Raster Miner: An Open-Source Toolkit for Knowledge Discovery in Satellite Imagery Data

by

RAGE Uday Kiran, Makiko Ohtake, Yoshiko Ogawa

The University of Aizu

Fukushima, Japan

Outline

- Introduction
- Importance of Imputation
- Problem Definition
- Proposed Approach
- Experimental Results
- Conclusions

Outline

- **Introduction**
- Importance of Imputation
- Problem Definition
- Proposed Approach
- Experimental Results
- Conclusions

Introduction

- Raster Miner is an open-source software for discovering knowledge hidden in Raster Data.
- Availability: GitHub
- Liscence: GNU V3
- Platforms: Windows and Mac
- Type of Execution: GUI, Terminal, and Python.

Knowledge Discovery Tasks

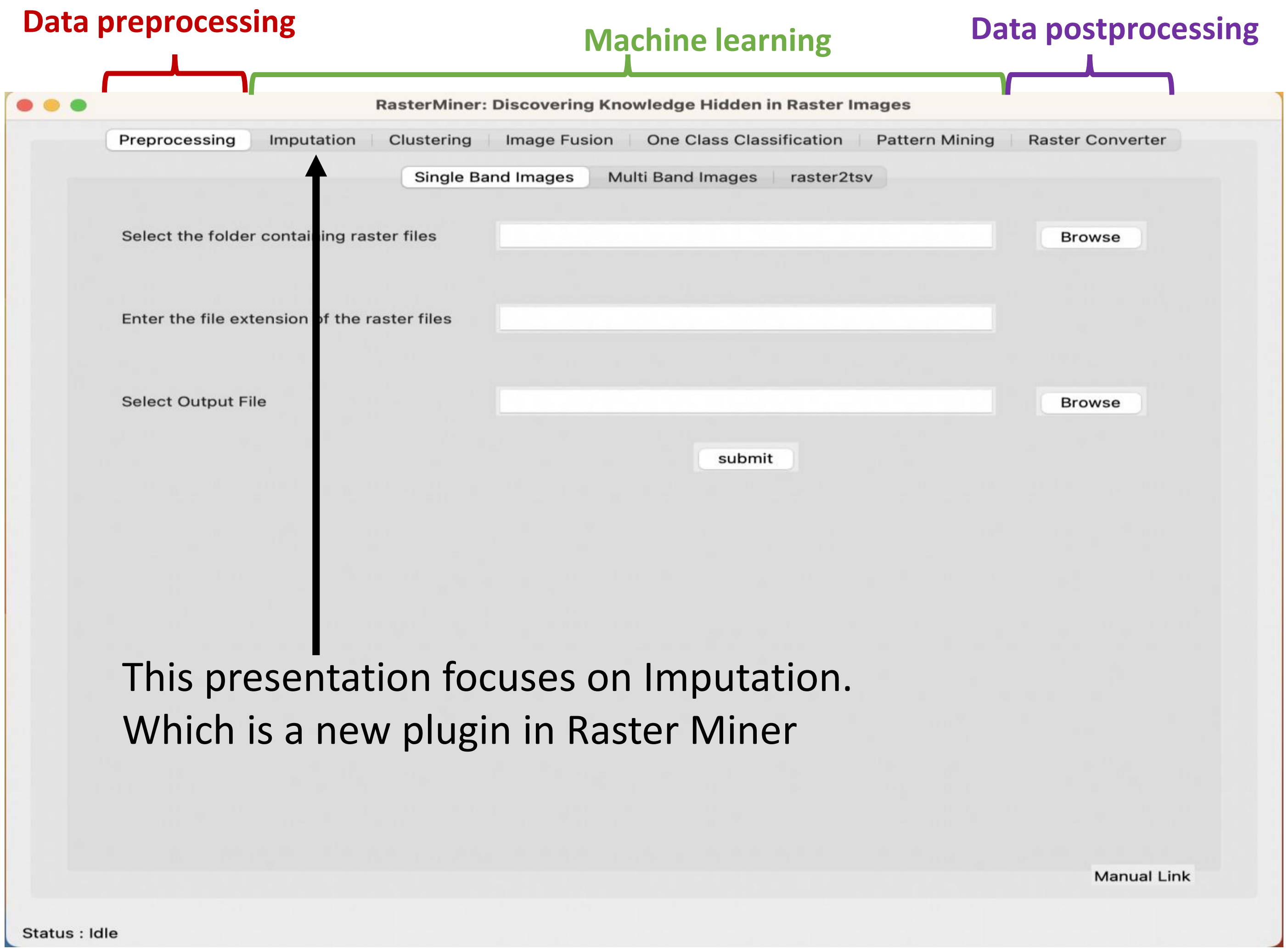


Fig. 1. Front-end of Raster Miner

Data preprocessing:

- Converts raster images into tsv files
- Support various formats of raster images
  - lbl, tiff, geotiff …

Machine learning:

- Contains techniques for various ML tasks
- ML topics currently available are:
  1. Imputation
  2. Clustering
  3. Image Fusion
  4. One Class Classification
  5. Pattern Mining

Data Postprocessing:

Converts the tsv data back into raster/tiff format

Imputation

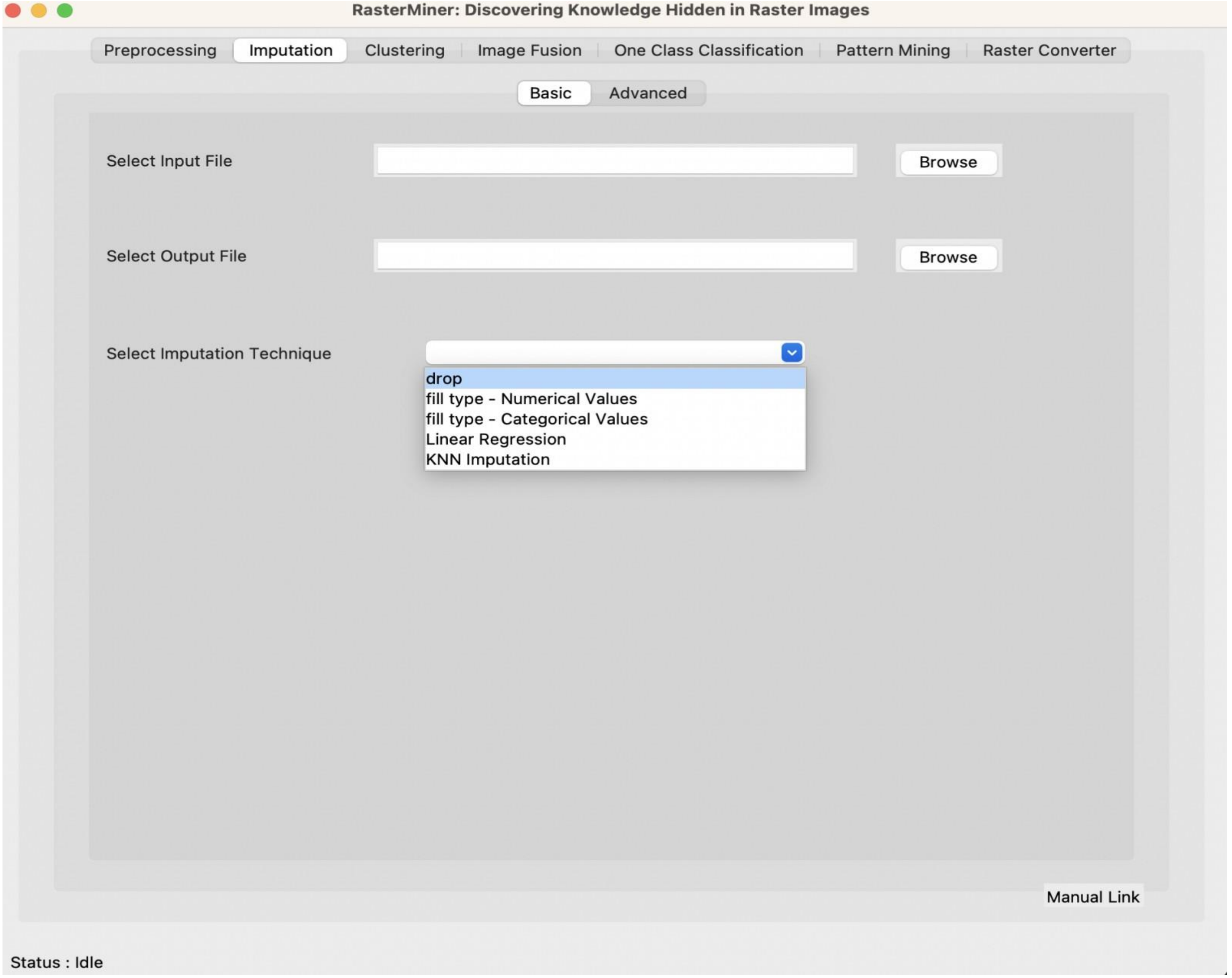


Fig. 2a. Basic Imputation Techniques

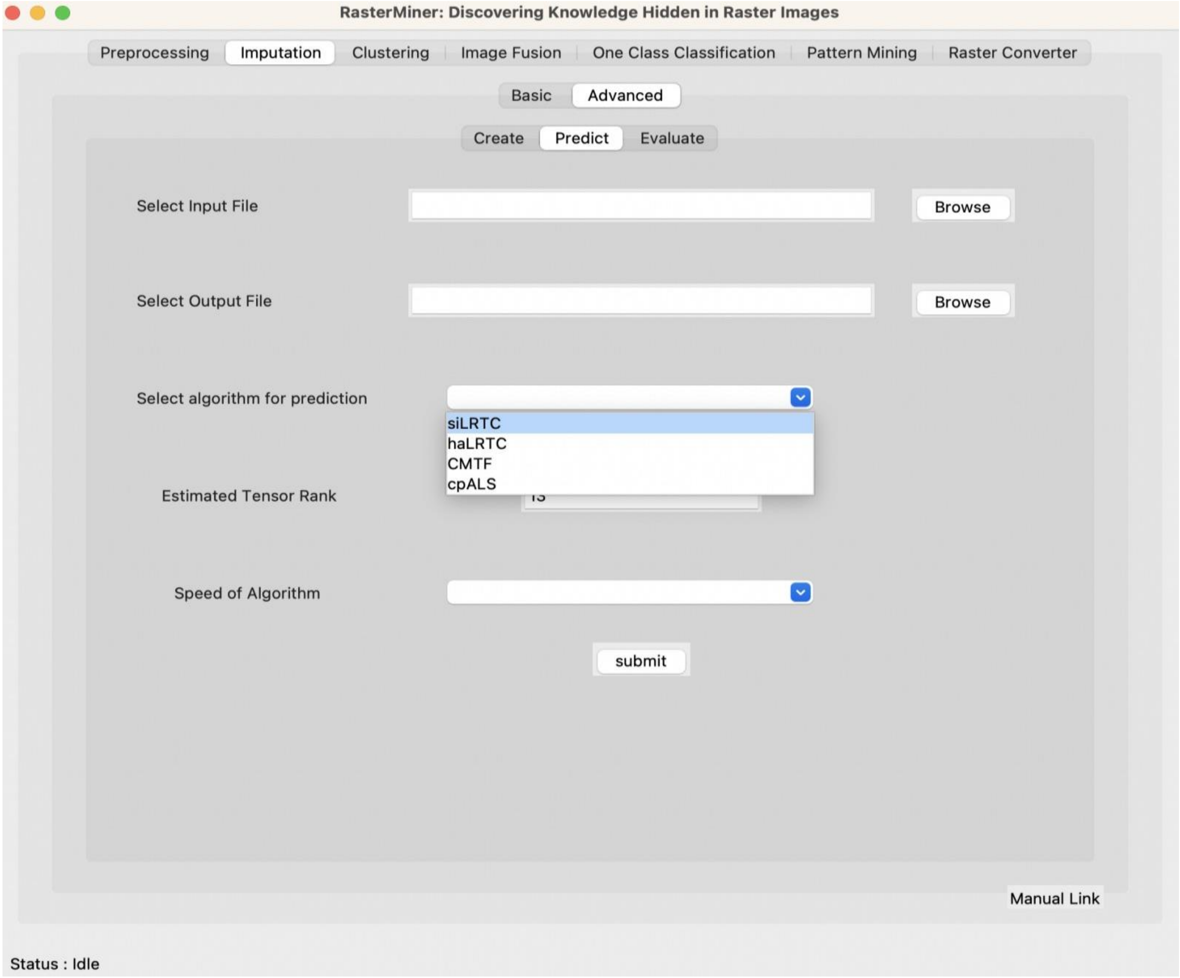


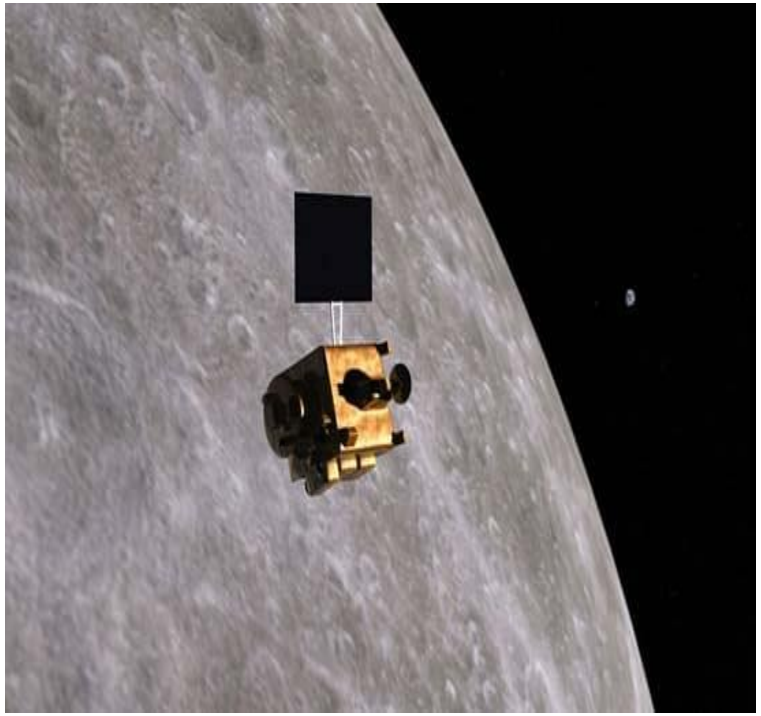
Fig. 2b. Advanced Tensor-based Imputation Techniques

Outline

- Introduction
- **Importance of Imputation**
- Problem Definition
- Proposed Approach
- Experimental Results
- Conclusions

Introduction: A New Space Race

Kaguya Satellite



Chandrayan-1 Satellite

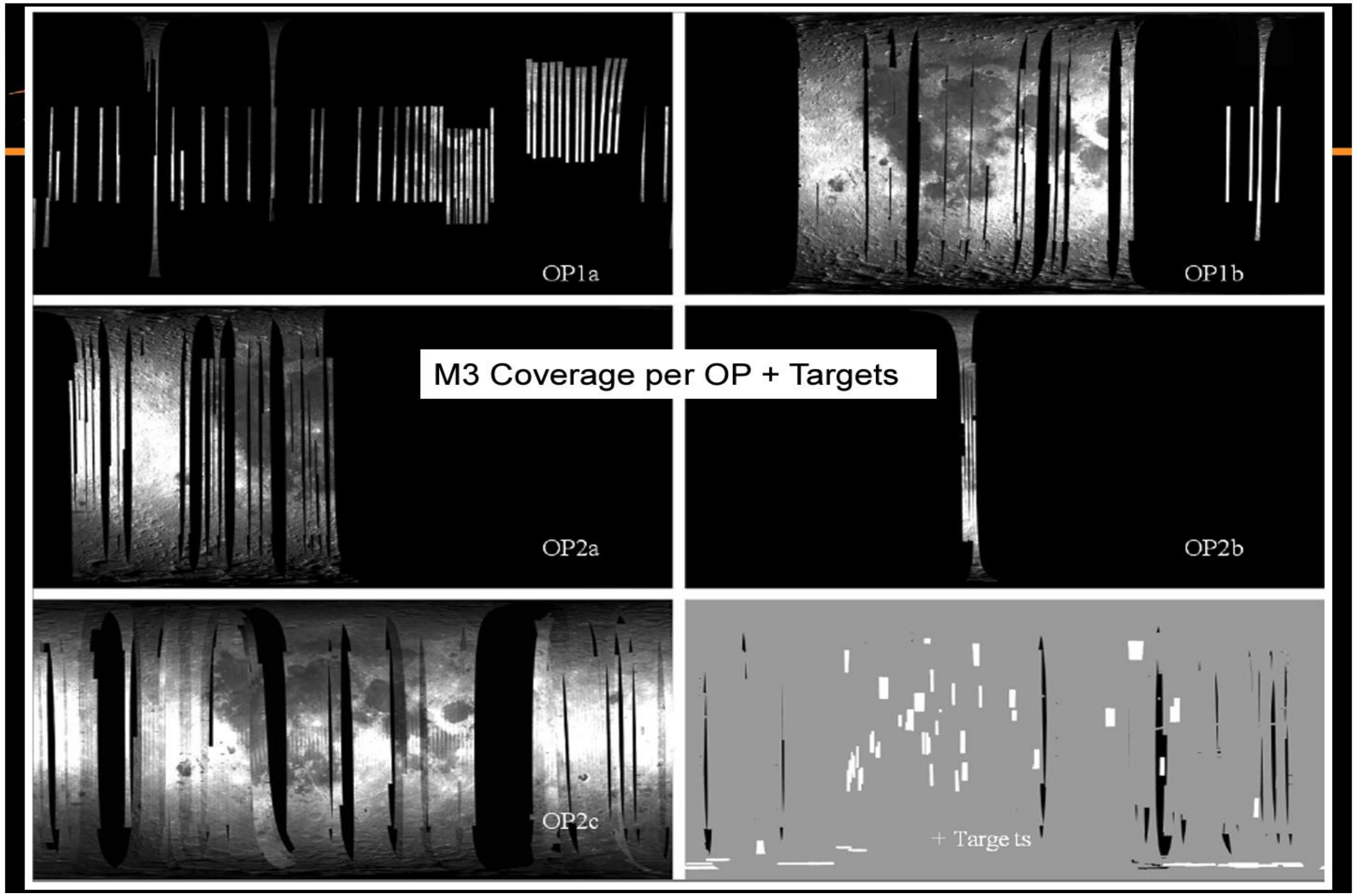
Chandrayan-2 Satellite

Fig. 1. Satellites Collecting Lunar Surface Data

Introduction: Missing Pixel Data Problem

Fig.4. Raster data of Moon’s surface gathered by Kaguya Satellite

Introduction: Missing Pixel Data Problem



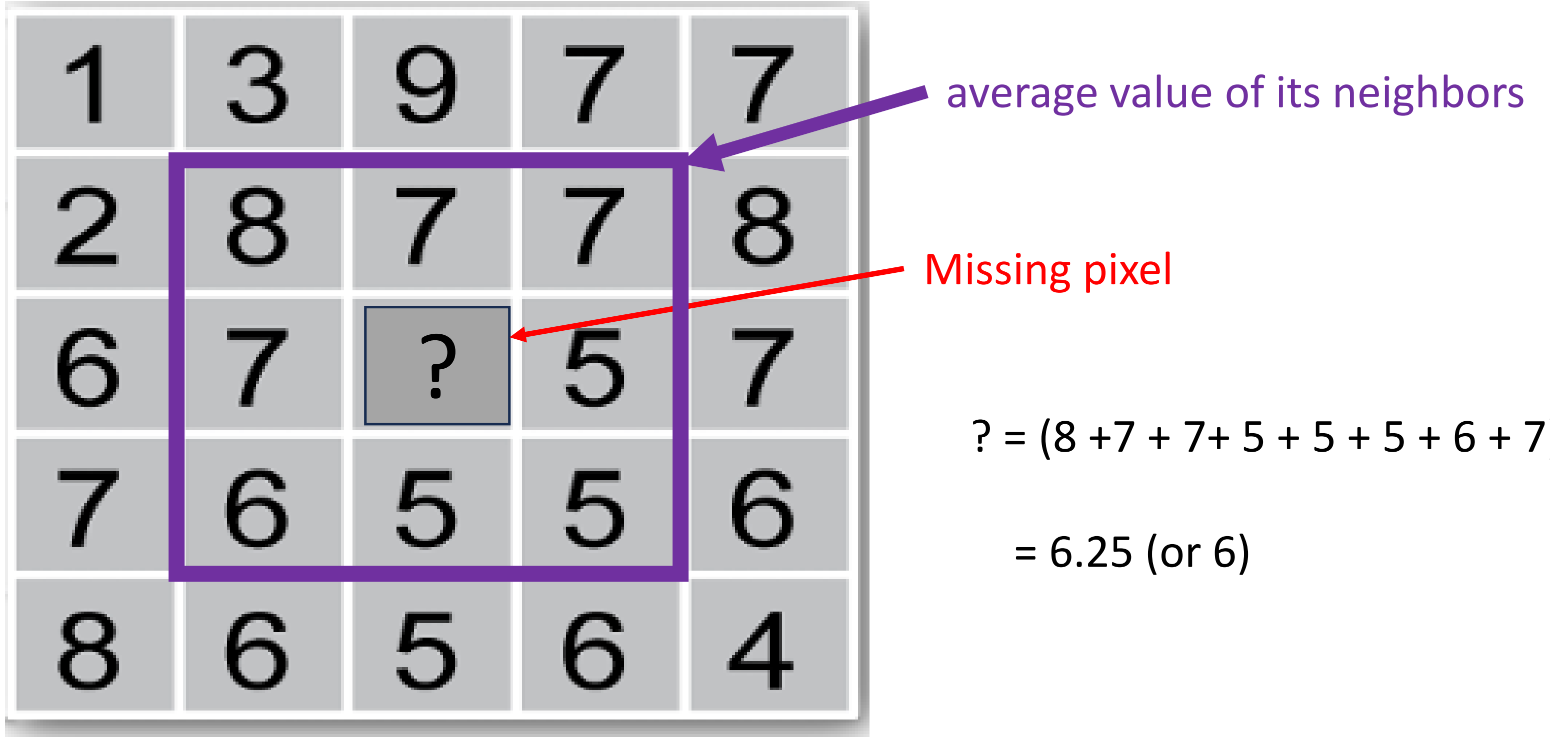
Missing data is shown in black

Fig. 5. Raster data of Moon's surface gathered by Chandrayan Satellite

Source: https://pds-imaging.jpl.nasa.gov/documentation/Isaacson_M3_Workshop_Final.pdf

Introduction: Existing Solutions and Their Limitations

- k-Nearest Neighbors Solution
  - fill in the missing pixel value using its neighbors value



? = (8 +7 + 7+ 5 + 5 + 5 + 6 + 7)/8

= 6.25 (or 6)

- **Limitation:** Cannot be applied if most of the data is missing

Introduction: Existing Solutions and Their Limitations

- Machine Learning
  - Neural Networks
  - GANs

- Limitation:
  - Need much data for model building.
  - Unfortunately, not much data is available

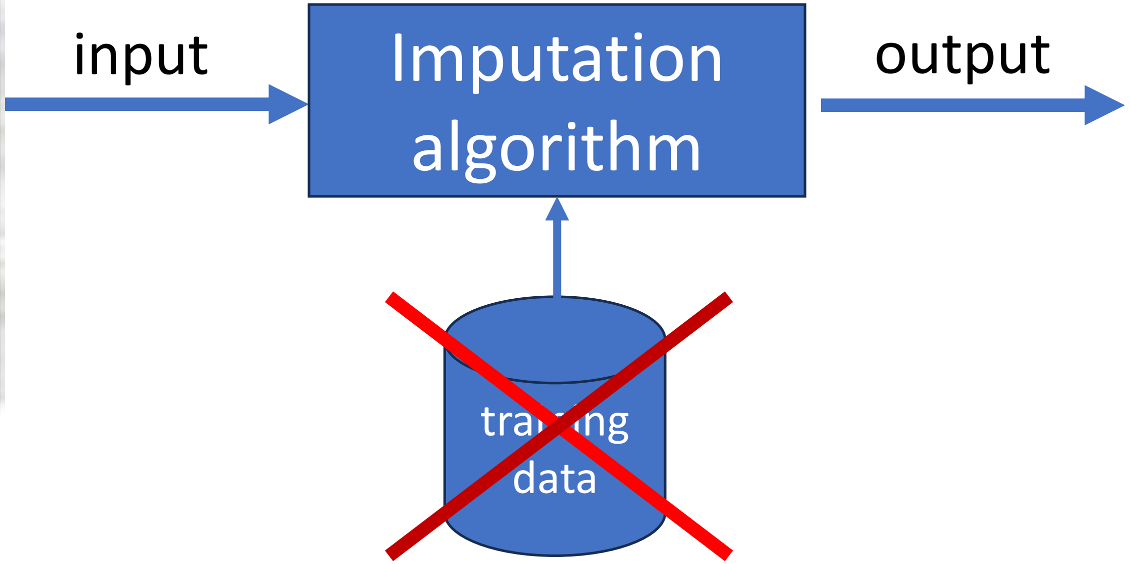
Outline

- Introduction
- Importance of Imputation
- **Problem Definition**
- Proposed Approach
- Experimental Results
- Conclusions

Problem Definition

- Predict the missing data in a highly corrupt image that has little training data

corrupt image



imputed image

Outline

- Introduction
- Importance of Imputation
- Problem Definition
- **Proposed Approach**
- Experimental Results
- Conclusions

Proposed Approach: Tensor Completion

- Idea:
  - Model raster image as a tensors using Tucker decomposition and Canonical Polyadic decomposition

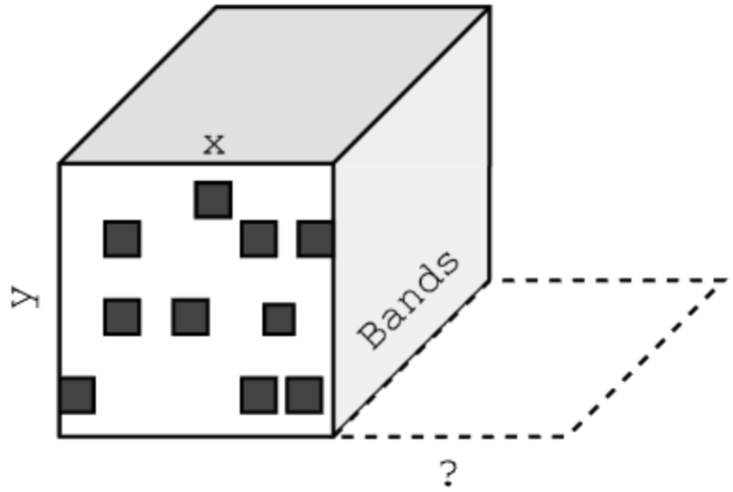


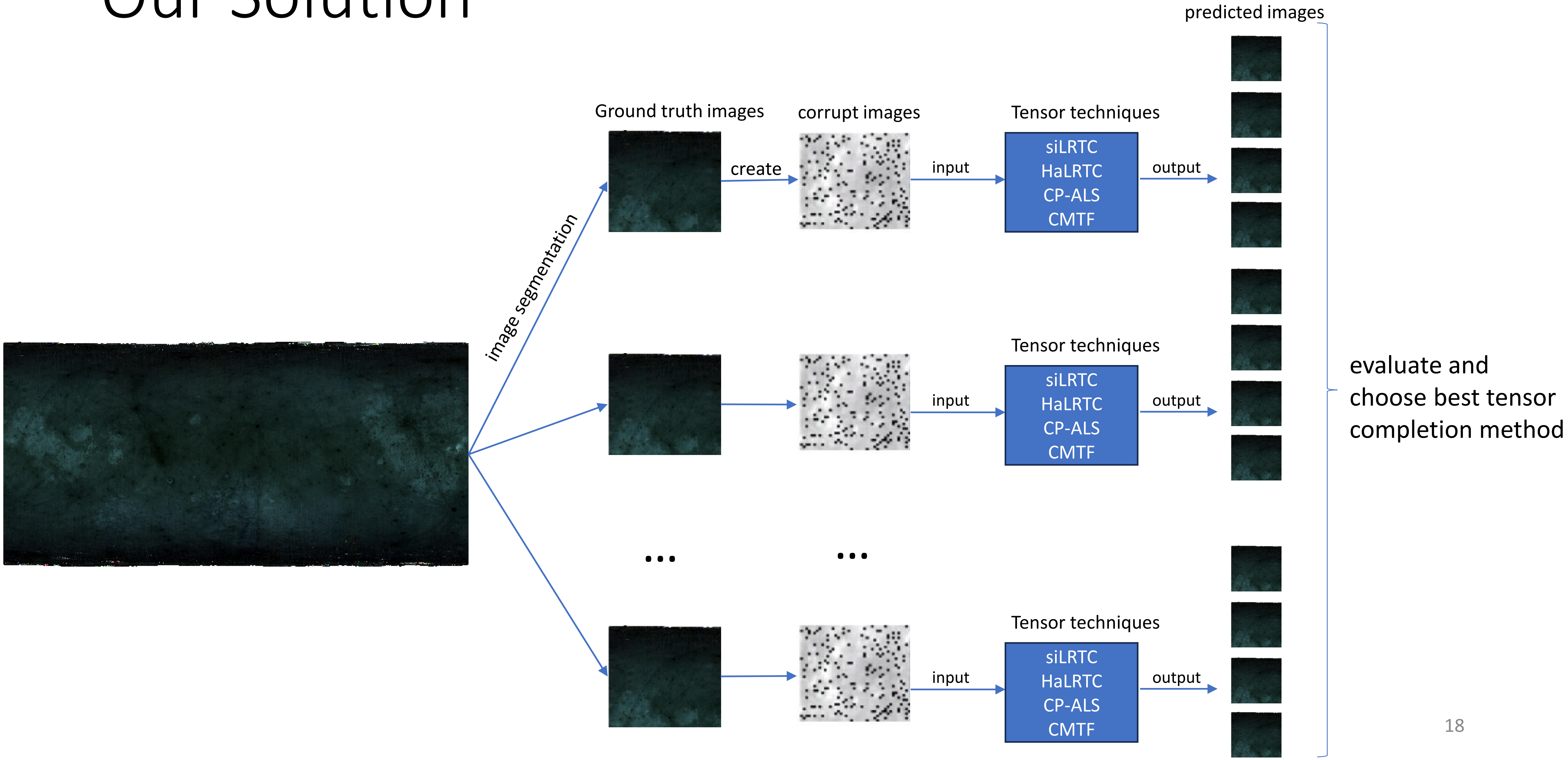
Fig. 6. Raster Data as Tensors

Challenges

- Several Tensor Solutions exists
  - siLRTC (simple Low-Rank Tensor Completion)
  - HaLRTC (High-accuracy Low-Rank Tensor Completion)
  - CP-ALS (Canonical Polyadic - Alternating Least Squares)
  - CMTF-OPT (Coupled Matrix and Tensor Factorization -Optimization)

- No universally accepted best solution exists for predicting missing data for any given dataset

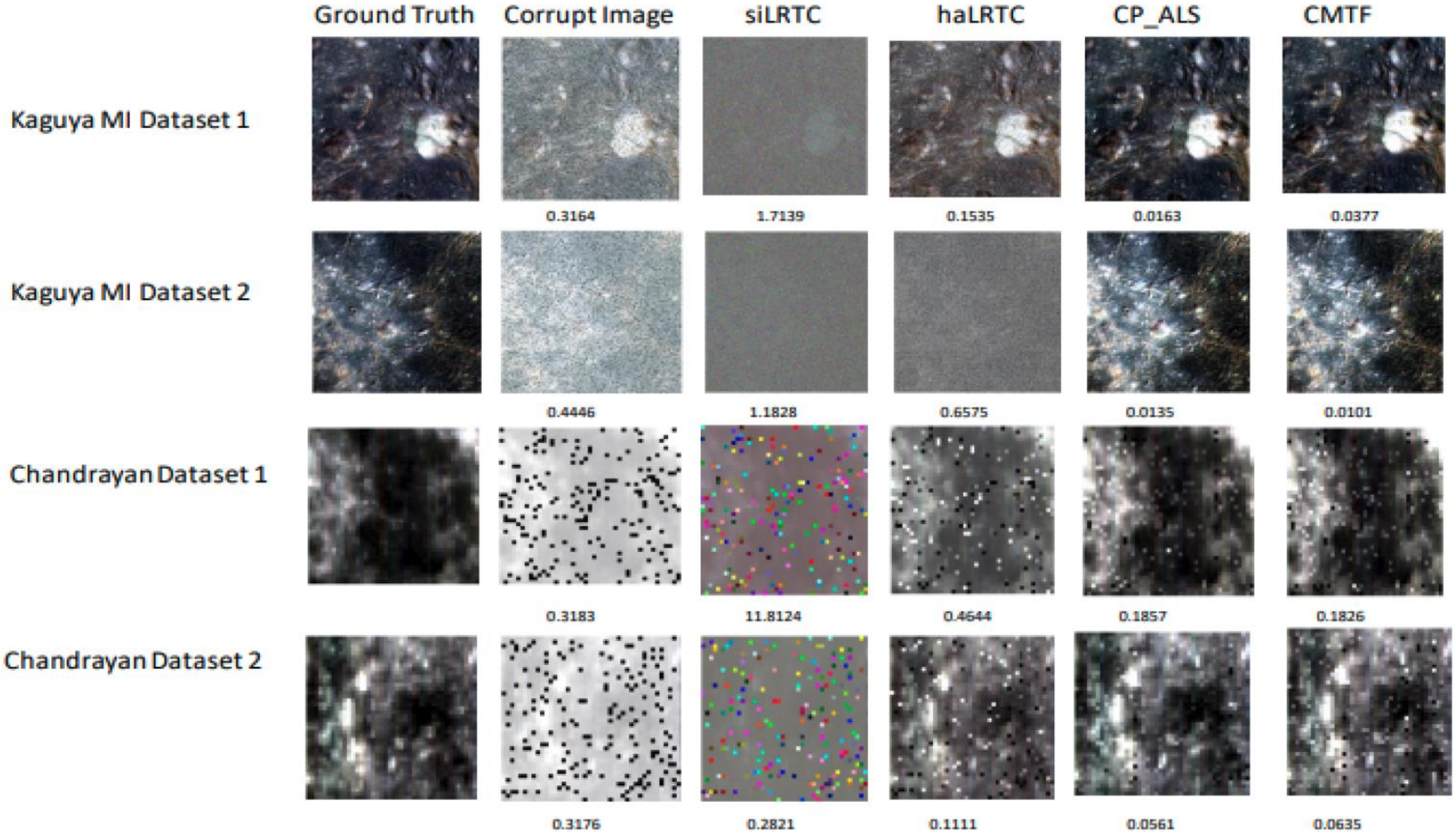
Our Solution



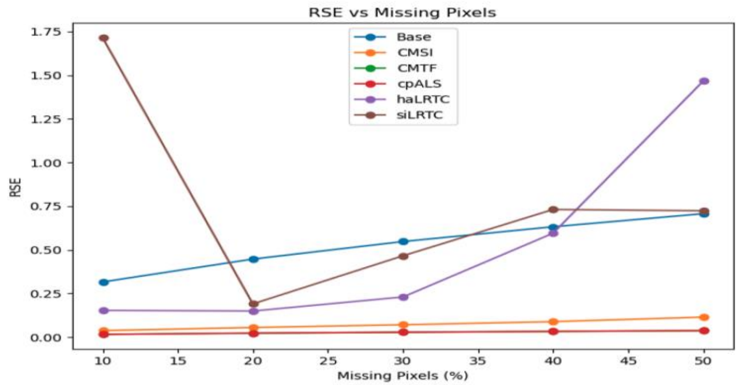
Outline

- Introduction
- Problem Definition
- Challenges and Proposed Approach
- **Experimental Results**
- Conclusions

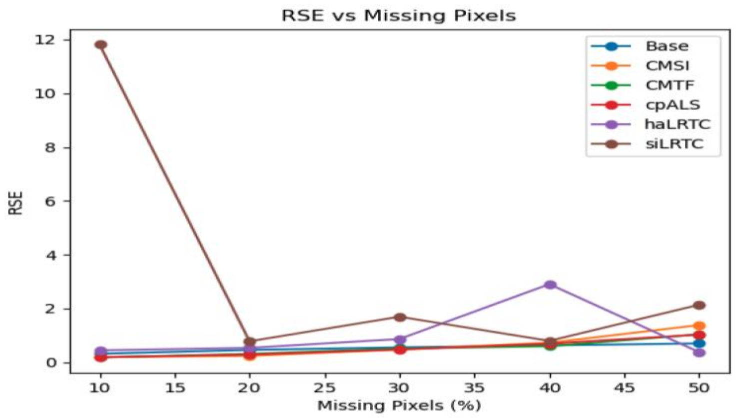
Experimental Results



Experimental Results

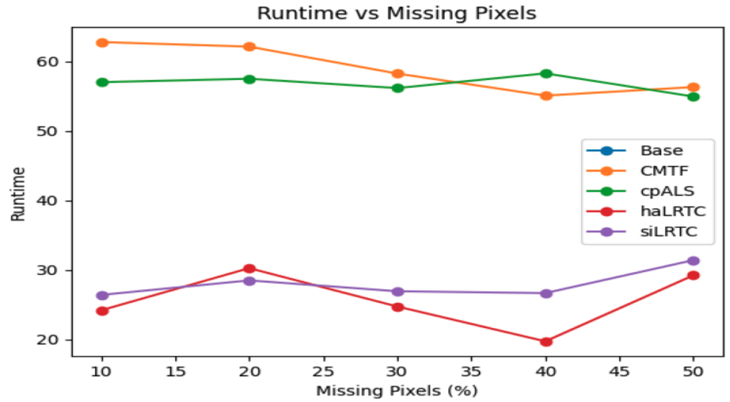


(a) Kaguya Dataset 1

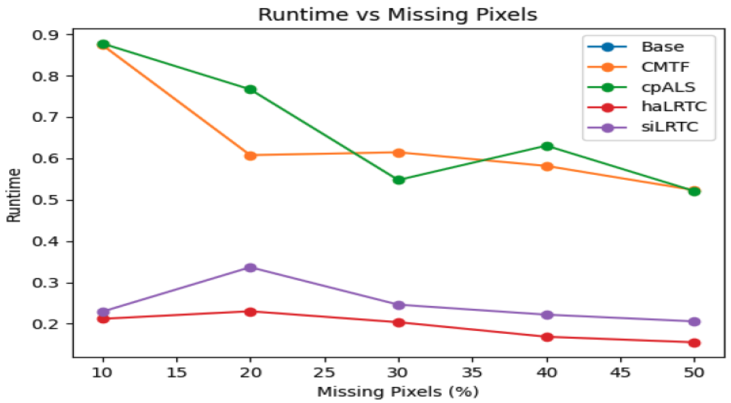


(b) Chandrayan Dataset 1

Fig. 4. Comparison between various approaches on varying the percentage of missing pixels on Kaguya Dataset 1(left) and Chandrayan Dataset 1(right)



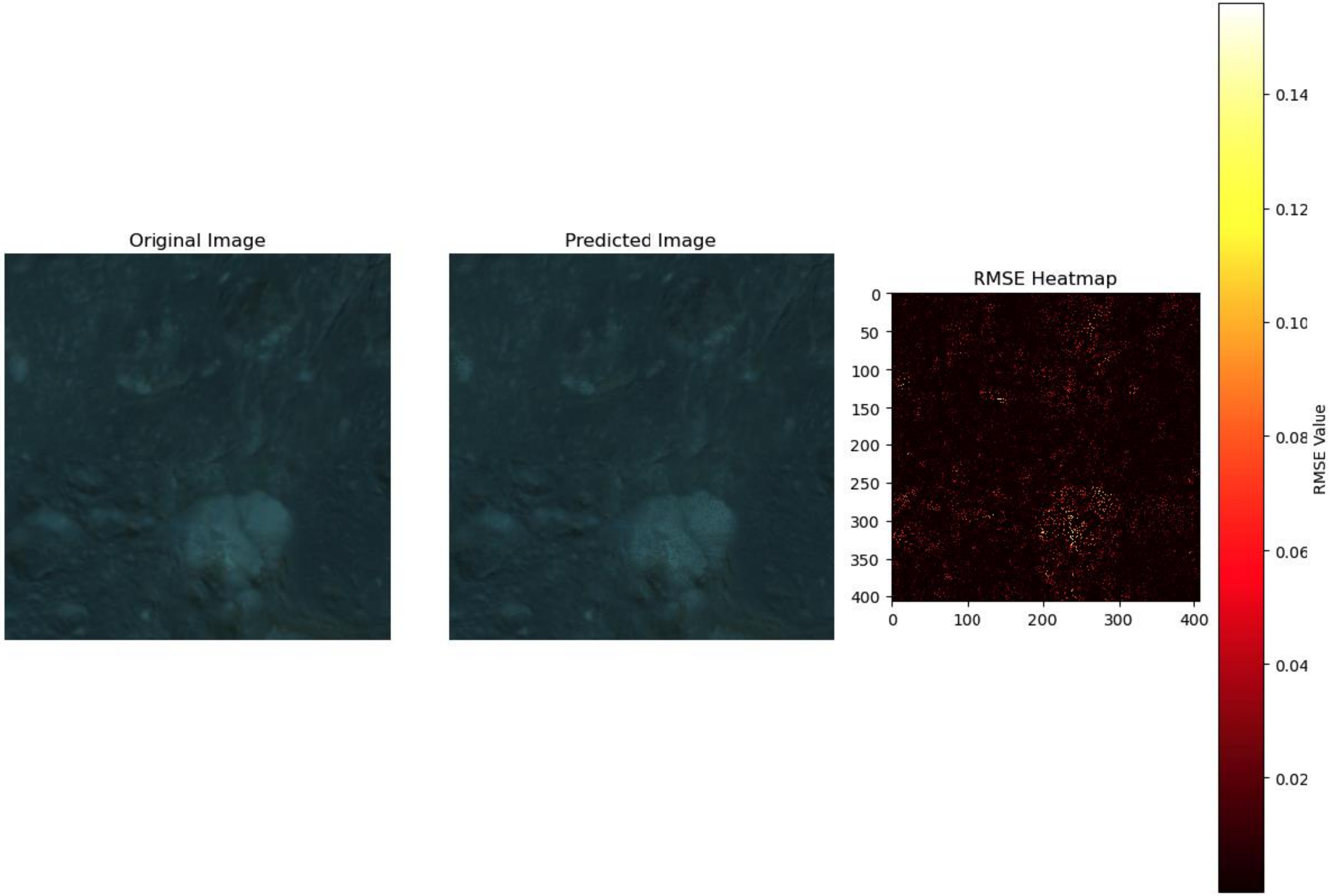
(a) Kaguya Dataset 1



(b) Chandrayan Dataset 1

Fig. 5. Study to compare the run time of different approaches on varying the percentage of missing pixels on Kaguya Dataset 1(left) and Chandrayan Dataset 1(right)

Experimental Results



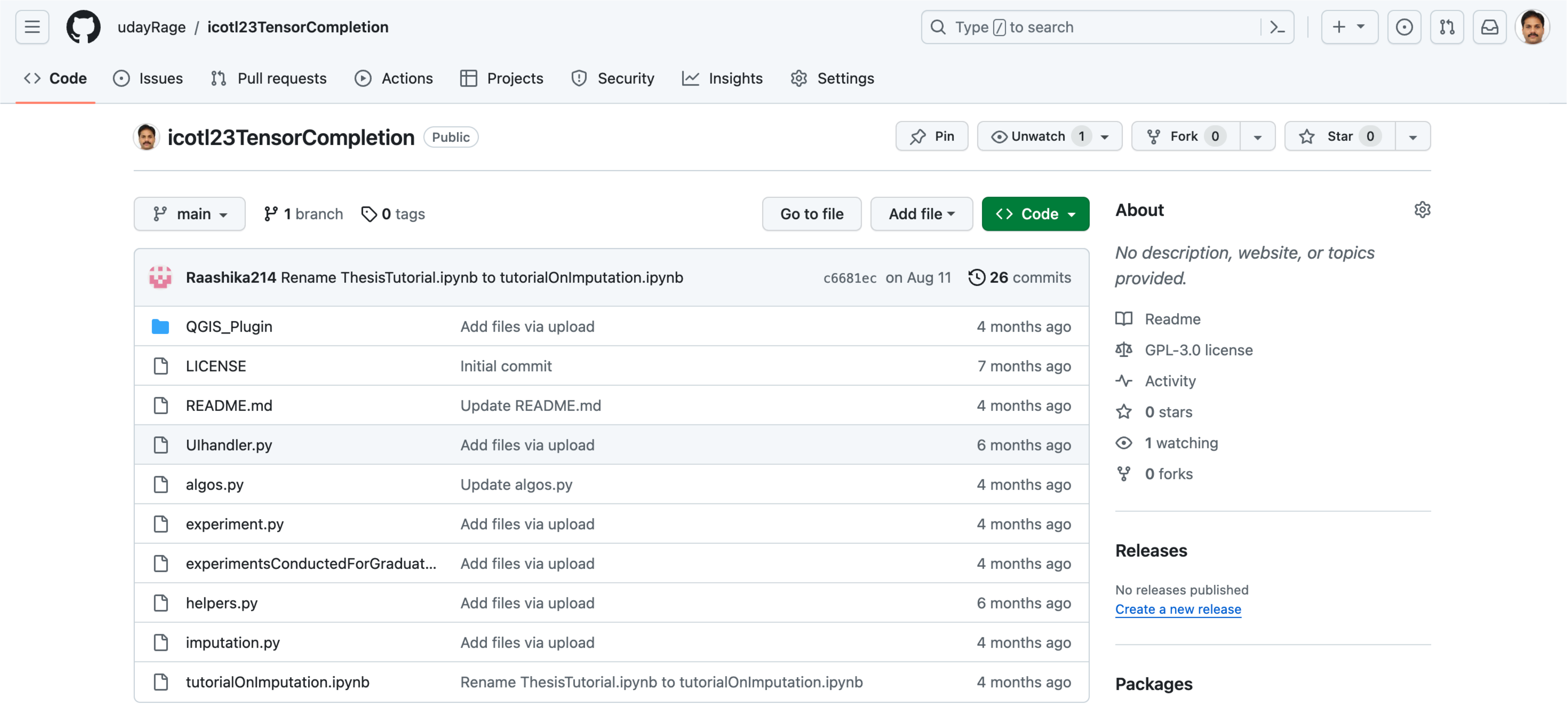
Outline

- Introduction
- Problem Definition
- Challenges and Proposed Approach
- Experimental Results
- **Conclusions**

Conclusions

- This paper tackled the missing pixel problem in raster images using Tensor Completion Technique.
- Experimental results demonstrate that CMTF-OPT technique performed better against other imputation techniques.
- CMTF-OPT was found to be computationally expensive than most of the tensor-based imputation techniques.

https://github.com/udayRage/icotl23TensorCompletion



Thank you

&

Questions