
Magellan: Guided MCTS for Latent Space Exploration and Novelty Generation

Anonymous Author(s)
Affiliation
Address
email

A Supplementary Material

A.1 Knowledge Corpus and Evaluation Dataset

The knowledge corpus serves as the semantic foundation for our agent. Its design was guided by the principles of representativeness and methodological rigor. To ensure it reflects the current state-of-the-art, we curated papers from premier venues. The detailed composition is in Table 1.

Table 1: Composition of the Knowledge Corpus.

Source	Years	Number of Papers
CVPR	2023–2025	7,937
ICML	2023–2025	7,695
Nature Medicine	2022–2025	950
Total		16,582

A.2 Theme Generation Methodology.

Our automated theme generation process, which produced the evaluation dataset, is rooted in conceptual clustering. We applied K-Means to the document embeddings, partitioning the semantic space into $K = 20$ clusters. This value was chosen to create a fine-grained conceptual map, allowing the theme generator to bridge specific and nuanced conceptual gaps. The generator then repeatedly sampled pairs of papers from distinct clusters and prompted an LLM to synthesize a bridging theme, using the prompt detailed in the A.2. A qualitative visualization of the corpus clusters is provided in Figure 1.

To illustrate the capability of our automated theme generation module (Section ??), we present a curated example below. The process begins by selecting two concepts from different, but related, conceptual clusters. These serve as inputs to the LLM, which then synthesizes a novel, bridging research theme.

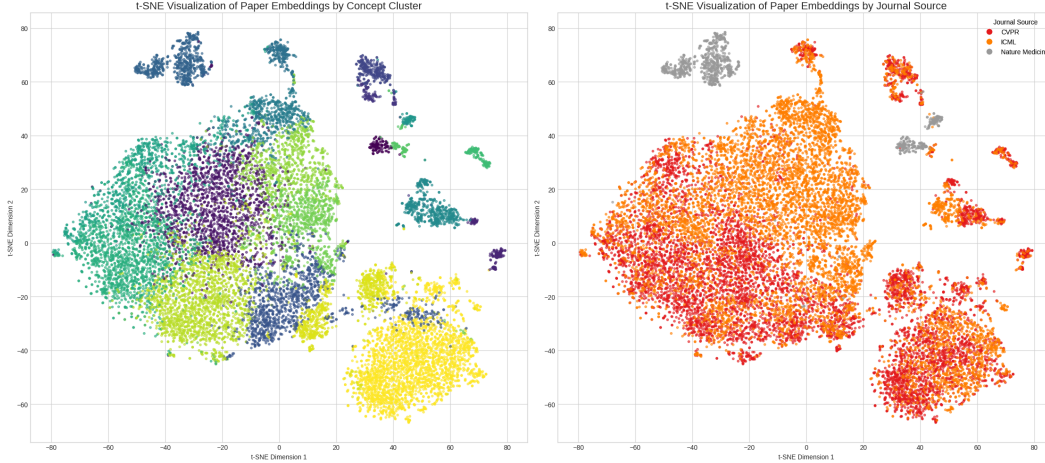


Figure 1: t-SNE visualization of the knowledge corpus embeddings. Each point represents a paper, colored by its assigned cluster ID (0-19). The plot shows clear semantic separation between conceptual groups, validating the basis of our cross-disciplinary theme generation strategy.

Prompt for Automated Theme Generation

You are a creative scientist tasked with generating a novel research proposal by synthesizing two concepts.

Concept 1 : *concept 1 text*

Concept 2 : *concept 2 text*

First, think step-by-step in a <think> block. Analyze both concepts. Find a plausible, insightful, and forward-looking connection. You could apply a technique from one to a problem in the other, find a shared principle, or use one as an analogy for the other.

After your thinking process, output the result as a JSON object with two keys:

1. **theme**: A concise, high-level research theme that captures the core idea. This should be a single, memorable sentence.
2. **elaboration**: A detailed, one-paragraph explanation of the theme. This should elaborate on the connection you found, outline the potential approach, and highlight the novelty. This will serve as the introductory context for the research proposal.

Example format:

```
```json
{
 "theme": "Leveraging Quantum-Inspired Tensor Networks for Explainable Large-Scale Graph Representation Learning.",
 "elaboration": "Current Graph Neural Networks (GNNs) often act as black boxes, limiting their trustworthiness in high-stakes domains. This research proposes a novel framework that adapts principles from quantum many-body physics, specifically tensor networks, to create a new class of GNNs. By representing graph structures and features as a tensor network, we can leverage efficient contraction algorithms (like DMRG) for node classification and link prediction, while the inherent structure of the network provides a direct, model-based explanation for its predictions, addressing the critical need for interpretability in complex graph data."
}
```

18

### 19 A.3 Implementation and Hyperparameter Details

20 To ensure full reproducibility, this section provides key details on the models and hyperparameters  
 21 used in our experiments.

22 **General LLM Configuration.** Across all experiments, for both our method and the baselines,  
23 the core Large Language Model was **Qwen3-1.7B ?**. For text generation, we consistently used a  
24 sampling temperature of 0.7 and a top-p value of 0.9 to encourage creative yet coherent outputs.

25 **Magellan Configuration.** The MCTS search was configured with a maximum of 30 iterations and  
26 an expansion width ( $K$ ) of 3. The early stopping mechanism was triggered if the best path remained  
27 stable for 2 consecutive iterations (Patience=2). The UCT formula was balanced with an exploration  
28 constant ( $C$ ) of 1.5 and a guidance weight ( $w_g$ ) of 1.0. The three components of our evaluation  
29 function were weighted as  $w_{\text{coh}} = 0.5$ ,  $w_{\text{nov}} = 0.3$ , and  $w_{\text{prog}} = 0.2$ . The guidance vector weights  
30 were set to  $\alpha = 1.0$  and  $\beta = 1.0$ .

31 **Baselines Configuration.** The baselines were configured to be strong competitors. For **Tree of**  
32 **Thoughts (ToT)**, we allowed the model to explore 5 candidates at each of a maximum of 5 steps.  
33 For **ReAct**, the agent was permitted a maximum of 10 steps to develop its reasoning and action plan.  
34 Detailed prompt settings for each method are shown in TableA.3.

#### Prompt for Chain of Thoughts

You are a research scientist. Based on the initial research idea below, write a complete and detailed research proposal. The proposal should be well-structured, clear, and scientifically plausible. Let's think step by step to ensure the logic is sound and the details are comprehensive.

Initial Research Idea:

Theme: *theme*

Elaboration: *elaboration*

First, I will analyze the core problem and the proposed approach. Then, I will outline the methodology, potential experiments, and expected outcomes.

My Detailed Research Proposal:

35

## Prompt for Tree of Thought

### Generator (first time)

You are a research scientist brainstorming a proposal.

Initial Idea: *input\_seq*

Based on this, generate 3 distinct and promising opening paragraphs for the proposal. Each paragraph should explore a slightly different angle or focus.

IMPORTANT: Present each paragraph separated by '—'.

Paragraph 1:

### Generator

You are a research scientist continuing a proposal draft.

Initial Idea: *input\_seq*

Proposal so far:

—

*state*

—

Based on the proposal so far, generate 3 distinct and logical next paragraphs to continue the proposal. Each should build upon the existing text in a unique way.

IMPORTANT: Present each paragraph separated by '—'.

Next Paragraph 1:

### Evaluator

You are a strict, expert peer reviewer. The original research theme is: *input\_seq*

Here is a partial research proposal draft:

—

*state*

—

Evaluate this draft on a scale of 1 to 10 based on its potential to become a high-impact paper. Consider its novelty, clarity, and scientific feasibility.

Your response MUST be a single integer from 1 to 10, with 10 being the best. Do not add any other text.

Score:

### Prompt for ReAct

You are a research scientist assistant. Your goal is to write a detailed research proposal based on an initial theme. You operate in a loop of Thought, Action, Observation.

At each step, you must first think about your plan, then choose ONE of the following actions:

- Search[topic]: Search for a specific topic in the internal knowledge base to get more information.
- Write[paragraph\_plan]: Write the next section of the proposal based on your plan.
- Finish[]: Conclude the process when the proposal is complete.

Here is an example:

—  
Initial Idea: Theme: Using Graph Neural Networks (GNNs) for protein-protein interaction (PPI) prediction. Elaboration: Current methods are slow and not interpretable.

Thought: The idea is good, but generic. I need to find a specific, novel GNN architecture to propose. I will search for limitations of current GNNs in this area.

Action: Search[GNN limitations for protein interaction]

Observation: Found documents mentioning scalability issues and problems with dynamic graphs.

Thought: The key challenges are scalability and dynamic interactions. I can propose a new model using a temporal GNN architecture. I will now write the introduction and methods section based on this plan.

Action: Write[Write the introduction explaining the problem and the proposed temporal GNN model. Then, detail the model architecture in the methods section.]

Observation: Paragraphs successfully written.

Thought: The proposal has an introduction and methods. Now I need to describe the experiments and expected outcomes.

Action: Write[Write the 'Experiments' section, describing datasets, metrics, and baselines. Then write the 'Expected Outcomes' section.]

Observation: Paragraphs successfully written.

Thought: The proposal is complete with all core sections. I will now finish.

Action: Finish[]

—  
Now, begin with the following task:

Initial Idea: *initial\_idea*

### Prompt for Magellan

You are a world-class Principal Investigator, known for writing clear, compelling, and fundable research proposals.

Your current task is to expand on the following research idea:

—  
*theme*  
—

You will now write the **\*\*next section\*\*** of this proposal.

Based on these principles, generate a distinct, detailed, well-reasoned and deepen "next section" for the research plan.

To do this, you must follow these core principles of scientific writing:

1. **Progressive Deepening:** Your new section **MUST** logically follow from the existing text. It should deepen the idea, moving from a general concept to specific details, or from a hypothesis to a method of testing it. Do not repeat existing information; build upon it.
2. **Concrete Detail:** Be specific and avoid vague language. When you are describing a mechanism, explain it with sufficient details for another expert to understand (with math, if needed).
3. **Critical Thinking:** Briefly acknowledge potential challenges, limitations, or alternative approaches to your proposed section. This demonstrates foresight.

First think about what is missing in the paragraph for a well-written research plan, or which part is not detailed enough. And then finish it.

**\*\* Make sure the whole article is coherent and logical. \*\***

38

### 39 **A.4 LLM-as-a-Judge Protocol**

40 To ensure a transparent and reproducible evaluation process, this section details the protocol used  
41 for our LLM-as-a-Judge methodology. We utilized the **DeepSeek-V3.1-Think** model ? as the core  
42 evaluator. The evaluation was guided by a prompt shown in A.4.

### Prompt for LLM-as-a-Judge

You are a distinguished professor and the chair of a top-tier academic conference, known for your rigorous, fair, and insightful reviews. Your task is to evaluate five scientific ideas generated by different AI models for a core research theme. You will score each proposal on a scale of 1-10 for the three dimensions below, providing a concise reason for each score. Finally, you must select a **single** best proposal overall.

#### Core Evaluation Dimensions:

1. **Plausibility:** Is the idea scientifically plausible? (1=Nonsense, 10=Highly Plausible)
2. **Structure & Clarity:** Is the structure complete and the logic coherent? (1=Chaotic, 10=Crystal Clear)
3. **Innovation Potential:** Does the idea present novel viewpoints, methods, or research paths? (1=Obsolete, 10=Highly Innovative)

#### Important Guidelines:

- These are preliminary ideas, not full research proposals. The absence of sections like introduction or methods is acceptable. Focus on the core value of the idea itself.
- Your scoring should be strict and discerning to reflect the quality differences between the proposals.
- It is acceptable and even encouraged for proposals to reasonably extend or innovate upon the initial theme. This should be considered a merit, not a deviation from the topic.

#### Output Requirement:

Please return your review strictly in the following JSON format, without any additional explanations or comments.

```
```json
{
  "evaluations": {
    "A": {
      "method": "...",
      "plausibility": <score_1_to_10>,
      "structure_clarity": <score_1_to_10>,
      "innovation_potential": <score_1_to_10>,
      "reason": "<A concise justification for the scores>"
    },
    ...
  },
  "final_decision": {
    "best_proposal": "<The single best proposal: A, B, C, D, or E>",
    "justification": "<A comprehensive explanation of why this proposal is the best overall>"
  }
}
```
```

#### Content to Evaluate:

**Core Theme:** *core theme*

**Elaboration:** *elaboration*

**Proposal A:**

*proposal A text*

**Proposal B:**

*proposal B text*

...

### Example of Automated Theme Generation 1

**Input Concept 1:** Co-SLAM: Joint Coordinate and Sparse Parametric Encodings for Neural Real-Time SLAM

**Input Concept 2:** Wrapped Gaussian on the manifold of Symmetric Positive Definite Matrices

#### Synthesized Theme:

**Theme:** Geometric Manifold Integration for Real-Time SLAM and Data Modeling

**Elaboration:** This research proposes integrating the geometric structure of symmetric positive definite (SPD) matrices with real-time SLAM systems to enhance robustness and efficiency. By leveraging the wrapped Gaussian distribution on the SPD manifold, we extend Co-SLAM's hybrid representation (hash-grid + one-blob encoding) to incorporate geometric constraints. The SPD manifold's inherent structure allows for efficient, low-dimensional parameterization of scene features, enabling real-time bundle adjustment while preserving surface coherence. The wrapped Gaussian's probabilistic framework ensures geometric consistency, addressing limitations in traditional SLAM's handling of high-frequency local features. This approach merges the efficiency of hash-grid representations with the geometric fidelity of SPD manifolds, enabling novel SLAM algorithms that dynamically adapt to complex, structured environments while maintaining high reconstruction accuracy and tracking robustness.

44

### Example of Automated Theme Generation 2

**Input Concept 1:** Clinical utility of targeted RNA sequencing in cancer molecular diagnostics

**Input Concept 2:** Masked and Adaptive Transformer for Exemplar Based Image Translation

#### Synthesized Theme:

**Theme:** Adaptive Transformer-Based Molecular Feature Matching for Precision Oncology and Image Translation

**Elaboration:** This research proposes integrating the masked and adaptive transformer (MAT) framework from exemplar-based image translation into RNA sequencing for cancer diagnostics. By leveraging the MAT's ability to learn cross-domain semantic correspondence and context-aware feature augmentation, we aim to enhance the accuracy of molecular feature matching in RNA-seq data, particularly for detecting fusion events and actionable alterations. The MAT's contrastive style learning principles are adapted to prioritize biologically relevant features in RNA sequences, improving diagnostic precision and therapeutic relevance. This approach bridges the gap between high-dimensional molecular data and actionable clinical insights, while also demonstrating the versatility of transformer-based architectures in tackling complex, domain-specific challenges across biology and computer vision.

45



### Example of Automated Theme Generation 3

**Input Concept 1:** WildlifeMapper: Aerial Image Analysis for Multi-Species Detection and Identification

**Input Concept 2:** MagicLens: Self-Supervised Image Retrieval with Open-Ended Instructions: Image retrieval

#### Synthesized Theme:

**Theme:** Integrating Open-Ended Text Instructions with Aerial Image Analysis for Enhanced Wildlife Species Detection and Retrieval

**Elaboration:** This research proposes a novel framework that merges the self-supervised, instruction-driven retrieval capabilities of MagicLens with the multi-species detection prowess of WildlifeMapper. By leveraging text instructions to encode complex ecological relationships (e.g., 'identify all large mammals in dense forest areas' or 'retrieve images of birds near water bodies'), the system enhances the contextual understanding of aerial imagery. The approach synthesizes MagicLens's implicit relation mining from web data with WildlifeMapper's aerial dataset, enabling the model to generalize across diverse species and environments. This integration allows for dynamic, open-ended queries that go beyond visual similarity, such as detecting species based on habitat context or ecological roles, while maintaining the efficiency and accuracy of automated wildlife monitoring. The novelty lies in combining text-based instruction learning with aerial image analysis to address the limitations of static, species-specific models, offering a scalable solution for real-time, adaptive environmental conservation.

46

## 47 A.5 Example Generated Themes

### 48 A.5.1 Example: Integrating Adversarial Prompt Tuning with Multi-Task Collaboration for 49 Robust Vision-Language Models

50 Integrating Adversarial Prompt Tuning with Multi-Task Collaboration for Robust Vision-Language  
51 Models This research proposes a novel framework that combines adversarial prompt tuning (TAPT)  
52 with multi-task collaboration (WeakMCN) to enhance the robustness and performance of vision-  
53 language models. By leveraging the adversarial training principles of TAPT, we design defensive  
54 prompts that dynamically adapt to task-specific requirements, while WeakMCN's dual-branch archi-  
55 tecture ensures collaborative learning between weakly supervised tasks (WREC and WRES). The  
56 integration of adversarial prompts in a multi-task setting allows the model to simultaneously optimize  
57 for task-specific objectives (e.g., grounding in WREC) and robustness against adversarial perturba-  
58 tions (e.g., visual attacks in TAPT). Key innovations include dynamic visual feature enhancement  
59 (DVFE) to adaptively combine pre-trained visual knowledge and a collaborative consistency module  
60 (CCM) to enforce cross-task alignment during optimization. This approach not only improves perfor-  
61 mance on benchmarks like RefCOCO but also ensures generalization in semi-supervised settings,  
62 demonstrating a novel synergy between adversarial defense and multi-task learning.

#### 63 Technical Framework and Methodology

To operationalize the synergy between adversarial prompt tuning (TAPT) and multi-task collaboration (WeakMCN), we propose a hierarchical framework that integrates adversarial prompt generation, multi-task learning, and dynamic feature adaptation. The core idea is to embed adversarial prompts into the multi-task learning pipeline as a regularization mechanism, ensuring that the model learns robust representations that are invariant to adversarial perturbations while maintaining task-specific performance. Specifically, we design a dual-branch architecture where the \*adversarial prompt branch\* generates task-agnostic defensive prompts via a gradient-based adversarial training process, and the \*multi-task branch\* processes task-specific inputs (e.g., visual-linguistic grounding, weakly supervised reasoning) using WeakMCN's dual-branch structure. The adversarial prompt generation is formalized as a game between the model and an adversary. For a given task, the model learns to minimize the loss  $L_{\text{task}}$ , while the adversary maximizes the perturbation  $\epsilon$  that degrades task

performance. This is modeled as a minimax optimization problem:

$$\min_{\theta} \max_{\delta} [L_{\text{task}}(\theta, \delta) + \lambda \cdot \|\delta\|_2],$$

where  $\theta$  represents the model parameters,  $\delta$  is the adversarial perturbation, and  $\lambda$  balances robustness and task accuracy. The adversarial prompts are then sampled from the distribution of  $\delta$ , and the model is trained to generalize across both clean and perturbed inputs. The multi-task collaboration is achieved through a collaborative consistency module (CCM), which enforces cross-task alignment by minimizing a cross-task consistency loss  $L_{\text{CCM}}$ . This loss is computed as the KL divergence between the outputs of the visual branch (for WREC) and the language branch (for WRES):

$$L_{\text{CCM}} = \mathcal{D}_{\text{KL}}(P_{\text{visual}} \parallel P_{\text{language}}),$$

where  $P_{\text{visual}}$  and  $P_{\text{language}}$  are the distributions of predictions from the visual and language branches, respectively. This ensures that the model’s visual and language modules are aligned during adversarial training, preventing task-specific biases from dominating the learning process. To adapt to task-specific requirements, we introduce dynamic visual feature enhancement (DVFE), which combines pre-trained visual features with task-specific prompts using a weighted fusion mechanism:

$$F_{\text{fusion}} = \alpha \cdot F_{\text{pretrained}} + (1 - \alpha) \cdot F_{\text{prompt}},$$

64 where  $\alpha$  is a learnable parameter that dynamically adjusts the contribution of pre-trained features  
65 versus adversarial prompts. This allows the model to leverage domain knowledge while remaining  
66 robust to adversarial attacks.

## 67 Challenges and Considerations

68 A critical challenge is balancing the adversarial training’s robustness with the multi-task learning’s  
69 specificity. Overly strong adversarial perturbations may degrade task performance, while weak  
70 perturbations may fail to enforce robustness. To mitigate this, we incorporate a gradient penalty  
71 term in the adversarial loss to ensure smoothness in the perturbation space. Another challenge is  
72 computational efficiency, as adversarial training increases the gradient computation cost. We address  
73 this by using a hybrid training strategy: adversarial prompts are generated during the validation  
74 phase, while the model is trained on clean data during the main training loop. This framework not  
75 only advances the state-of-the-art in robust vision-language models but also provides a scalable  
76 approach for deploying models in adversarial environments, such as real-time applications with noisy  
77 or manipulated inputs.

78 **Experimental Evaluation and Validation** To validate the effectiveness of our framework, we  
79 design a comprehensive evaluation plan that systematically tests the synergy between adversarial  
80 prompt tuning (TAPT) and multi-task collaboration (WeakMCN) across diverse vision-language  
81 tasks. The experiments are structured to address three core objectives: (1) benchmark performance  
82 on standard vision-language tasks, (2) evaluate robustness against adversarial perturbations, and (3)  
83 assess generalization in semi-supervised and low-data settings.

84 **Benchmark Tasks and Datasets** We evaluate our framework on three representative tasks:

85 (1) **Visual-Text Grounding (WREC)**, which involves aligning visual regions with text descriptions  
86 (e.g., RefCOCO and RefCOCO+);

87 (2) **Weakly Supervised Reasoning (WRES)**, which requires reasoning over visual-linguistic rela-  
88 tionships (e.g., Visual Reasoning Benchmarks); and

89 (3) **Adversarial Robustness**, where we test the model’s ability to maintain performance under visual  
90 perturbations (e.g., Gaussian noise, JPEG compression, and adversarial attacks from the \*AdvProp\*  
91 dataset). For benchmarking, we compare our framework against state-of-the-art methods, including:

92 - **TAPT-only models**: Adversarial prompt tuning without multi-task collaboration.

93 - **WeakMCN-only models**: Multi-task collaboration without adversarial defense.

94 - **Hybrid baselines**: Existing methods that combine adversarial training with multi-task learning (e.g.,  
95 \*Adversarial Prompt Learning\* and \*Multi-Task Robust Learning\*).

96 **Performance Metrics** We measure performance using task-specific metrics:

97 - For WREC, we use **Intersection over Union (IoU)** and **Mean Average Precision (mAP)**. We  
98 evaluate **reasoning accuracy** (e.g., correct inference on logical questions) and **visual-linguistic**  
99 **consistency** (e.g., KL divergence between visual and language predictions).

100 - For adversarial robustness, we compute **task accuracy under perturbation** (e.g., accuracy on clean  
 101 data vs. perturbed data) and **robustness margin** (e.g., maximum perturbation strength before task  
 102 failure).

103 **Semi-Supervised and Low-Data Evaluation** To assess generalization, we conduct experiments in  
 104 semi-supervised settings where the model is trained on a large pre-training corpus and fine-tuned on  
 105 small task-specific datasets. We evaluate:

106 - **Domain adaptation:** Performance on out-of-distribution tasks (e.g., medical imaging, low-light  
 107 scenes).

108 - **Data efficiency:** Training on 10% of the full dataset while maintaining performance.

109 **Implementation Details and Baseline Comparisons** The framework is implemented using PyTorch  
 110 and Hugging Face Transformers, with adversarial prompts generated via the \*FGSM\* (Fast Gradient  
 111 Sign Method) and \*PGD\* (Projected Gradient Descent) algorithms. Key hyperparameters include:  
 112 - Adversarial perturbation budget  $\epsilon = 0.03$  (L2 norm). - Gradient penalty coefficient  $\lambda = 0.1$   
 113 in the minimax optimization. - Dynamic fusion weight  $\alpha$  trained via a softmax distribution over  
 114 task-specific tasks. We also compare against alternative approaches:

115 - **Task-specific adversarial training:** Applying TAPT to individual tasks without multi-task collabora-  
 116 tion.

117 - **Multi-task baseline without adversarial defense:** WeakMCN trained on clean data.

118 - **Multi-task baseline with weak adversarial training:** WeakMCN trained with minimal perturba-  
 119 tions.

## 120 **Potential Challenges and Mitigations**

A critical challenge is the trade-off between robustness and task specificity: excessive adversarial training may degrade performance on clean data, while weak perturbations may fail to enforce robustness. To address this, we incorporate a **smoothness constraint** in the adversarial loss:

$$L_{\text{adv}} = \min_{\theta} \max_{\delta} [L_{\text{task}}(\theta, \delta) + \lambda \cdot \|\delta\|_2 + \mu \cdot \|\nabla_{\delta} L_{\text{task}}\|_2],$$

121 where  $\mu$  penalizes abrupt changes in perturbation gradients, ensuring smooth adversarial examples.  
 122 Additionally, we use a **hybrid training strategy** where adversarial prompts are generated during  
 123 validation, while the model is trained on clean data during the main loop to reduce computational  
 124 overhead. This evaluation plan not only quantifies the framework’s performance but also rigorously  
 125 tests its scalability, robustness, and adaptability to real-world scenarios, providing a clear path to  
 126 practical deployment in adversarial environments."

## 127 **A.5.2 Example: Geometric Manifold Integration for Real-Time SLAM and Data Modeling**

128 **Geometric Manifold Integration for Real-Time SLAM and Data Modeling** This research proposes  
 129 integrating the geometric structure of symmetric positive definite (SPD) matrices with real-time  
 130 SLAM systems to enhance robustness and efficiency. By leveraging the wrapped Gaussian distribution  
 131 on the SPD manifold, we extend Co-SLAM’s hybrid representation (hash-grid + one-blob encoding)  
 132 to incorporate geometric constraints. The SPD manifold’s inherent structure allows for efficient,  
 133 low-dimensional parameterization of scene features, enabling real-time bundle adjustment while  
 134 preserving surface coherence. The wrapped Gaussian’s probabilistic framework ensures geometric  
 135 consistency, addressing limitations in traditional SLAM’s handling of high-frequency local features.  
 136 This approach merges the efficiency of hash-grid representations with the geometric fidelity of  
 137 SPD manifolds, enabling novel SLAM algorithms that dynamically adapt to complex, structured  
 138 environments while maintaining high reconstruction accuracy and tracking robustness.

## 139 **Mathematical Framework and Algorithm Design**

140 To formalize the geometric manifold integration, we will develop a rigorous mathematical frame-  
 141 work that bridges the geometric structure of symmetric positive definite (SPD) matrices with the  
 142 probabilistic constraints of the wrapped Gaussian distribution. The SPD manifold, defined as the  
 143 set of  $n \times n$  matrices with positive eigenvalues, is parameterized via the Cholesky decomposition,  
 144 where each matrix  $X \in \mathbb{R}^{n \times n}$  is represented as  $X = \text{Cholesky}(Z)$ , with  $Z \in \mathbb{R}^{n \times n}$  and  $Z^T Z = X$ .  
 145 This parameterization ensures that the manifold is smooth, compact, and equipped with a natural  
 146 Riemannian metric, enabling efficient optimization over the space of features. The wrapped Gaussian

distribution, a key component of the framework, is defined on the SPD manifold by leveraging the eigenvalue decomposition of the matrix. Specifically, the probability density function (PDF) of a point  $X$  on the manifold is given by:  $p(X) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{1}{2}(y - \mu)^T \text{trace}^{-1}(y - \mu)\right)$ , where  $\Sigma$  is the covariance matrix of the distribution and  $\mu$  is the mean. This formulation ensures that the distribution is invariant under orthogonal transformations, preserving geometric consistency during SLAM estimation. The wrapped Gaussian’s ability to model local features with low-dimensional parameterization aligns with the SPD manifold’s structure, enabling real-time bundle adjustment while maintaining surface coherence. To integrate this into Co-SLAM, we extend the hash-grid encoding to operate on the SPD manifold. The hash-grid parameterizes the spatial distribution of features by discretizing the manifold’s geometry into a grid, where each cell corresponds to a region of the manifold. This reduces the dimensionality of the feature space, allowing for faster updates during SLAM. The one-blob encoding, which captures the local geometry of features, is adapted to enforce geometric constraints by ensuring that the estimated pose and feature positions remain consistent with the manifold’s curvature.

The algorithm design involves three key steps: **data preprocessing**, **manifold embedding**, and **optimization**. During data preprocessing, feature points are projected onto the SPD manifold using the logarithmic map  $\log(X)$ , which maps the matrix to its log-determinant space. This step ensures that the features are represented in a coordinate system compatible with the manifold’s geometry. The manifold embedding step involves updating the hash-grid and one-blob encodings dynamically as new features are added, leveraging the SPD manifold’s structure to maintain spatial coherence. The optimization process employs a variational approach, minimizing the difference between the estimated pose and the observed data using the wrapped Gaussian distribution. This is achieved by formulating the problem as a constrained optimization:  $\min_{X, \theta} \sum_{i=1}^N [\log p(X_i) + \text{tr}(\Sigma_i^{-1}(\mathbf{z}_i - \pi(\theta, X_i))(\mathbf{z}_i - \pi(\theta, X_i))^T)]$ , where  $i, \theta$  represents the pose parameters and  $X_i$  are the estimated feature positions. The constraints ensure that the optimization respects the SPD manifold’s structure, preventing degenerate solutions and maintaining the integrity of the geometric constraints.

#### 174 Critical Considerations:

175 - **Computational Efficiency:** The logarithmic map and low-dimensional parameterization reduce the  
176 computational burden, but high-frequency local features may necessitate additional constraints to  
177 prevent numerical instability.

178 - **Robustness to Noise:** The wrapped Gaussian’s probabilistic framework inherently accounts for  
179 noise, but its effectiveness in high-dimensional spaces requires careful tuning of the covariance matrix  
180  $\Sigma$ .

181 - **Alternative Approaches:** While the SPD manifold offers geometric fidelity, alternative methods like  
182 differential geometry or manifold learning could provide flexibility. However, these approaches often  
183 require more complex preprocessing or may not align as closely with the probabilistic constraints of  
184 the wrapped Gaussian. By combining the geometric rigor of the SPD manifold with the probabilistic  
185 robustness of the wrapped Gaussian, this framework enables real-time SLAM systems that dynam-  
186 ically adapt to complex environments while preserving geometric consistency and computational  
187 efficiency.

#### 188 Implementation and Optimization of the Geometric Manifold Framework

189 To operationalize the geometric manifold integration framework, we will develop a modular algorithm  
190 that integrates the SPD manifold’s geometric structure with the wrapped Gaussian distribution’s  
191 probabilistic constraints. The core of the implementation lies in two critical components: **hash-grid**  
192 **parameterization** and **one-blob encoding**, which together enforce geometric consistency and enable  
193 real-time optimization.

194 **Hash-Grid Parameterization and One-Blob Encoding** The hash-grid encoding, adapted to the  
195 SPD manifold, discretizes the manifold’s geometry into a hierarchical structure. Each grid cell  
196 corresponds to a region of the manifold, and feature points are assigned to cells based on their  
197 spatial distribution. This reduces the dimensionality of the feature space, enabling efficient updates  
198 during SLAM. The one-blob encoding, a geometric representation of local features, is modified to  
199 enforce constraints on the SPD manifold. Specifically, the encoding ensures that the estimated pose  
200 and feature positions remain consistent with the manifold’s curvature by incorporating a geometric

constraint term in the optimization objective. This term penalizes deviations from the manifold’s intrinsic curvature, preventing the system from overfitting to local features and preserving surface coherence. The implementation of the hash-grid and one-blob encodings requires careful handling of the SPD manifold’s logarithmic map. The logarithmic map  $\log(X)$  maps a matrix  $X \in \mathbb{R}^{n \times n}$  to its log-determinant space, ensuring compatibility with the manifold’s Riemannian metric. For real-time performance, we employ a spatially adaptive hash-grid, where grid cells are dynamically resized based on the density of features, minimizing redundancy while maintaining resolution. The one-blob encoding is parameterized using a **geometric kernel** that incorporates the manifold’s curvature, allowing for efficient updates during SLAM.

**Optimization Strategy** The optimization process is formulated as a **stochastic variational problem** to balance geometric fidelity and computational efficiency. The objective function combines two terms:

1. **Geometric fidelity**: A term derived from the wrapped Gaussian distribution, ensuring the estimated pose and feature positions align with the manifold’s curvature.

2. **Data consistency**: A term that minimizes the discrepancy between observed feature positions and the estimated positions, enforced via a **stochastic gradient descent (SGD)** algorithm. To accelerate convergence, we employ **batched SGD** with a **dynamic learning rate** that adapts to the manifold’s curvature. The optimization is further optimized using numerical linear algebra techniques, such as Cholesky decomposition for the SPD manifold’s metric and sparse matrix operations to handle large-scale feature data. The algorithm is implemented in a **CUDA-accelerated framework** to ensure real-time performance, with each iteration involving:

- A **logarithmic map** for feature projection,
- A **geometric constraint update** for the one-blob encoding,
- A **stochastic gradient step** for pose optimization.

**Computational Efficiency and Real-Time Constraints** The framework’s computational efficiency is critical for real-time SLAM. The hash-grid parameterization reduces the effective dimensionality of the feature space from  $O(n^2)$  to  $O(n)$ , enabling rapid updates. However, high-frequency local features may introduce numerical instability, necessitating a **dynamic regularization term** in the optimization. This term scales with the feature density, preventing overfitting while maintaining geometric consistency. To ensure real-time performance, we implement low-latency communication between the hash-grid and one-blob encodings, leveraging parallel processing and memory-efficient data structures. The use of **logarithmic maps** and **sparse matrices** further reduces memory overhead, allowing the system to handle large-scale environments with minimal computational resource usage.

### Critical Considerations

- **Numerical Stability**: The logarithmic map and SPD manifold’s curvature may introduce numerical errors, particularly for ill-conditioned matrices. We address this by incorporating a **numerical stabilization term** in the optimization, which dampens oscillations in the gradient.

- **Scalability**: The framework’s performance scales with the number of features, but high-dimensional data (e.g., 3D point clouds) may require **approximate manifold learning** techniques to maintain efficiency.

- **Alternative Approaches**: While the SPD manifold provides geometric fidelity, methods like **differential geometry** or **manifold learning** offer flexibility in handling non-Euclidean data. However, these approaches often require more complex preprocessing or may not align as closely with the probabilistic constraints of the wrapped Gaussian. By integrating the SPD manifold’s geometric structure with the wrapped Gaussian’s probabilistic framework, this implementation enables a real-time SLAM system that dynamically adapts to complex environments while preserving geometric consistency and computational efficiency. The framework’s modular design allows for further extensions, such as incorporating **multi-sensor fusion** or **dynamic environment modeling**.

**Conclusion** The proposed algorithm combines the geometric rigor of the SPD manifold with the probabilistic robustness of the wrapped Gaussian, offering a novel approach to real-time SLAM. Through detailed implementation and optimization strategies, the framework addresses key challenges in high-frequency feature tracking and geometric consistency, paving the way for scalable and accurate SLAM systems in dynamic environments.