

Machine Learning Experiments on Multi-Modal Astronomical Data using HiSS-Cube Framework

Farukh Rustamov
Supervised by: RNDr. Petr Škoda, CSc
CVUT
FIT

December 16, 2024

Contents

1	Introduction	2
1.1	Motivation	2
1.2	Objectives	2
2	Background	3
2.1	Astronomical Surveys and SDSS	3
2.2	HiSS-Cube Framework	3
2.3	Visual Genome: Preliminary Multi-Modal Experiments	3
2.4	Machine Learning in Astronomy	4
3	Methodology	5
3.1	Installation and Data Preparation	5
3.2	Experiment Design	5
3.3	Machine Learning Experiments	5
4	Results	6
4.1	Performance Comparison	6
4.2	Scalability Analysis	6
5	Discussion and Future Work	7
5.1	Findings	7
5.2	Future Directions	7
6	Conclusion	8
A	Appendices	9
A.1	HiSS-Cube Installation Guide	9
A.2	Scripts and Code Listings	9

Chapter 1

Introduction

1.1 Motivation

- The growth of astronomical data and the need for machine learning.
- Introducing the Sloan Digital Sky Survey (SDSS).
- Challenges in multi-modal and multi-dimensional data analysis.

1.2 Objectives

- Developing a pipeline for analyzing SDSS data using HiSS-Cube.
- Multi-modal machine learning experiments.
- Evaluate performance and scalability.

Chapter 2

Background

2.1 Astronomical Surveys and SDSS

- Overview of SDSS (images, spectra, and filters).
- Challenges in analyzing large-scale datasets.

2.2 HiSS-Cube Framework

- Architecture of HiSS-Cube.
- Features: multi-scale, semi-sparse data handling, and uncertainty estimates.

2.3 Visual Genome: Preliminary Multi-Modal Experiments

- Overview of Visual Genome dataset:
 - Combines images and text data (multi-modal).
 - Contains 108,000 high-resolution images with over 5 million annotations.
- Relevance to astronomical data analysis:
 - Images with annotated regions (similar to SDSS sky images).
 - Textual descriptions (analogous to numerical spectra).
- Example tasks:
 - Object classification and detection.
 - Semantic relationship analysis between objects.
 - Image captioning and description generation.
- Use of frameworks like TensorFlow, PyTorch for CNN and spaCy/BERT for text analysis.

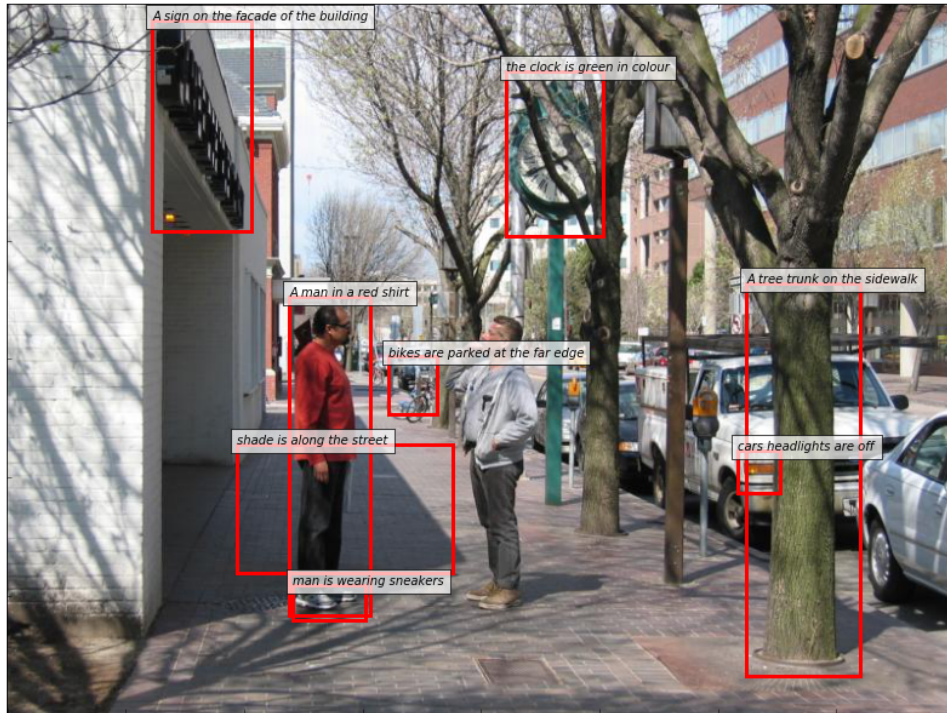


Figure 2.1: An example of Visual Genome data.

2.4 Machine Learning in Astronomy

- Role of ML methods in modern astronomy.
- Examples: classification, regression, clustering.

Chapter 3

Methodology

3.1 Installation and Data Preparation

- Step-by-step HiSS-Cube installation.
- Download and pre-process SDSS data.

3.2 Experiment Design

- Identify science cases for multi-modal analysis.
- Compare multi-modal vs single-modal performance.

3.3 Machine Learning Experiments

- Overview of ML methods to be used (e.g., CNN, t-SNE).
- Experimental setup for classification and clustering.

Chapter 4

Results

4.1 Performance Comparison

- Results of multi-modal experiments.
- Comparison with single-modal experiments.

4.2 Scalability Analysis

- Execution time on different scales.
- Precision vs execution time trade-offs.

Chapter 5

Discussion and Future Work

5.1 Findings

- Summary of key results.
- Interpretation of multi-modal ML advantages.

5.2 Future Directions

- Extend experiments to the full SDSS dataset.
- Leverage large clusters for scalability.

Chapter 6

Conclusion

- Summarize the importance of HiSS-Cube for astronomy.
- Highlight the contributions of this work.

Appendix A

Appendices

A.1 HiSS-Cube Installation Guide

A.2 Scripts and Code Listings

Bibliography