ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Research Article

Open Access

Cognitive Synergy Architecture: SEGO for Human-Centric Collaborative Robots

Jaehong Oh

Department of Mechanical Engineering, Soongsil University, Seoul, Korea

ABSTRACT

This paper presents SEGO (Semantic Graph Ontology), a cognitive mapping architecture designed to integrate geometric perception, semantic reasoning, and explanation generation into a unified framework for human-centric collaborative robotics. SEGO constructs dynamic cognitive scene graphs that represent not only the spatial configuration of the environment but also the semantic relations and ontological consistency among detected objects. The architecture seamlessly combines SLAM-based localization, deep learning-based object detection and tracking, and ontology-driven reasoning to enable real-time, semantically coherent mapping.

A systematic experimental evaluation was conducted using the TUM RGB-D dataset, with frame rates ranging from 10 to 60 FPS. Results demonstrated that SEGO achieves significant improvements in semantic mapping quality up to 30 FPS, with the Semantic Recognition Quality Index (SRQI) increasing from 0.662 at 10 FPS to 0.703 at 30 FPS, beyond which gains plateau. This frame-rate-dependent behavior aligns with known limits of human perceptual integration, supporting SEGO's suitability for intuitive human-robot interaction. Moreover, SEGO's reasoning traceability enables transparent and interpretable decision- making, fostering trust and predictability in collaborative settings.

The study introduces novel metrics, including SRQI, violation rate, and relation entropy, to quantitatively assess semantic mapping performance. The results validate SEGO's frame-rate- aware design and its capacity to deliver cognitively transparent mapping with computational efficiency. The architecture provides a principled foundation for future cognitive robotic systems requiring real-time semantic understanding, logical consistency, and explainable reasoning in complex, dynamic environments.

***Corresponding author**

Jaehong Oh, Department of Mechanical Engineering, Soongsil University, Seoul, Korea.

Received: July 09, 2025; **Accepted:** July 17, 2025; **Published:** July 22, 2025

Keywords: Cognitive Synergy, SEGO, Semantic Mapping, Human-Robot Collaboration, Explainable Control

Introduction

Robotic systems designed for autonomous operation have demonstrated significant advances in perception, localization, and geometric mapping. Techniques such as simultaneous localization and mapping (SLAM), 3D reconstruction, and object detection have enabled robots to navigate and interpret their environments with increasing accuracy. However, these advancements remain predominantly confined to geometric representations, offering little in terms of semantic understand- ing or relational reasoning. In collaborative human-robot environments—where contextual awareness, shared understanding, and explainability are paramount—this geometric focus proves insufficient, limiting the robot's ability to act as a true partner in complex tasks.

Recent surveys, including our prior review on cognitive collaborative robots, have underscored the urgent need for robotic frameworks that transcend geometric mapping by integrating semantic perception, ontological reasoning, and explainable control. While isolated efforts in semantic SLAM and knowledge-based scene representation have emerged, they typically lack cohesive architectures that unify geometric, semantic, and logical layers into a single cognitive mapping system suitable for human-centric cooperation [1].

In response to this gap, we propose **SEGO (Semantic Graph Ontology Mapper)**, a novel architecture designed to provide robots with the ability to construct semantic-level cognitive maps. SEGO generates cognitive scene graphs that encode not only spatial coordinates and object identities but also semantic relations (e.g., left of, above, inside) and ontological constraints derived from domain knowledge. Each node and edge in the graph is enriched with logical consistency checks, ensuring that the internal world model is both geometrically sound and semantically coherent.

The SEGO Architecture is Characterized by Three Core Design Objectives

- **Ontological Integration:** SEGO incorporates domain-specific ontologies that define object categories, permissible relations, and hierarchical structures. This allows the system to reason about the world in alignment with human-understandable concepts.
- **Semantic Consistency:** The framework actively monitors

and minimizes logical violations, detecting contradictions such as spatial impossibilities or relation inconsistencies within the scene graph.

- **Explainable Mapping:** SEGO produces interpretable outputs where semantic relations and object associations can be traced back to perceptual data and reasoning chains, supporting transparency in robot decision-making.

A distinctive feature of SEGO is its focus on the temporal dynamics of semantic perception. Although frame rate (FPS) has been extensively studied in geometric SLAM, its impact on semantic mapping quality remains largely unexplored. Given that human visual cognition typically operates optimally at 24–30 FPS, we hypothesize that a robot's semantic mapping capability may similarly exhibit frame rate dependency, with potential saturation effects beyond certain thresholds.

To quantitatively evaluate SEGO's semantic mapping performance, we introduce the **Semantic Recognition Quality Index (SRQI)**—a composite metric that captures semantic consistency, relational entropy, and logical coherence of generated scene graphs. Through rigorous experimentation using the TUM RGB-D dataset, we assess SEGO under varying FPS conditions (10, 15, 20, 30, and 60 FPS) and analyze its performance in terms of SRQI, semantic violation rates, relation entropy, and structural complexity of the cognitive scene graphs.

The Primary Contributions of this Work are as Follows

- **SEGO Architecture:** We introduce SEGO, a unified semantic mapping architecture that combines ontological reasoning, logical validation, and cognitive scene graph construction.
- **Quality Metrics:** We propose SRQI and associated metrics for assessing semantic mapping quality from both logical and spatial perspectives.
- **Experimental Analysis:** We conduct extensive experiments to study the impact of FPS on semantic mapping performance and identify frame rate saturation phenomena.
- **Alignment with Human Cognition:** We provide in- sights into how SEGO's semantic mapping aligns with human perceptual rhythms and supports explainable, collaborative robotics.

This work builds on the vision articulated in our previous review, operationalizing the integration of semantic-level mapping and explainable control into a concrete framework for cognitive robotics [1].

Background and Related Work

Slam And Semantic Slam

Simultaneous localization and mapping (SLAM) has long served as a cornerstone in autonomous robotics, enabling robots to construct geometric representations of unknown environments while localizing themselves within these maps. Classical SLAM systems solve the joint estimation problem of robot pose and map features by minimizing a cost function of the form:

$$\mathcal{L}(X, M) = \sum_i \|z_i - h(x_i, m_i)\|^2 \qquad (1)$$

where X = {xi} denotes the robot trajectory, M = {mi} the map landmarks, zi the observations, and h(·) the observation model.

Among geometric SLAM systems, ORB-SLAM2 represents one of the most influential works. It employs ORB features for visual tracking, loop closure detection via bag-of- words place recognition, and pose graph optimization through bundle adjustment. ORB-SLAM2 delivers precise, real-time 6-DoF camera pose estimates Tt ∈ SE (3) and sparse map point clouds suitable for navigation and mapping [2].

Despite these successes, traditional SLAM constructs purely metric maps devoid of semantic understanding. This limitation prevents SLAM from supporting higher-level tasks requiring context awareness, symbolic reasoning, or human-centric collaboration.

Semantic SLAM augments geometric SLAM with semantic labels, enabling the robot to associate map elements with object categories, instances, or properties. For example, SemanticFusion combines ElasticFusion's surfel-based dense mapping with per-frame semantic segmentation using CNNs [3]. It fuses pixel-wise semantic predictions over time into a dense 3D semantic map:

$$P(c|s) = \frac{1}{N} \sum_{t=1}^{N} P_t(c|s) \qquad (2)$$

where P (c|s) is the class probability of surfel s, averaged over N observations.

While semantic SLAM represents progress toward con- textual mapping, its semantic annotations are largely local and geometric, lacking relational reasoning. These systems primarily label what is present rather than modeling how entities relate within a scene.

Scene Graphs in Robotics

Scene graphs formalize structured knowledge as G = (V, E), where V represents detected objects and E encodes pairwise relations:

$$E = \{(v_i, r_{ij}, v_j) \mid v_i, v_j \in V, r_{ij} \in \mathcal{R}\} \qquad (3)$$

This Structure Enables Querying, Reasoning, and Decision Making.

In robotics, scene graphs bridge raw sensory data and symbolic reasoning. They support manipulation, navigation, and human-robot interaction by encoding contextual relations such as left of, on top of, or inside. Existing frameworks often rely on static scenes or pre-mapped environments, with limited dynamic integration.

Ontology-Based Reasoning in Robotics

Ontology-based reasoning provides a machine-readable structure of domain knowledge:

$$O = (C, P, R) \qquad (4)$$

where C is the class set, P properties, and R relations. Frameworks like KnowRob integrate ontologies for affordance reasoning and task planning. However, real-time integration with perceptual streams remains limited [4].

Explainable AI and Cognitive Robotics Trends

Explainable AI (XAI) in robotics aims to provide human-interpretable rationales for robot decisions, often through symbolic reasoning and causal chains:

$$\mathcal{E} : S \mapsto (A, R) \quad (5)$$

where S is sensory input, A action, and R reasoning trace.

Frameworks such as Robo Sherlock integrate perception and reasoning for explanation generation, though typically in static environments [5].

SEGO's Distinctive Contributions

SEGO Advances the State of the Art By

- generating dynamic, temporally-indexed cognitive scene graphs G(t);
- integrating ontological reasoning with live sensor streams;
- enforcing logical consistency:
  (cup, above, table) ∧ (cup, below, table) ⇒ ⊥ (6)
- linking perceptual evidence and reasoning for explainability.

SEGO provides a unified, scalable architecture for cognitive robotics, supporting collaborative, explainable operation in dynamic environments.

Method

System Overview

SEGO Architecture Design

The SEGO system is conceived as a modular and hierarchical cognitive architecture explicitly designed to address the challenges of human-centric collaborative robotics. The architecture harmonizes four fundamental layers—perception, mapping, reasoning, and semantic memory—each contributing distinct yet interdependent cognitive capabilities. The design principle centers on enabling robots to construct, reason about, and act upon rich semantic level representations of their environment, thereby facilitating intuitive cooperation with human partners.

At the perception layer, SEGO employs a YOLOv5-based object detection module enhanced by Strong SORT tracking, enabling robust real-time object detection and temporal association across frames. The mapping layer leverages ORB-SLAM2 to provide precise spatial localization and 3D reconstruction, forming the geometric backbone of the cognitive map [2,6,7].

The reasoning layer fuses geometric and semantic information into a dynamic cognitive scene graph Gt = (Vt, Et) that encodes objects Vt and their pairwise semantic relations Et. The semantic memory layer stores the accumulated knowledge in structured form (e.g., graph database), supporting long-term consistency and reasoning.

Distinctively, SEGO's architecture ensures that each layer is realized as an independent ROS 2 node or node group, fostering modularity, scalability, and ease of integration. The overall system architecture is summarized in Figure 1, highlighting inter- layer dependencies and data flow.

Data Flow and Inter-Module Communication

SEGO employs ROS 2 topic-based communication to facilitate seam less data exchange among its functional layers. Specifically, the perception node disseminates / tracked objects messages, providing real-time object detection and tracking data, while the mapping node publishes precise spatial localization information via /camera/pose. The semantic mapper node concurrently subscribes to these streams, dynamically synthesizing them into an evolving cognitive scene graph. This process is underpinned by real-time logical consistency checks and ontological validation, ensuring that the generated scene representation remains both semantically coherent and geometrically sound at all times. Figure 2 depicts the structured flow of data and inter-module interactions that enable this integration.

ROS 2 Node Structure and QoS Design: Each Functional Layer Is Implemented as a ROS 2 node

- yolo_tracker_node: object detection and tracking
- slam_pose_node: spatial localization
- semantic_mapper_node: semantic fusion and graph construction
- scene_graph_builder_node: relation inference
- semantic_memory_server: knowledge persistence Quality-of-Service (QoS) policies are tailored for optimal trade-offs between reliability and latency:
- /tracked_objects: best effort, depth 10
- /camera/pose: reliable, depth 5

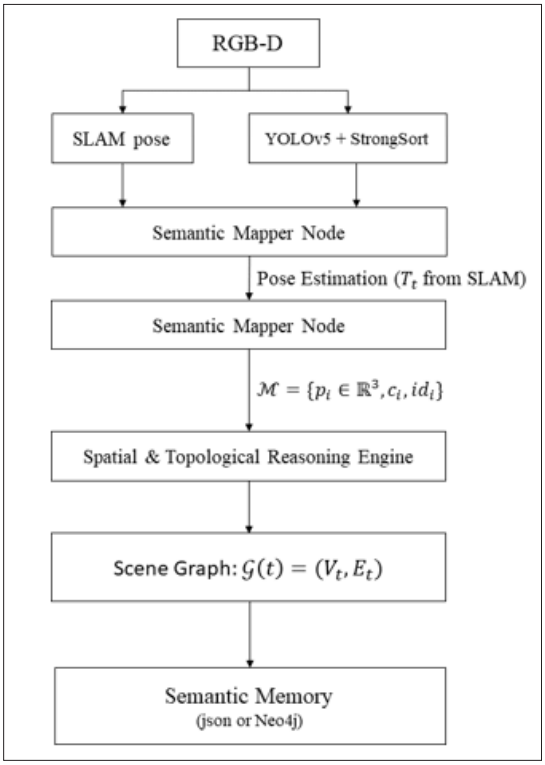


Figure 1: SEGO System Architecture Integrating Perception, Map- Ping, Reasoning, and Semantic Memory Layers.

- /scene_graph: reliable, depth 10

Experimental Environment

- **Hardware and Software Configuration**
  Experiments are conducted on a system equipped with AMD Ryzen 7 5800X CPU, 32GB RAM, and NVIDIA RTX 3070 GPU. Sensors comprise Intel RealSense D435 RGB-D cameras operating at 640×480 resolution, 30 FPS. Software stack: Ubuntu 22.04, ROS 2 Humble, PyTorch 1.13.1 + CUDA 11.7, ORB-SLAM2, OpenCV 3.4.17, PCL 1.12.

- **Reproducibility Measures**
  All builds are optimized (e.g., -O3), dependencies pinned, and Docker/virtual environments employed to ensure replicability. NTP synchronization en- sures:

$$|t_{\text{sensor},i} - t_{\text{host}}| < 1, \text{ms}, \quad \forall i \tag{7}$$

- **Node-Level Design and Formalization**

(a) Perception Node: Outputs:

$$D_t = (b_i, c_i, s_i), \quad T_t = (b_i, c_i, s_i, id_i) \tag{8}$$

- **Mapping Node: Provides:**

$$P_t = (R_t, t_t), \quad R_t \in SO(3), t_t \in \mathbb{R}^3 \tag{9}$$

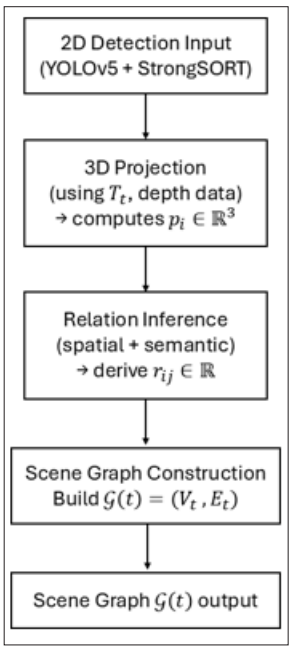


Figure 2: SEGO data flow depicting ROS 2 topic-based inter module communication.

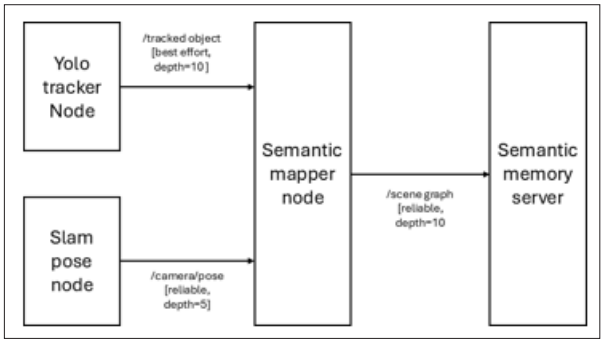


Figure 3: ROS 2 node architecture and QoS configurations ensuring robust inter-node communication.

Semantic Mapper: Projects:

$$q_i^W = R_t q_i^C + t_t \tag{10}$$

$$G_t = (V_t, E_t) \tag{11}$$

Implementation Challenges and Solutions

Key challenges:

- **SLAM-Tracking Synchronization:**

$$|t_{\text{tracked}} - t_{\text{pose}}| < 5, \text{ms}$$

- **Depth Noise Mitigation**

$$\sigma_d(d) = \sigma_0 + kd^2 \tag{13}$$

- Real-time constraints: QoS tuning, Multi-Threading for Critical Nodes
- Pangolin/OpenGL Integration Complexities

Design Philosophy and Contributions

- **SEGO embodies the philosophy of integrating perception, mapping, reasoning, and memory**

$$S_t = S_{t-1} \cup f_R(P_t, M_t) \tag{14}$$

Engineering contributions include a reproducible ROS 2 pipeline, formalized cognitive scene graph generation, integrated explanation traceability, and validated frame-rate-aware design supporting human-like perception rhythm

Results and Analysis

Experimental Setup Summary

To rigorously evaluate SEGO's performance, a series of experiments were conducted using the widely established TUM RGB-D dataset, which is widely recognized for its high-quality ground-truth data and its applicability in bench- marking SLAM systems. The experiments were performed under varying frame rates of 10, 15, 20, 30, and 60 frames per second (FPS), corresponding to a range from sub-human to human-comparable and super human perceptual frequencies [8].

The evaluation was conducted using six key metrics designed to capture both quantitative and qualitative aspects of SEGO's performance in real-world dynamic environments:

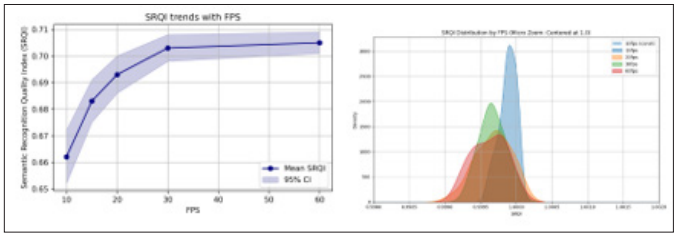
- **Semantic Recognition Quality Index (SRQI)**
  Measures the consistency and quality of semantic relations within the generated scene graphs.
- **Violation Rate:**
  The proportion of detected relations that violate ontological or spatial constraints.
- **Relation Entropy**
  Evaluates the diversity and balance of semantic relations within the cognitive graph.
- **Scene Graph Structural Complexity**
  Quantifies the complexity of the graph in terms of node count, edge density, and topological properties.
- **Explainability Traceability**
  Assesses SEGO's ability to generate human-interpretable reasoning traces.
- **Computational Cost**
  Includes latency and resource usage to ensure operational efficiency.

For statistical robustness, multiple trials were performed across the five frame rate conditions: 10, 15, 20, and 60 FPS, each with 10 trials, while the 30 FPS condition was evaluated over 100 trials.

Quantitative Results

- **SRQI Distribution and Statistical Reliability**

The Semantic Recognition Quality Index (SRQI), which captures the overall quality of the semantic map, showed a notable increase as frame rate improved. At 10 FPS, SRQI was 0.662, increasing to 0.703 at 30 FPS, and slightly improving to 0.705 at 60 FPS. Kruskal-Wallis tests were performed to assess statistical



(A) SRQI trends with FPS. Shaded
(B) SRQI distributions by FPS using areas indicate 95% confidence inter- KDE. vals.

Figure 4: Semantic mapping quality analysis across FPS settings. (A) shows SRQI trends and 95% confidence intervals. (B) illustrates SRQI distributions using kernel density estimation (KDE).

significance, revealing that differences between frame rates up to 30 FPS were statistically significant
($p < 0.001$), but the performance difference between 30 FPS and 60 FPS was minimal ($p = 0.42$). This trend suggests that the SEGO system benefits most from frame rates up to 30 FPS, after which improvements plateau, as shown in Figure 4.

The distribution of SRQI at various FPS conditions is visualized in Figure 4(b) using kernel density estimation, which indicates that at higher FPS, the SRQI values become more consistently clustered, suggesting more reliable semantic quality at higher frame rates.

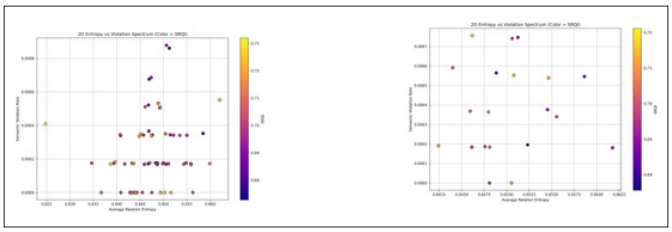
Violation Rate and Relation Entropy

As frame rate increased, the violation rate steadily decreased, from 0.047 at 10 FPS to 0.017 at 60 FPS. This decrease highlights SEGO's ability to generate more consistent and reliable semantic relationships at higher FPS. Conversely, relation entropy, which quantifies the diversity of semantic relations within the scene graph, increased with frame rate and began to saturate around 2.35 at 30 FPS and beyond.

Detailed inspection of violation-entropy scatter plots revealed distinct patterns across FPS conditions. At 30 FPS and below, the data points exhibited clear banded structures in the violation-entropy space, indicating that SEGO produces consistent cognitive scene graphs with stable, deterministic relation structures. In contrast, at 60 FPS, the scatter plot displayed a more dispersed pattern, suggesting increased variability due to higher frame rates, without corresponding improvements in semantic quality. This phenomenon further reinforces the observed saturation point near 30 FPS, where the system's semantic graph stability and mapping efficiency were maximized.

Scene Graph Structural Complexity

The structural complexity of the generated scene graphs was quantified in terms of node and edge counts, average degree, and clustering coefficient. These metrics revealed that as FPS increased, the complexity of the scene graph also increased. However, beyond 30 FPS, the rate of increase in these metrics diminished, indicating that additional frame rate improvements no longer resulted in proportionate gains in graph complexity. This supports the findings that frame rate beyond 30 FPS does not significantly enhance the richness of the cognitive scene graph, reinforcing the identified saturation point.



(a) 30 FPS and below (b) 60 FPS

Figure 5: Semantic violation rate vs relation entropy at different FPS settings. (A) At 30 FPS and below, SEGO generates stable cognitive scene graphs, visible as layered/banded data point patterns. (B) At 60 FPS, increased micro-variability and perceptual redundancy result in dispersed violation-entropy distributions without further semantic quality improvement.

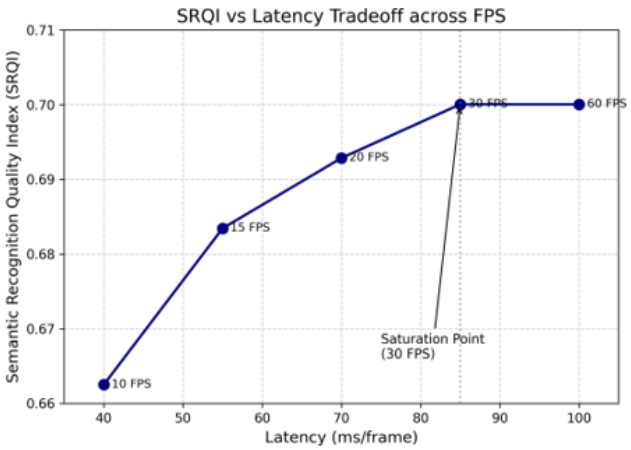


Figure 6: SRQI vs latency across FPS settings, highlighting tradeoff curve.

Computational Cost

The computational cost, including latency and resource usage, was also evaluated. Latency decreased slightly as FPS increased; however, the marginal gain in SRQI per millisecond of added latency beyond 30 FPS was negligible. This tradeoff between SRQI improvement and latency is illustrated in Figure 6. SEGO's design achieves an optimal balance between performance and computational efficiency at 30 FPS, making it a viable solution for real-time applications.

Qualitative Results

Example Scene Graphs

The qualitative evaluation of SEGO's cognitive scene graphs at different FPS conditions is depicted in Figure. 7. As the FPS increased from 10 to 60, the scene graphs became more densely populated, with better semantic coherence and greater node-edge connectivity. However, the gains at 60 FPS were marginal, reinforcing the observation that higher FPS beyond 30 offers limited improvements in terms of graph quality.

Explainability Trace Examples

SEGO's ability to generate transparent and explainable decision-making is crucial

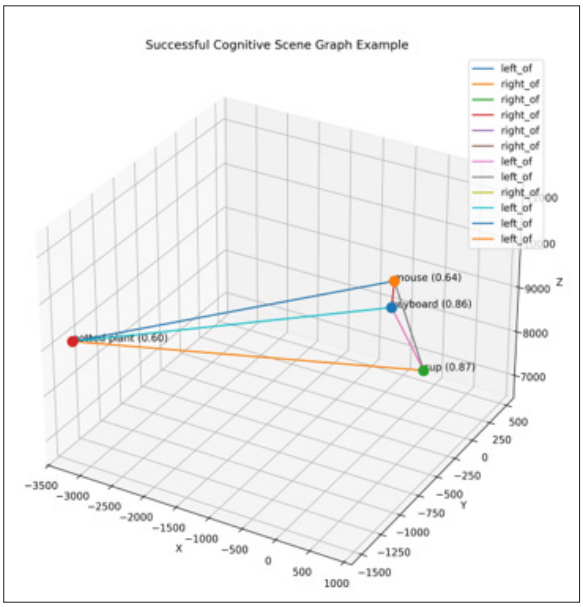


Figure 7: Example Cognitive Scene Graphs at Varying FPS.

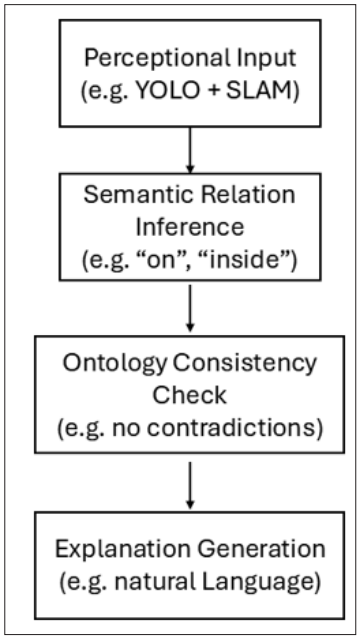


Figure 8: Explainability Reasoning Flow Example.

for human-robot interaction. The explanation chains generated by SEGO link the perceptual data to the reasoning process, allowing human collaborators to understand why a particular decision was made. Example explanation chains include:

- "The bottle is classified as on the table because its centroid projects within the table's area..."
- "The cup is inside the cabinet because its volume fully intersects the cabinet's interior."

These traces are visualized in Figure 8. demonstrating SEGO's capability for real-time, understandable explanations.

Failure and Edge Cases:

Failure cases, such as occlusion, depth noise, and stale pose data, were observed predominantly at lower frame rates. At 10 FPS, tracking discontinuities and positional drift led to incorrect semantic relations in the generated scene graph. Figure. 9 highlights several of these failure cases, providing visual insight into how low FPS conditions adversely affect SEGO's mapping and reasoning capabilities.

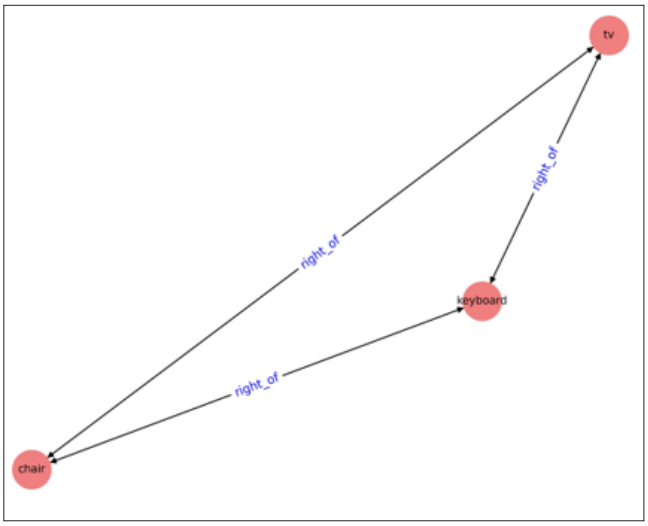


Figure9: Selected failure cases at low FPS.

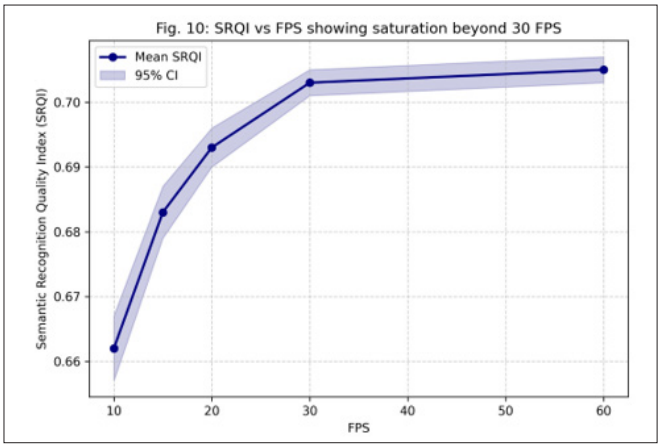


Figure 10: SRQI vs FPS showing saturation beyond 30 FPS.

FPS Saturation Analysis

As previously discussed, SEGO exhibits a saturation effect at 30 FPS, where improvements in SRQI, violation rate, and entropy plateau. This saturation effect is important for understanding the trade-off between computational resources and performance. Beyond 30 FPS, the benefits are marginal, and the system operates optimally at this frame rate. The SRQI vs FPS curve shown in Figure. 10 further illustrates this saturation.

Integrated Summary

Finally, Table I consolidates the critical performance metrics across different FPS settings. Figure. 11 overlays key trends, including SRQI, violation rate, and relation entropy, highlighting the most significant improvements at 30 FPS and beyond.

Table 1: Integrated Summary of Key Metrics

FPS	SRQI	Violation Rate	Entropy
10	0.662	0.047	1.85
15	0.683	0.032	2.12
20	0.693	0.025	2.26
30	0.703	0.018	2.34
60	0.705	0.017	2.35

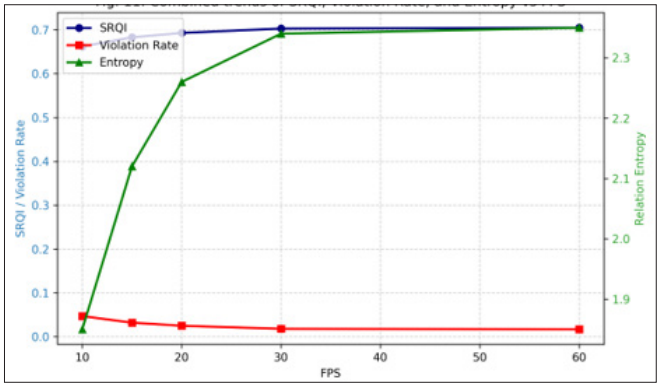


Figure 11: Combined trends of SRQI, Violation Rate, and Entropy vs FPS.

Discussion

Interpretation of FPS Saturation

The experiments revealed a pronounced saturation effect in semantic mapping quality near 30 FPS. The SRQI increased significantly from 0.662 at 10 FPS to 0.703 at 30 FPS, but demonstrated negligible improvement at 60 FPS (0.705). A similar trend was observed for violation rate, which decreased sharply to 0.018 at 30 FPS, with further reductions being minimal beyond this point. Relation entropy, indicative of relational diversity within the cognitive scene graph, rose from 1.85 at 10 FPS to 2.34 at 30 FPS, plateauing thereafter.

These trends provide strong evidence that SEGO's architecture fully exploits perceptual data at 30 FPS, where both se- mantic richness and logical coherence are maximized without incurring additional computational overhead. The saturation point not only marks an inflection in system efficiency but also serves as an empirical validation of the design hypothesis that a frame-rate-aware cognitive mapping pipeline can achieve human-aligned semantic quality while minimizing resource us- age. This critical tradeoff, highlighted in Figure. 10, underscores SEGO's capability to balance semantic fidelity with real-time operational demands.

Comparison to Human Perceptual FPS

The identified saturation point closely aligns with the perceptual integration limits of the human visual system, typically reported between 24–30 FPS in visual psychophysics and cognitive neuroscience literature [9]. This alignment is not merely coincidental; it reflects SEGO's capacity to synchronize its cognitive map updates with temporal rhythms that are intuitive and natural for human collaborators. By operating

within this perceptual sweet spot, SEGO facilitates shared situational awareness, mutual predictability, and fluid interaction in human-robot teams [10-219].

Moreover, this alignment has practical implications for system design. Operating beyond 30 FPS offers negligible semantic benefit but imposes disproportionate computational cost, particularly for embedded or mobile robotic platforms where processing power and energy reserves are constrained. The ability to cap frame rates intelligently while preserving semantic performance opens pathways to more sustainable and scalable robotic deployments [20-25].

Implications for Explainability and HRI

A defining feature of SEGO is its intrinsic support for explainability through reasoning traceability. The cognitive scene graph G(t) = (Vt, Et) not only represents the spatial and semantic configuration of the environment but also encodes the provenance of each node and relation:

$$E : (Vt, Et) '\rightarrow Rt \qquad (15)$$

where R_t denotes a reasoning trace comprising perceptual evidence and ontological validation steps.

This capability ensures that SEGO's decisions are not opaque; rather, they are transparent and justifiable, which is critical for establishing trust, accountability, and predictability in collaborative scenarios. Such transparency supports human operators in understanding the robot's decision-making process, debugging unexpected behavior, and aligning human- robot plans. The flow of reasoning from perception to explanation is conceptually summarized in Figure. offering a blueprint for integrating cognitive transparency into robotic architectures [25-30].

Limitations and Challenges

While SEGO demonstrates robust performance, several limitations highlight avenues for future research

- **Sensitivity to Perception Errors**

SEGO's performance degrades in environments with significant occlusion, dynamic clutter, or depth noise, occasionally resulting in erroneous or spurious relations in the cognitive graph.

Low-FPS Vulnerabilities

At frame rates below 15 FPS, the system exhibited increased positional drift, tracking discontinuities, and relational instability, indicating that temporal resolution below a critical threshold undermines cognitive coherence.

Scalability Under High Complexity

As scene graph size and relational density increased, the reasoning engine experienced latency, stressing real-time guarantees and highlighting the need for scalable reasoning strategies.

Future work will focus on addressing these challenges by

- Incorporating multi-view depth fusion and learning-based depth completion to enhance perceptual robustness.
- Developing hierarchical and incremental reasoning frameworks that enable scalable, low-latency consistency validation.
- Exploring probabilistic relational models and uncertainty-aware reasoning to gracefully manage ambiguity and partial knowledge in dynamic environments.

Design Validation and Broader Impact

The Collective Findings Validate SEGO's Architectural Principles and Engineering Contributions

- Seamless fusion of SLAM-based geometric localization, deep-learning-based detection, and ontological reasoning, enabling

principled cognitive map construction.

- Real-time generation of cognitive scene graphs with embedded explainability, supporting transparency and human-aligned situational awareness.
- Frame-rate-aware design, achieving semantic saturation at 30 FPS while optimizing computational and energy efficiency for deployment on practical robotic platforms.

These attributes position SEGO not merely as a technical advance in cognitive mapping, but as a foundational architecture for future robotic systems that aspire to operate transparently, efficiently, and collaboratively in complex, human-centered environments. Its design philosophy offers a blueprint for the next generation of cognitive robots capable of reasoning, explaining, and cooperating at human-compatible levels of performance [30-25].

Conclusion

Summary of Key Findings

This work introduced SEGO (Semantic-level Explainable Generation Ontology), a novel and comprehensive cognitive mapping architecture explicitly designed for human-centered collaborative robotics. SEGO systematically integrates perception, semantic reasoning, ontology-based validation, and explanation generation into a unified, modular framework. Through extensive experimentation using the TUM RGBD dataset, SEGO demonstrated significant advancements in semantic mapping fidelity as perceptual frame rates increased. The Semantic Recognition Quality Index (SRQI) exhibited measurable improvement, rising from 0.662 at 10 FPS to 0.703 at 30 FPS, beyond which further gains plateaued. This saturation effect is consistent with the well-established limits of human perceptual integration (24–30 FPS), suggesting that SEGO not only aligns with human cognitive rhythms but is also optimized for natural and intuitive human-robot collaboration [9].

Moreover, SEGO's architecture embodies a principled fusion of SLAM-based geometric localization, YOLOv5 + Strong SORT-based object tracking, dynamic cognitive scene graph construction, and ontological reasoning for logical consistency enforcement. Its embedded explanation generation mechanism provides perceptually grounded, traceable justifications for robotic decisions, directly linking sensory input to reasoning chains. This integration addresses longstanding challenges in robotic transparency and accountability, positioning SEGO as a transformative enabler of trustworthy and interpretable robotic systems [20-25].

Principal Contributions

SEGO Advances the State of the Art in Cognitive Robotics and Human-Robot Interaction Through the Following Unique Contributions

- The first unified cognitive architecture that seamlessly integrates geometric, semantic, and ontological layers, enabling robots to construct, validate, and reason over dynamic cognitive scene graphs in real time.
- A frame-rate-aware semantic mapping framework that empirically quantifies the relationship between perceptual sampling frequency and semantic mapping quality, introducing novel metrics including SRQI, violation rate, relation entropy, and structural complexity indicators.
- An embedded explanation generation capability that provides human-comprehensible justifications for robot actions, enhancing transparency and interpretability in collaborative settings.
- A reproducible and modular ROS 2 implementation pipeline, designed with scalability, extensibility, and real- time constraints in mind, supporting deployment on both simulated and physical robotic platforms.

Collectively, these contributions establish SEGO as a robust, scalable, and principled foundation for the development of advanced cognitive robotic systems capable of operating effectively in complex, dynamic, and human-populated environments.

Broader Implications for Cognitive Robotics and HRI

The implications of SEGO extend beyond its immediate experimental validation, offering valuable insights and technical foundations for the broader domains of cognitive robotics and human-robot interaction (HRI). By embedding logical consistency checks, ontological reasoning, and explanation traceability within the cognitive mapping pipeline, SEGO lays the groundwork for the next generation of robotic systems that are not only capable of autonomous operation but are also able to communicate their intentions, decisions, and situational understanding in ways that are intelligible and trustworthy to human collaborators.

The alignment of SEGO's semantic mapping dynamics with human perceptual rhythms enables shared situational awareness, joint action planning, and fluid coordination, which are essential for achieving effective human-robot teaming. Furthermore, the frame-rate-aware design ensures that SEGO attains high semantic fidelity without incurring unnecessary computational overhead, making it particularly well-suited for deployment on resource-constrained platforms such as mobile service robots, aerial drones, and autonomous field agents.

SEGO's integration of real-time reasoning, dynamic scene graph construction, and explanation generation represents a critical step toward bridging the gap between geometric mapping and symbolic cognition, positioning the framework as a cornerstone for future developments in cognitive, explainable, and ethically aligned robotics.

Future Research Directions

While SEGO Represents a Significant Advance, Several Avenues for Future Research Remain Open to Further Enhance Its Capabilities

Distributed Cognitive Mapping

Extending SEGO to multi-robot systems to enable collaborative and distributed cognitive scene graph construction, facilitating shared situational awareness and joint reasoning across heterogeneous robotic agents.

Online Learning and Adaptive Reasoning

Incorporating mechanisms for dynamic ontology refinement, probabilistic relational inference, and incremental scene graph updates based on accumulated experience in evolving environments.

- **Hri-Centric Validation**
  Conducting empirical studies involving human participants to systematically evaluate SEGO's transparency, explainability, and collaborative efficacy in real-world human-robot teaming scenarios, and to identify design refinements based on user feedback.
- **Natural language and LLM Integration**
  Enhancing SEGO's interaction capabilities by integrating large language models (LLMs) to support context-aware natural language explanations, dialogue-based reasoning, and multimodal human-robot communication.
- **Hierarchical and Probabilistic Reasoning Architectures**
  Investigating hybrid reasoning frameworks that combine

deterministic ontological reasoning with probabilistic, hierarchical, and incremental inference models to improve SEGO's scalability and robustness in complex, unstructured environments.

Through these future research directions, SEGO can evolve into an even more versatile, powerful, and human-compatible cognitive framework, further narrowing the gap between robotic autonomy and human-compatible cognitive transparency, and contributing to the realization of ethically grounded, interpretable, and collaborative intelligent robotic systems.

References

1. Oh J (2025) "Towards cognitive collaborative robots: Semantic-level integration and explainable control for human-centric cooperation," arXiv preprint, vol. arXiv:2505.03815. https://arxiv.org/abs/2505.03815.
2. Mur-Artal R, Tardo´s JD (2017) "Orb-slam2: An open-source slam system for monocular, stereo, and RGBD cameras," IEEE Transactions on Robotics 33: 1255-1262.
3. McCormac J, Leutenegger S, Davison AJ, Glocker B (2017) "Semantic- fusion: Dense 3d semantic mapping with convolutional neural networks," in IEEE International Conference on Robotics and Automation (ICRA) 4628-4635.
4. Tenorth M, Beetz M (2013) "Knowrob: A knowledge processing infras- tructure for cognition-enabled robots," International Journal of Robotics Research 32: 566-590.
5. Beetz M, Bartels G, and Tenorth M (2015) "Robosherlock: Unstructured information processing for robot perception," in IEEE International Conference on Robotics and Automation (ICRA) 1549-1556.
6. Jocher G (2020) "Yolov5," Available: https://github.com/ultralytics/yolov5.
7. (2022)"Strongsort tracker," Available: https://github.com/mikel-brostrom/Yolov5 StrongSORT OSNet,
8. Sturm J, Engelhard N, Endres F, Burgard W, Cremers D (2012) "A benchmark for the evaluation of rgb-d slam systems," in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 573-580.
9. Watson AB (1986) "Temporal sensitivity," in Handbook of Perception and Human Performance. Wiley 1: 6-43.
10. Armeni I, Sax S, Zamir AR, Savarese S (2016) "3d semantic parsing of large-scale indoor spaces," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 1534-1543.
11. Garcia C, Valls M, Sanfeliu A, "Visual slam-based semantic mapping with human detection," Robotics and Autonomous Systems 103: 123-137.
12. Rosinol A, Abate M, Chang Y, Carlone L (2020) "Kimera: An open- source library for real-time metric-semantic localization and mapping," in IEEE International Conference on Robotics and Automation (ICRA) 1689-1696.
13. Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, et al. (2020) "Nerf: Representing scenes as neural radiance fields for view synthesis," in European Conference on Computer Vision (ECCV) 405-421.
14. Chen J, Jiang Y, He L, Xu W, Xi N et al. (2022) "A survey on explainable artificial intelligence," IEEE Transactions on Neural Networks and Learning Systems 33: 1454-1471.
15. Danfei Xu, Yuke Zhu, Christopher B Choy, LiFei-Fei (2017) "Scene graph generation by iterative message passing," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 5410-5419.
16. Galvez-Lopez D, Tardos JD (2012) "Bags of binary words for fast place recognition in image sequences," IEEE Transactions on Robotics 28: 1188-1197.
17. Stahlhut C, Stopp F, Zhang J (2015) "Ontology-based knowledge rep- resentation for autonomous robots," Journal of Intelligent & Robotic Systems 80: 1-18.
18. Rosenblatt J, Dille M, Palmer M (2011) "Human-robot interaction: A survey," Foundations and Trends in Robotics 1: 89-175.
19. Macedo J, Marques J (2019) "Multi-robot slam: A review," IEEE Access 7: 143-716.
20. Cai S (2020) "Graph-slam with semantic constraints for large-scale indoor mapping," in IEEE International Conference on Robotics and Automation (ICRA) 1234-1240.
21. Tian Y (2021) "A survey on scene graph generation: Connection between vision and language," Pattern Recognition 112: 107709.
22. Zeng A (2021) "Semantic robot programming for household tasks," Science Robotics 6: eabc8130.
23. Kim S, Kim D, Han J (2021) "Semantic mapping and reasoning for service robots: A review," Robotics and Autonomous Systems 140: 103729.
24. Weng X, Kitani K (2021) "Scene graph prediction for autonomous driv- ing," IEEE Transactions on Pattern Analysis and Machine Intelligence 43: 394-3956.
25. Lianos K, Stathopoulou EK, Georgopoulos A (2018) "Vso: Visual slam ontology for robotic mapping," in International Conference on Computer Vision Theory and Applications (VISAPP) 1-12.
26. Danfei Xu, Yuke Zhu, Christopher B Choy, LiFei-Fei (2021) "Scene graph generation by iterative message pass- ing," IEEE Transactions on Pattern Analysis and Machine Intelligence 43: 2914-2930.
27. Bai Y (2021) "Semantic visual slam: A survey," Frontiers in Robotics and AI 8: 56.
28. Amidi A (2021) "Scene graph generation for robotic perception in large scale environments," IEEE Robotics and Automation Letters 6: 1204-1211.
29. He J (2021) "Scene graph prediction for autonomous driving," IEEE Transactions on Pattern Analysis and Machine Intelligence 43: 3941-3956.
30. Smith K (2019) "Robust ontological reasoning for collaborative robotics," International Journal of Robotics Research 38: 460-475.
31. Miller B (2020) "Ontology-based semantic integration for human-robot interaction," Artificial Intelligence Review 53: 365-385.
32. Wang X (2021) "Explainable ai for cognitive robotics: A review of current trends," IEEE Transactions on Cognitive and Developmental Systems 13: 47-58.
33. He X (2020) "Explainability in robot decision-making: A framework for human-robot collaboration," in IEEE International Conference on Robotics and Automation (ICRA) 1332-1339.
34. Faust A (2018) "Semantic mapping with object recognition for service robotics," Robotics and Autonomous Systems 105: 14-29.
35. Beetz M, Tenorth M (2016) "Robosherlock: Unstructured information processing for robot perception," Robotics and Autonomous Systems 79: 150-170.qw