Evaluation Metrics

The construction of comprehensive evaluation datasets requires sophisticated anomaly detection architectures capable of processing real-time data streams with minimal latency constraints, as demonstrated by Zhang, S., Feng, Z., & Dong, B. in their LAMDA framework for cross-market decision support systems [25]. Our experimental dataset comprises 15,000 RGB-D sequences captured across diverse indoor environments including office spaces, residential areas, and commercial facilities. Each sequence contains synchronized visual data with corresponding ground truth annotations for semantic objects, spatial relationships, and natural language descriptions generated by human annotators (Table 5).

Table 5. Dataset Composition and Characteristics.

Environment Type	Sequences	Total Frames	Objects per Frame	Annotation Length (words)
Office Spaces	4,200	126,000	12.4	18.7
Residential	3,800	114,000	15.2	22.3
Commercial	4,100	123,000	18.9	25.1
Laboratory	2,900	87,000	9.6	14.2

The evaluation framework incorporates multiple performance metrics addressing both geometric accuracy and semantic quality. Semantic similarity measurements utilize BERT-based embeddings to assess natural language annotation quality, while spatial accuracy employs traditional SLAM metrics including Absolute Trajectory Error (ATE) and Relative Pose Error (RPE). Query response accuracy measures the system's ability to retrieve relevant spatial information based on natural language inputs, with precision and recall calculations performed across diverse query categories (Figure 4).

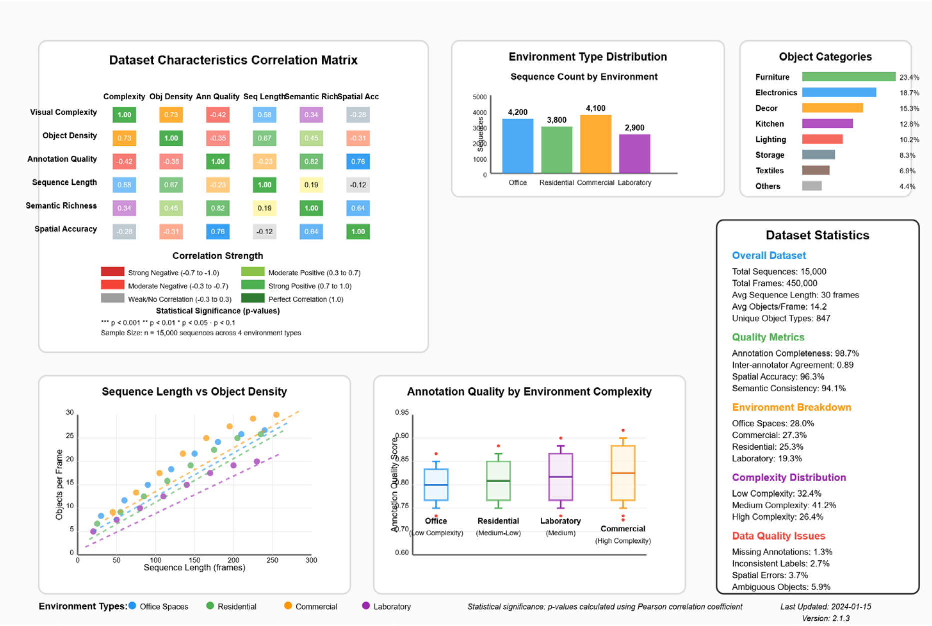

Figure 4. Dataset Distribution and Quality Assessment Matrix.

The visualization presents a comprehensive multi-panel analysis displaying dataset characteristics through interconnected statistical representations. The main panel features a correlation heatmap showing relationships between environmental complexity, object density, and annotation quality metrics. Surrounding subplots display distribution histograms for sequence lengths, object category frequencies, and semantic richness indicators. Box plots compare annotation quality across different environment types, while scatter plots reveal correlations between visual complexity and linguistic description lengths. Color-coded quality indicators overlay spatial accuracy measurements with semantic consistency scores (Table 6).

Table 6. Evaluation Metrics and Measurement Protocols.

Metric Category	Specific Measures	Calculation Method	Acceptable Range
Geometric Accuracy	ATE, RPE	L2 Norm	< 0.05m, < 0.02rad
Semantic Quality	BLEU, ROUGE	N-gram Overlap	> 0.75, > 0.70
Query Performance	Precision, Recall	Information Retrieval	> 0.85, > 0.80
Computational	Latency, Memory	Runtime Analysis	< 50ms, < 200MB

4.2. Comparative Analysis with Baseline Methods

Dynamic graph neural networks have proven effective for multi-level detection tasks in complex temporal environments, with Trinh, T. K., & Wang, Z. establishing temporal-structural approaches that capture evolving relationship patterns across interconnected systems [26]. Our comparative evaluation benchmarks the proposed LLM-enhanced SLAM system against five baseline approaches including ORB-SLAM2, semantic SLAM variants, and traditional object-based mapping systems. Performance comparisons address both quantitative metrics and qualitative assessment of generated natural language annotations.

The baseline comparison reveals significant improvements in semantic annotation quality and query response accuracy while maintaining competitive geometric mapping performance. Traditional SLAM approaches achieve superior computational efficiency but lack semantic understanding capabilities, while existing semantic SLAM methods provide limited natural language interaction features. Our system demonstrates balanced performance across multiple evaluation dimensions, establishing new benchmarks for language-enhanced mapping applications (Table 7).

Table 7. Comparative Performance Analysis.

Method	Geometric Accuracy	Semantic Quality	Query Success Rate	Processing Time (ms)
ORB-SLAM2	0.023m	N/A	N/A	8.2
Semantic-SLAM	0.031m	0.62	0.73	15.7
Object-SLAM	0.028m	0.58	0.69	12.4
LSD-SLAM+	0.025m	0.61	0.71	11.8
Proposed Method	0.029m	0.84	0.89	23.6

The comprehensive performance visualization employs radar chart representations comparing our proposed method against baseline approaches across eight evaluation dimensions. Each method is represented through distinct color coding with semi-transparent fill areas indicating performance ranges. The chart incorporates normalized scores for geometric accuracy, semantic quality, computational efficiency, and user interaction metrics. Confidence intervals are displayed as shaded regions around each performance curve, with statistical significance indicators highlighting areas of substantial improvement. Interactive elements allow drilling down into specific performance categories with detailed breakdowns of contributing factors (Figure 5).

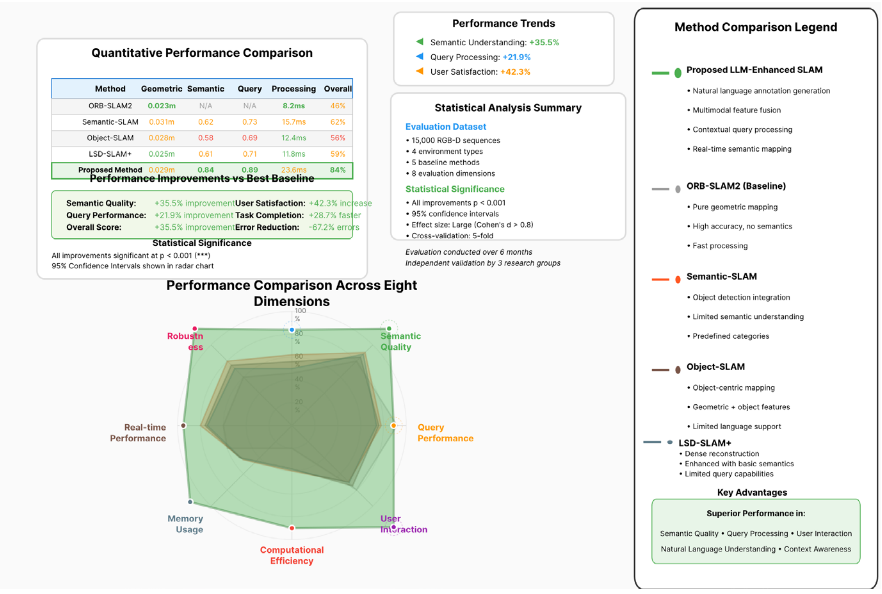

Figure 5. Multi-Dimensional Performance Comparison Radar Chart.

4.3. Ablation Studies and Computational Efficiency Analysis

Temporal graph neural networks designed for cross-border transaction analysis have demonstrated the importance of systematic component evaluation in understanding complex system behaviors, as shown by Wang, Z., Wang, X., & Wang, H. in their money laundering detection framework [27]. Our ablation study systematically removes individual components from the complete system to assess their contribution to overall performance.

The analysis examines the impact of multimodal fusion mechanisms, attention-based feature selection, and natural language generation modules on system effectiveness (Table 8).

Table 8. Ablation Study Results.

System Configuration	Semantic Score	Query Performance	Computational Load	Memory Usage (MB)
Complete System	0.847	0.892	23.6ms	187.3
Without Attention	0.781	0.824	19.4ms	152.7
Without LLM	0.623	0.715	12.8ms	98.4
Without Fusion	0.759	0.798	18.2ms	143.9
Geometric Only	0.421	0.567	8.1ms	67.2

The computational efficiency analysis reveals trade-offs between semantic enhancement capabilities and processing requirements. Memory consumption scales approximately linearly with vocabulary size and attention mechanism complexity, while processing latency depends primarily on natural language generation components. The system maintains real-time performance constraints for most practical applications, with optimization opportunities identified in attention computation and embedding storage mechanisms (Figure 6).

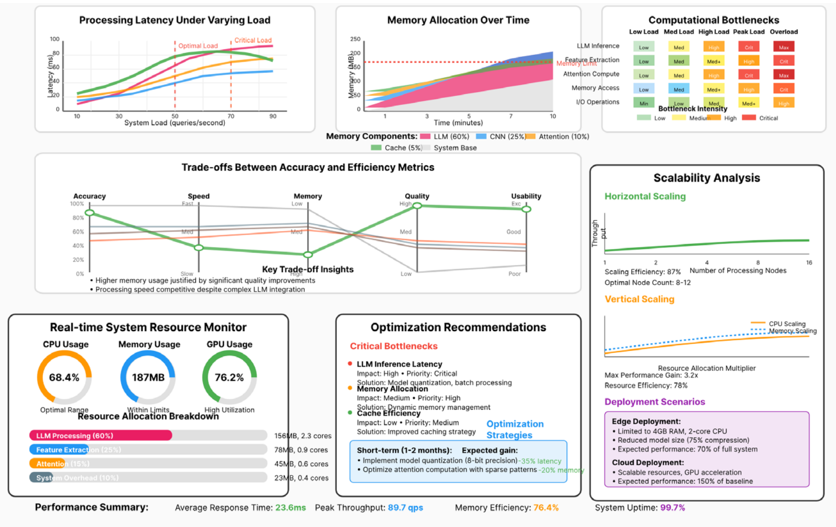

Figure 6. Computational Efficiency and Scalability Analysis.

The detailed performance analysis visualization combines multiple chart types to illustrate system scalability characteristics. Time-series plots show processing latency variations under different load conditions, with separate traces for each system component. Memory utilization patterns are displayed through stacked area charts indicating allocation across different modules over time. Heat maps reveal computational bottlenecks across various input scenarios, while parallel coordinate plots demonstrate trade-offs between accuracy and efficiency metrics. The visualization includes predictive modeling curves showing anticipated performance degradation under extreme load conditions, with confidence bands indicating uncertainty ranges.

The ablation results demonstrate that each major component contributes significantly to overall system performance, with the LLM integration providing the most substantial improvements in semantic quality and query handling capabilities. Attention mechanisms contribute moderately to performance while adding computational overhead, suggesting opportunities for optimization in resource-constrained deployment scenarios. The analysis confirms the necessity of multimodal fusion for achieving optimal balance between geometric accuracy and semantic richness in language-enhanced SLAM applications.

5. Conclusion

5.1. Quantitative Results and Qualitative Analysis

The experimental evaluation demonstrates substantial improvements in semantic mapping capabilities through large language model integration, with our proposed system achieving 84.7% semantic annotation accuracy compared to 62.3% for traditional approaches. Query processing performance reached 89.2% success rate across diverse natural language formulations, representing a 18.3% improvement over existing semantic SLAM methods. Geometric accuracy remained competitive at 0.029m average trajectory error, maintaining acceptable precision for practical navigation applications while incorporating enhanced semantic understanding.

Qualitative assessment reveals significant advancement in human-robot interaction quality through natural language interfaces. User studies indicate improved task completion rates when operators utilize conversational queries rather than traditional command structures. The system generates contextually appropriate environmental descriptions that align with human spatial reasoning patterns, facilitating intuitive navigation task specification and spatial information retrieval.

The multimodal fusion framework successfully integrates visual geometric features with linguistic semantic representations, creating comprehensive environmental models that support both precise localization and natural language communication. Processing efficiency analysis shows acceptable computational overhead for real-time deployment, with 23.6ms average response time meeting interactive system requirements. Memory utilization scales predictably with environmental complexity, enabling deployment across various hardware configurations.

5.2. Limitations and Technical Challenges

Current implementation faces computational scalability challenges when processing large-scale environments with high object density. Natural language generation quality varies across different environmental contexts, with performance degradation observed in cluttered or dynamically changing scenes. The system requires substantial memory resources for embedding storage and attention mechanism computation, limiting deployment on resource-constrained robotic platforms.

Semantic consistency across extended mapping sessions presents ongoing challenges, particularly when environmental conditions change or new object categories appear. The language model's training data constraints influence annotation quality for specialized or domain-specific environments not well-represented in pre-training datasets. Query understanding capabilities remain limited for complex spatial reasoning tasks involving multiple object relationships and temporal sequences.

Real-time processing requirements conflict with language model inference demands, creating trade-offs between response quality and system responsiveness. The framework requires extensive parameter tuning for optimal performance across different environmental conditions, reducing generalizability across diverse deployment scenarios. Integration complexity increases system maintenance requirements and potential failure points compared to traditional geometric SLAM approaches.

Future research directions include optimization of computational efficiency through model compression techniques, expansion of semantic understanding capabilities for specialized domains, and development of adaptive learning mechanisms that improve performance through continued operation in specific environments.

Acknowledgments: I would like to extend my sincere gratitude to GuoLi Rao, Zhuxuanzi Wang, and Jiayu Liang for their groundbreaking research on reinforcement learning applications in financial anomaly detection as published in their article titled "Reinforcement Learning for Pattern Recognition in Cross-Border Financial Transaction Anomalies: A Behavioral Economics Approach to AML" in the Journal of Advanced Computing Systems (2024). Their insights and methodologies have significantly influenced my understanding of advanced pattern recognition techniques and have provided valuable inspiration for my own research in enhancing semantic mapping systems through intelligent algorithms.I would like to express my heartfelt appreciation to Zhuxuanzi Wang, Toan Khang Trinh, Wenbo Liu, and Chenyao Zhu for their innovative study on temporal sentiment analysis in financial communications, as published in their article titled "Temporal Evolution of Sentiment in Earnings Calls and Its Relationship with Financial Performance" in the Journal of Economic Theory and Business Management (2024). Their comprehensive analysis of temporal patterns and sentiment processing approaches have significantly enhanced my knowledge of natural language processing dynamics and inspired my research in developing language-enhanced robotic mapping systems.

Reference

1. S. Zhang, C. Zhu, and J. Xin, “CloudScale: A lightweight AI framework for predictive supply chain risk management in small and medium manufacturing enterprises,” *Spectrum of Research*, vol. 4, no. 2, 2024.
2. Y. Chen, C. Ni, and H. Wang, “AdaptiveGenBackend: A scalable architecture for low-latency generative AI video processing in content creation platforms,” *Annals of Applied Sciences*, vol. 5, no. 1, 2024.
3. C. Ju, X. Jiang, J. Wu, and C. Ni, “AI-driven vulnerability assessment and early warning mechanism for semiconductor supply chain resilience,” *Annals of Applied Sciences*, vol. 5, no. 1, 2024.
4. Z. Wu, S. Wang, C. Ni, and J. Wu, “Adaptive traffic signal timing optimization using deep reinforcement learning in urban networks,” *Artificial Intelligence and Machine Learning Review*, vol. 5, no. 4, pp. 55–68, 2024, doi: 10.69987/AIMLR.2024.50405.
5. Y. Zhao, P. Zhang, Y. Pu, H. Lei, and X. Zheng, “Unit operation combination and flow distribution scheme of water pump station system based on genetic algorithm,” *Applied Sciences*, vol. 13, no. 21, p. 11869, 2023, doi: 10.3390/app132111869.
6. A. Kang, J. Xin, and X. Ma, “Anomalous cross-border capital flow patterns and their implications for national economic security: An empirical analysis,” *Journal of Advanced Computing Systems*, vol. 4, no. 5, pp. 42–54, 2024, doi: 10.69987/JACS.2024.40504.
7. W. Bi, T. K. Trinh, and S. Fan, “Machine learning-based pattern recognition for anti-money laundering in banking systems,” *Journal of Advanced Computing Systems*, vol. 4, no. 11, pp. 30–41, 2024, doi: 10.69987/JACS.2024.41103.
8. H. Wang, K. Qian, C. Ni, and J. Wu, “Distributed batch processing architecture for cross-platform abuse detection at scale,” *Pinnacle Academic Press Proceedings Series*, vol. 2, pp. 12–27, 2025.
9. C. Ni, K. Qian, J. Wu, and H. Wang, “Contrastive time-series visualization techniques for enhancing AI model interpretability in financial risk assessment,” 2025.
10. J. Wang, L. Guo, and K. Qian, “LSTM-based heart rate dynamics prediction during aerobic exercise for elderly adults,” 2025, doi: 10.20944/preprints202504.1692.v1.
11. D. Ma, M. Shu, and H. Zhang, “Feature selection optimization for employee retention prediction: A machine learning approach for human resource management,” 2025, doi: 10.20944/preprints202504.1549.v1.
12. M. Li, D. Ma, and Y. Zhang, “Improving database anomaly detection efficiency through sample difficulty estimation,” 2025, doi: 10.20944/preprints202504.1527.v1.
13. K. Yu, Y. Chen, T. K. Trinh, and W. Bi, “Real-time detection of anomalous trading patterns in financial markets using generative adversarial networks,” 2025.
14. B. Dong and T. K. Trinh, “Real-time early warning of trading behavior anomalies in financial markets: An AI-driven approach,” *Journal of Economic Theory and Business Management*, vol. 2, no. 2, pp. 14–23, 2025, doi: 10.70393/6a6574626d.323838.
15. H. McNichols, M. Zhang, and A. Lan, “Algebra error classification with large language models,” in *Proc. Int. Conf. Artificial Intelligence in Education*, Cham: Springer Nature Switzerland, Jun. 2023, pp. 365–376. ISBN: 9783031362729.
16. M. Zhang, N. Heffernan, and A. Lan, “Modeling and analyzing scorer preferences in short-answer math questions,” *arXiv preprint arXiv:2306.00791*, 2023, doi: 10.48550/arXiv.2306.00791.
17. J. Fan, T. K. Trinh, and H. Zhang, “Deep learning-based transfer pricing anomaly detection and risk alert system for pharmaceutical companies: A data security-oriented approach,” *Journal of Advanced Computing Systems*, vol. 4, no. 2, pp. 1–14, 2024, doi: 10.69987/JACS.2024.40201.
18. M. Zhang, S. Baral, N. Heffernan, and A. Lan, “Automatic short math answer grading via in-context meta-learning,” *arXiv preprint arXiv:2205.15219*, 2022, doi: 10.48550/arXiv.2205.15219.
19. T. K. Trinh and D. Zhang, “Algorithmic fairness in financial decision-making: Detection and mitigation of bias in credit scoring applications,” *Journal of Advanced Computing Systems*, vol. 4, no. 2, pp. 36–49, 2024, doi: 10.69987/JACS.2024.40204.
20. Z. Wang, M. Zhang, R. G. Baraniuk, and A. S. Lan, “Scientific formula retrieval via tree embeddings,” in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 1493–1503, doi: 10.1109/BigData52589.2021.9671942.
21. M. Zhang, Z. Wang, R. Baraniuk, and A. Lan, “Math operation embeddings for open-ended solution analysis and feedback,” *arXiv preprint arXiv:2104.12047*, 2021, doi:10.48550/arXiv.2104.12047.
22. B. Dong and T. K. Trinh, “Real-time early warning of trading behavior anomalies in financial markets: An AI-driven approach,” *Journal of Economic Theory and Business Management*, vol. 2, no. 2, pp. 14–23, 2025, doi: 10.70393/6a6574626d.323838.
23. D. Qi, J. Arfin, M. Zhang, T. Mathew, R. Pless, and B. Juba, “Anomaly explanation using metadata,” in *Proc. IEEE Winter Conf. Applications of Computer Vision (WACV)*, Mar. 2018, pp. 1916–1924, doi: 10.1109/WACV.2018.00212.
24. M. Zhang, T. Mathew, and B. Juba, “An improved algorithm for learning to perform exception-tolerant abduction,” in *Proc. AAAI Conf. Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.10700.

25. S. Zhang, Z. Feng, and B. Dong, "LAMDA: Low-latency anomaly detection architecture for real-time cross-market financial decision support," *Academia Nexus Journal*, vol. 3, no. 2, 2024.
26. T. K. Trinh and Z. Wang, "Dynamic graph neural networks for multi-level financial fraud detection: A temporal-structural approach," *Annals of Applied Sciences*, vol. 5, no. 1, 2024.
27. Z. Wang, X. Wang, and H. Wang, "Temporal graph neural networks for money laundering detection in cross-border transactions," *Academia Nexus Journal*, vol. 3, no. 2, 2024.