A Survey of Visual Prompt Tuning Methods: Analyzing Retrieval, Efficiency and Memory Systems

Your Name
Department of Computer Science
University of Illinois at Urbana-Champaign
netid@illinois.edu

February 13, 2025

Abstract

We present a comprehensive survey of visual prompt tuning (VPT) methods and their applications in building efficient and scalable language-vision systems. This survey specifically examines the intersection of VPT with retrieval mechanisms and memory architectures in large language models (LLMs). We analyze various approaches including Learning to Prompt (L2P), CoDA-Prompt, and related methods through three key lenses: cost efficiency, scalability, and privacy preservation. Our analysis reveals important patterns in how different architectures handle the retrieval-computation trade-off and manage memory updates. We also propose a taxonomy for categorizing these systems based on their retrieval methods and memory update schemas, providing insights for future research directions in building more efficient and private visual-language systems.

1 Introduction

Visual prompt tuning has emerged as a powerful paradigm for adapting large vision-language models to downstream tasks while maintaining efficiency and scalability. As these systems become more prevalent, understanding their memory mechanisms and retrieval methods becomes crucial for building practical applications. This survey examines the landscape of VPT methods with a particular focus on:

- The retrieval mechanisms employed by different architectures

- Memory update schemas and their implications for continual learning
- Trade-offs between computational efficiency and model performance
- Privacy considerations in retrieval and storage systems

Recent work in LLM-agent memory systems provides valuable insights for analyzing these aspects of VPT methods. By examining this connection, we aim to bridge the gap between visual prompting and efficient memory management in multi-modal systems.

2 Research Questions

This survey aims to address the following key questions:

1. How do different VPT methods implement and optimize their retrieval mechanisms?
2. What are the efficiency-performance trade-offs in various memory update schemas?
3. How do different architectures handle privacy concerns in their retrieval systems?
4. What patterns emerge in the relationship between memory architecture and model scalability?

3 Methodology

Our analysis will focus on three main categories:

3.1 Retrieval Method Analysis

We will examine different retrieval mechanisms including:

- Direct prompt retrieval methods (L2P)
- Continuous prompt optimization (CoDA-Prompt)
- Hybrid approaches combining discrete and continuous methods

3.2 Memory Schema Comparison

We will analyze memory update mechanisms across different architectures:

- Static vs. dynamic memory structures
- Efficient update strategies
- Memory compression techniques

3.3 Efficiency Metrics

We will evaluate systems based on:

- Computational cost per inference
- Memory usage and scaling properties
- Privacy preservation capabilities

4 Expected Contributions

This survey aims to make the following contributions:

1. A comprehensive taxonomy of VPT methods based on their retrieval and memory architectures
2. Quantitative analysis of efficiency-performance trade-offs across different approaches
3. Guidelines for selecting appropriate VPT methods based on specific application requirements
4. Identification of open challenges and future research directions

5 Timeline and Milestones

- Week 1-2: Literature review and initial categorization
- Week 3-4: implement related backbone and dataset
- Week 5-8: loading all method into one code-base
- Week 9-10: running experiments
- Week 11-14: writing results

6 Broader Impact

This research will provide valuable insights for building more efficient and private visual-language systems. Understanding these trade-offs is crucial as these systems become more prevalent in real-world applications.