COSMO LENS

FINE-TUNING VISION-LANGUAGE MODELS FOR COMPREHENSIVE ASTRONOMICAL ANALYSIS

MOTIVATION

The astronomical community generates vast amounts of visual data through telescopes, satellites, and observatories. While Vision-Language Models (VLMs) have achieved state-of-the-art performance in general computer vision tasks, there exists a significant gap in their application to specialized astronomical data. Current approaches primarily focus on narrow classification tasks using fine-tuned vision encoders like CLIP, but lack the comprehensive capabilities required for diverse astronomical analysis tasks.

THE FIELD NEEDS A UNIFIED, FOUNDATIONAL VLM SPECIFICALLY ADAPTED FOR ASTRONOMY THAT CAN HANDLE MULTIPLE TASKS INCLUDING:

- VISUAL QUESTION ANSWERING (VQA) FOR ASTRONOMICAL PHENOMENA
- GALAXY MORPHOLOGY CLASSIFICATION
- SUPERNOVA ANALYSIS AND CHARACTERIZATION
- REDSHIFT ESTIMATION
- TIME SERIES DATA INTERPRETATION
- OBJECT DETECTION IN CELESTIAL IMAGES
- MULTI-MODAL UNDERSTANDING OF ASTRONOMICAL DATA

OBJECTIVES

1. Develop a specialized astronomical VLM by fine-tuning an open-source model (Qwen 2.5 VL) on comprehensive astronomical data
2. Create a multi-task learning framework that can simultaneously handle various astronomical analysis tasks
3. Establish new benchmarks for astronomical visual understanding tasks
4. Demonstrate the model's capability to generalize across different types of astronomical data
5. Provide an open-source solution for the astronomical community

NOVELTY OF APPROACH

1. Comprehensive Multi-task Capability: Unlike existing solutions that focus on single tasks (primarily classification), our approach aims to create a foundational model capable of handling multiple astronomical analysis tasks simultaneously.

2. Unified Architecture: We propose integrating various astronomical analysis capabilities into a single model, reducing the need for multiple specialized models.

3. Time Series Integration: Including time series data analysis capabilities in a VLM for astronomy is a novel approach that could benefit various astronomical studies, particularly in supernova and variable star research.

4. Large-scale Astronomical Dataset Utilization: Leveraging the multimodal universe dataset for comprehensive fine-tuning, going beyond the limited datasets used in previous studies.

METHODOLOGY

1. Data Preparation
- Curate and preprocess data from the multimodal universe dataset
- Create task-specific subsets for different astronomical analyses
- Develop standardized input formats for various types of astronomical data

2. Model Architecture
- Base model: Qwen 2.5 VL
- Modifications to handle astronomical-specific features:
  - Enhanced time series processing capabilities
  - Multi-scale feature extraction for varying celestial objects

3. Fine-tuning Strategy
- Multi-task learning approach with task-specific heads
- Progressive fine-tuning stages:
  - Initial adaptation to astronomical domain
  - Task-specific fine-tuning
  - Multi-task optimization

4. Evaluation Framework
- Develop comprehensive evaluation metrics for each task
- Create benchmark datasets for astronomical VQA
- Implement comparison with existing specialized models

LITERATURE SURVEY

Multimodal Universe dataset

E. Angeloudi et al., "The Multimodal Universe: Enabling Large-Scale Machine Learning with 100TB of Astronomical Scientific Data," in The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2024.

AstroCLIP

F. Lanusse et al., "AstroCLIP: Cross-Modal Pre-Training for Astronomical Foundation Models," in NeurIPS 2023 AI for Science Workshop, 2023.

Maven: A Multimodal Foundation Model for Supernova Science

G. Zhang, T. Helfer, A. T. Gagliano, S. Mishra-Sharma, and V. A. Villar, "Maven: A Multimodal Foundation Model for Supernova Science," in NeurIPS Workshop on Time Series in the Age of Large Models, 2024.

LITERATURE SURVEY

Qwen2-VL

P. Wang et al., "Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution," arXiv:2409.12191 [cs.CV], 2024.

LoRA Fine Tuning

E. J. Hu et al., "LoRA: Low-Rank Adaptation of Large Language Models," arXiv:2106.09685 [cs.CL], 2021.

PaperCLIP

S. Mishra-Sharma, Y. Song, and J. Thaler, "PAPERCLIP: Associating Astronomical Observations and Natural Language with Multi-Modal Models," in First Conference on Language Modeling, 2024.

LITERATURE SURVEY

CosmoCLIP

R. Imam, M. T. Alam, U. Rahman, M. Guizani, and F. Karray, "CosmoCLIP: Generalizing Large Vision-Language Models for Astronomical Imaging," arXiv:2407.07315 [cs.CV], 2024.

ColPali

M. Faysse, H. Sibille, T. Wu, B. Omrani, G. Viaud, C. Hudelot, and P. Colombo, "ColPali: Efficient Document Retrieval with Vision Language Models," in Proc. 13th Int. Conf. Learn. Represent. (ICLR), 2025.

The Well Dataset

R. Ohana et al., "The Well: a Large-Scale Collection of Diverse Physics Simulations for Machine Learning," in Proc. 38th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track (NeurIPS), 2024.

EXPECTED OUTCOMES

1. Technical Outcomes
- A unified VLM capable of handling multiple astronomical analysis tasks
- State-of-the-art performance on astronomical VQA tasks
- Improved accuracy in galaxy morphology classification
- Robust performance on time series analysis tasks

2. Research Contributions
- Novel architecture for multi-task astronomical analysis
- New benchmarks for astronomical visual understanding
- Methodology for adapting general-purpose VLMs to specialized scientific domains

3. Practical Impact
- Open-source implementation for the astronomical community
- Reduced computational overhead through unified model
- Improved accessibility of advanced astronomical analysis tools

THANK YOU