Content-Based Search of Large Image Archives at the PDS Imaging Node Motivation of Content-Based Search

Steven Lu[1], Kiri L Wagstaff[1], Emily Dunkel[1], Bhavan Vasu[1,2], Kevin Grimes[1], and Michael Mcauley[1]

[1]Jet Propulsion Laboratory, California Institute of Technology
[2]Oregon State University

January 2, 2023

Abstract

The Planetary Data System (PDS) maintains archives of data collected by NASA missions that explore our solar system. The PDS Cartography and Imaging Sciences Node (Imaging Node) provides access to millions of images of planets, moons, and other bodies. Given the large and continually growing volume of data, there is a need for tools that enable users to quickly search for images of interest. Each image archived at the PDS Imaging Node is described by a rich set of searchable metadata properties, such as the time it was collected and the instrument used. However, users often wish to search on the content of the image to find those images most relevant to their scientific investigation or individual curiosity.

To enable the content-based search of the large image archives, we utilized machine learning techniques to create convolution neural network (CNN) classification models. The initial CNN classification results for rover missions (i.e., Mars Science Laboratory and Mars Exploration Rover) and orbiter missions (i.e., Mars Reconnaissance Orbiter, Cassini, and Galileo) were deployed at the PDS Image Atlas (`https://pds-imaging.jpl.nasa.gov/search`) in 2017. With the content-based search capability, users of the PDS Image Atlas can search using a list of pre-defined classes and quickly find relevant images. For example, users can search “Impact ejecta” and find the images containing impact ejecta from the archive of the Mars Reconnaissance Orbiter mission.

All of the CNN classification models were trained using the transfer learning approach, in which we adapted a CNN model pretrained on Earth images to classify planetary images. Over the past several years, we employed the following three techniques to improve the efficiency of collecting labeled data sets, the accuracy of the models, and the interpretability of the classification results:

· First, we used the marginal-probability based active learning (MP-AL) algorithm to improve the efficiency of collecting labeled data sets.

· Second, we used the classifier chain and ensemble approaches to improve the accuracy of the classification results.

· Third, we incorporated the prototypical part network (ProtoPNet) architecture to improve the interpretability of the classification results.

Content-Based Search of Large Image Archives at the PDS Imaging Node

Steven Lu[1], Kiri L. Wagstaff[1], Emily Dunkel[1], Bhavan Vasu[1,2], Kevin Grimes[1], Michael McAuley[1]

[1]Jet Propulsion Laboratory, California Institute of Technology; [2]Oregon State University

Poster ID 1164476, AGU Fall Meeting, Chicago IL & Online Everywhere, 13 December 2022

Contact author: you.lu@jpl.nasa.gov

Motivation of Content-Based Search

Images collected by NASA planetary science missions are curated by the Cartography and Imaging Sciences Discipline Node (Imaging Node) of the Planetary Data System (PDS). These holdings currently include more than 39.3 million products that span 22 missions, with targets that include Mars, the Moon, Mercury, Jupiter, Saturn, Venus, and more. Some example Mars surface images taken by the Curiosity rover of Mars Science Laboratory (MSL) mission are shown in Figure 1.

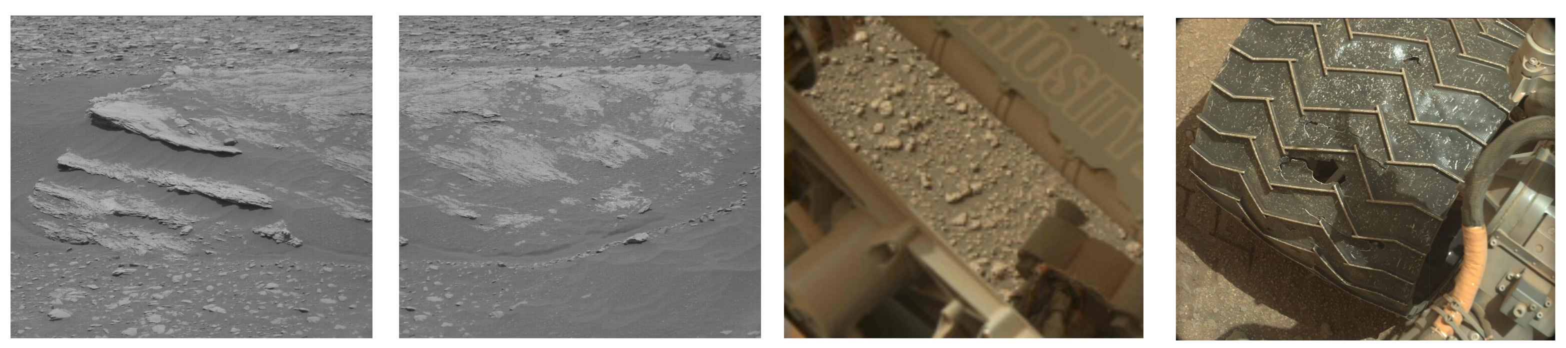


Figure 1. Example Mars images taken by the Curiosity rover of Mars Science Laboratory Mission.

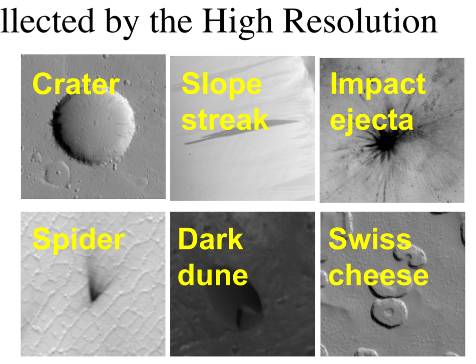
Problem: How to effectively find images of interest from 39.3 million images archived at the PDS Imaging Node? For example, how to find images taken by the Curiosity rover of MSL mission that contain wheels (the right-most image shown in Figure 1).

Solution: Content-based image classification to quickly find images of interest.

1. Machine Learning Classifiers

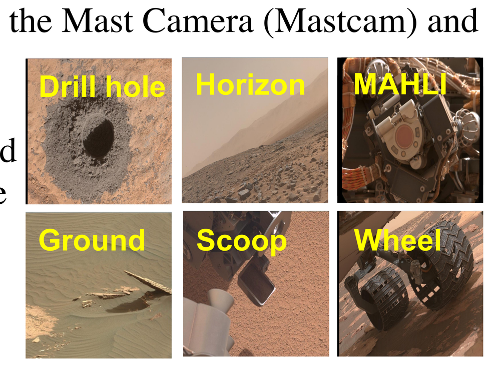
HiRISENet: CNN Classifier for Mars Orbital Images

HiRISENet (Wagstaff et al., 2021) was created to classify images collected by the High Resolution Imaging Science Experiment (HiRISE) camera onboard the Mars Reconnaissance Orbiter (MRO) mission.



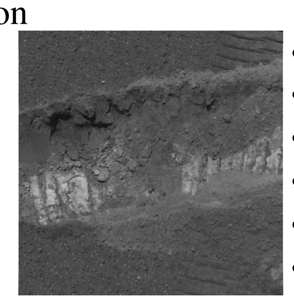
MSLNet: Hybrid Classifier for Mars Surface Images

MSLNet (Wagstaff et al., 2021) was created to classify images collected by the Mast Camera (Mastcam) and Mars Hand Lens Imager (MAHLI) instruments mounted on the Mars Science Laboratory (MSL) mission's Curiosity rover.



MERNet: Ensemble Classifier for Mars Surface Images

MERNet (Lu et al., 2022) was a multi-label convolution neural network (CNN) that classifiers images collected by the Panoramic Camera (Pancam) instrument mounted on Mars Exploration Rover (MER) mission's Spirit and Opportunity rovers. MERNet is an ensemble of 5 CNN classifiers.



- Rover Tracks
- Soil Trench
- Clasts
- Bright Soil
- Spherules
- Nearby Surface

Early work: Cassini and Galileo Classifiers

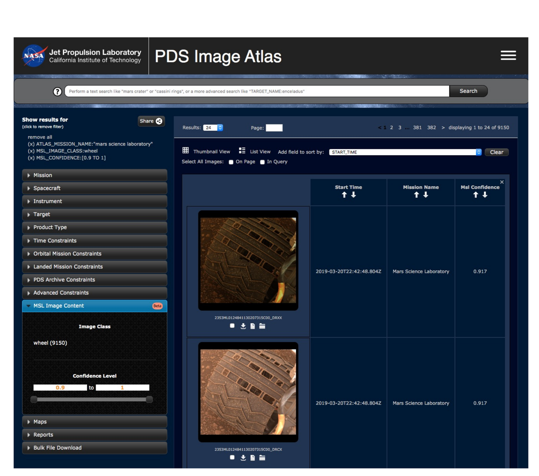
The Cassini classifier (Stanboli et al., 2017) was created for classifying images collected by the ISS instrument from the Cassini orbiter. The Galileo classifier was created to classify Europa images collected by the SSI instrument from the Galileo orbiter.

Ongoing work: LROC and Mars 2020 classifiers

For more information about the Lunar Reconnaissance Orbiter Camera (LROC) classifier, please check AGU poster 109717.

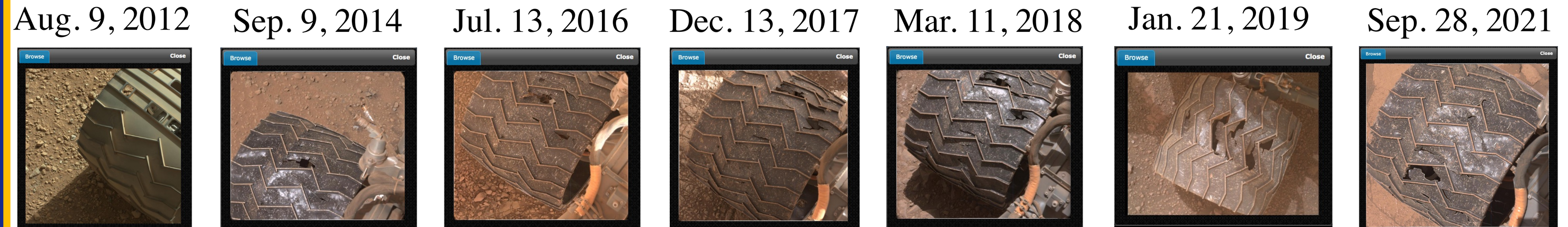
2. PDS Image Atlas

- High confidence classification results are deployed on the PDS Image Atlas.
- The PDS Image Atlas provides unique capabilities for users to find data of interest.
  - Support metadata-based search
  - Support content-based search
- The PDS Image Atlas is publicly accessible at URL https://pds-imaging.jpl.nasa.gov/search
- The next generation PDS Image Atlas will be launched soon



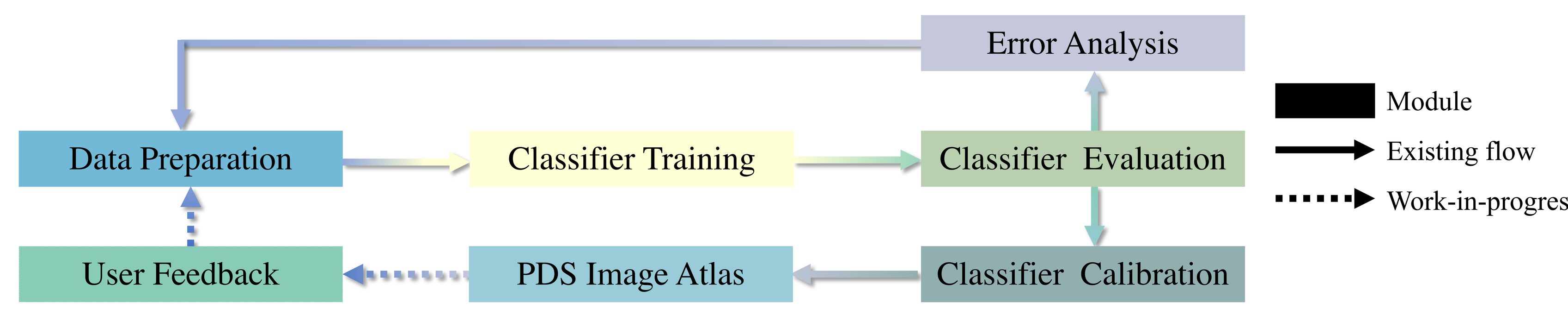
MSL rover wheel degradation over time

Aug. 9, 2012 | Sep. 9, 2014 | Jul. 13, 2016 | Dec. 13, 2017 | Mar. 11, 2018 | Jan. 21, 2019 | Sep. 28, 2021



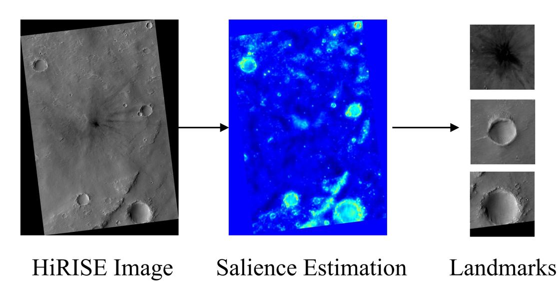
3. Machine Learning Techniques

Machine Learning Pipeline Overview



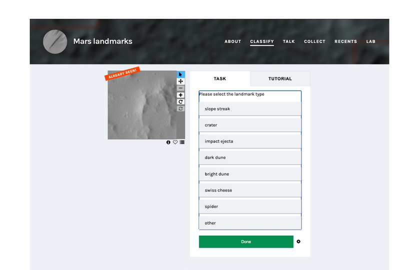
Machine Learning Data Preparation

Salient landmark detection
(Wagstaff et al., 2012)

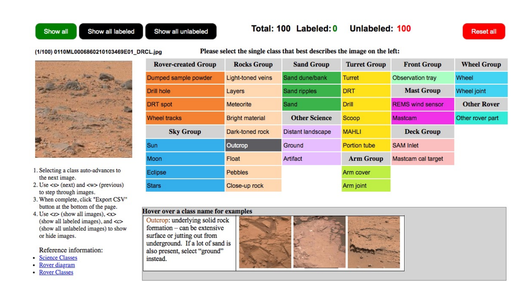


Data labeling

Zooniverse crowd source labeling



IDAR labeling tool



https://github.com/stevenlujpl/IDAR

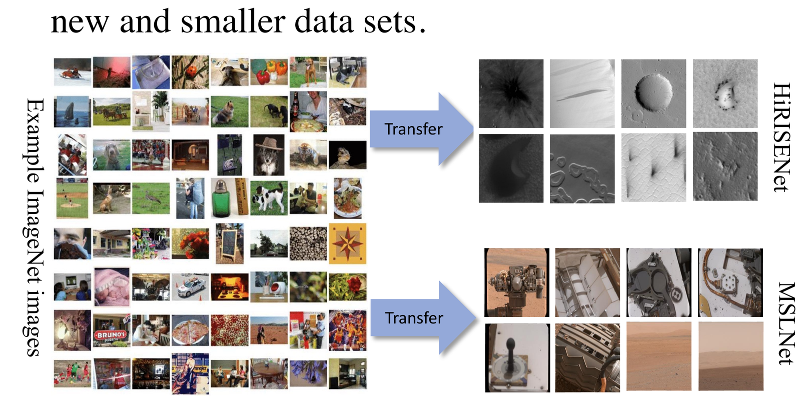
Classifier Training

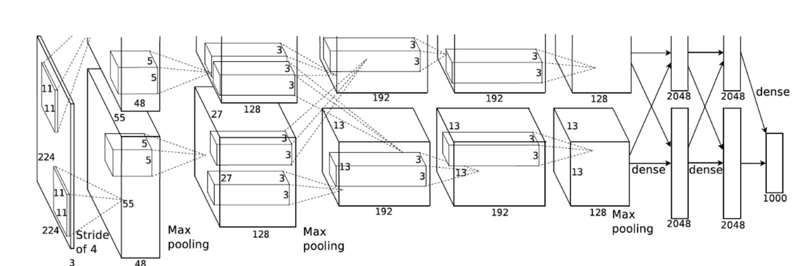
Transfer learning

- A machine learning method where a classifier developed for a task is reused and adapted as the starting point for a classifier on another task.
- In general, this is how deep learning models can be effectively trained with reasonable performance on new and smaller data sets.

Classifier fine-tuning

- Employ several thousands of labeled images from PDS archives
- Freeze or allow small variations on the weights of initial layers.
- Boost learning rate on the last layer.

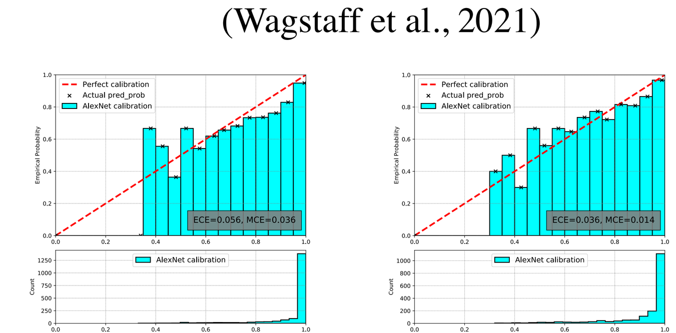




AlexNet architecture (Credit: Alex Krizhevsky)

Classifier Calibration

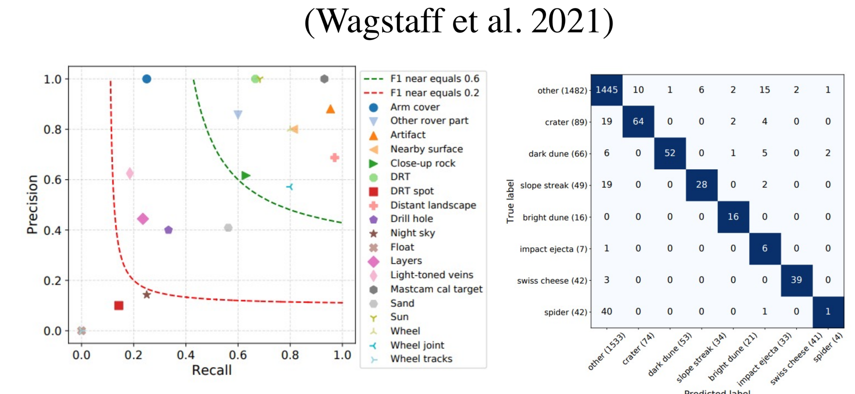
(Wagstaff et al., 2021)



Before calibration (HiRISENet) | After calibration (HiRISENet)

Classifier Evaluation

(Wagstaff et al. 2021)

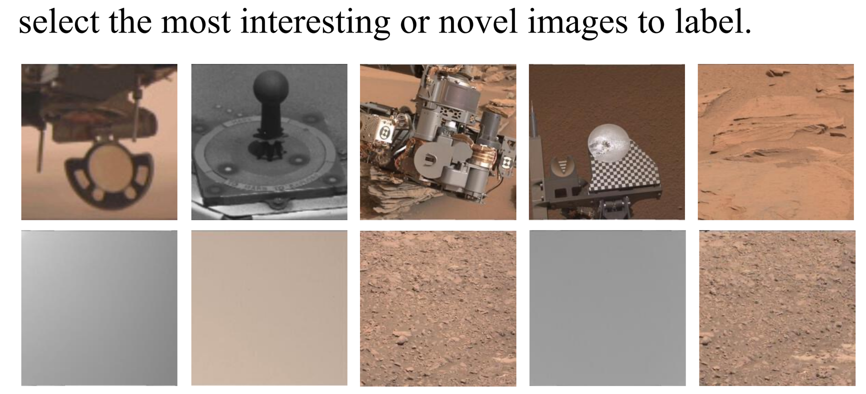


Precision and recall (MSLNet) | Confusion matrix (HiRISENet)

Data driven image labeling approach

(Wagstaff and Lu, 2020)

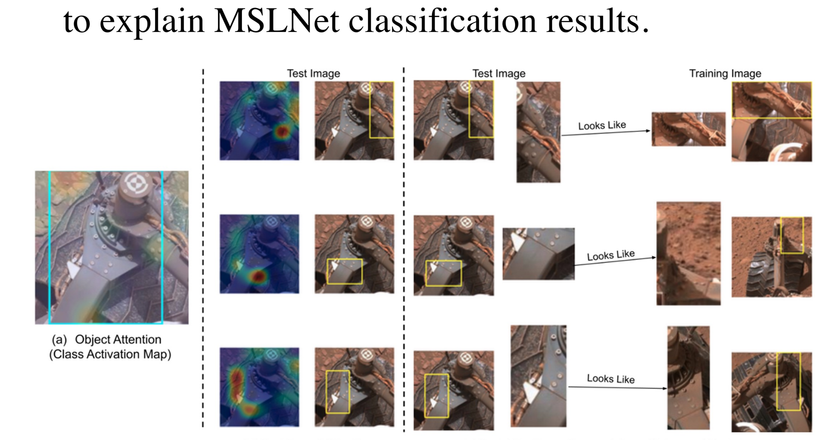
Employed DEMUD (Wagstaff et al., 2013) to incrementally select the most interesting or novel images to label.



Top row: 5 **most** novel MSL images selected by DEMUD
Bot row: 5 **least** novel MSL images selected by DEMUD

Classifier Explainability

Employed ProtoPNet architecture (Chen et al., 2019) to explain MSLNet classification results.



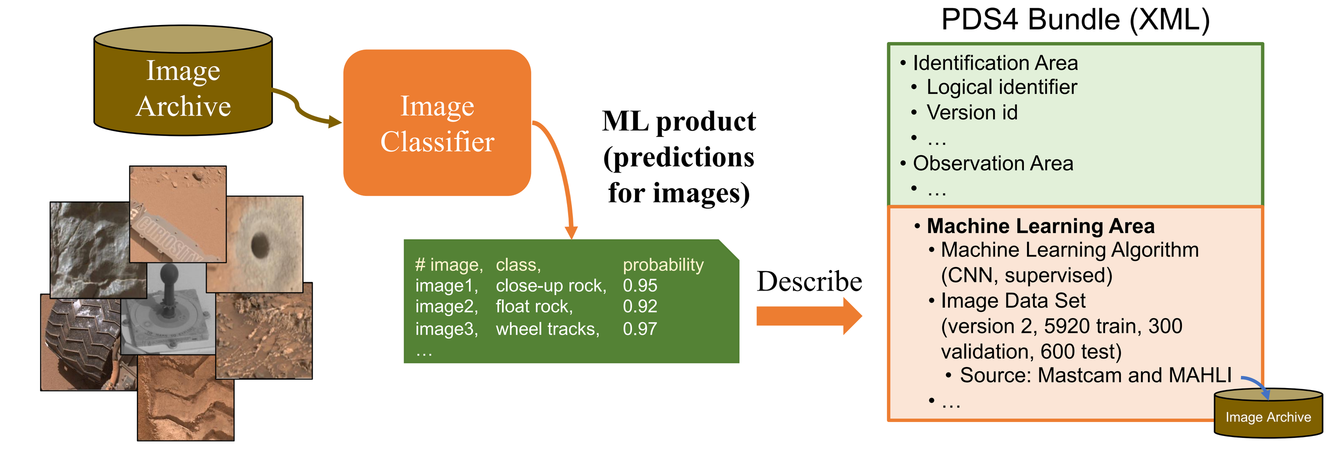
MSLNet Case based classifier result reasoning

4. Machine Learning Operationalization

Step 1: Machine Learning Products Delivery

Machine learning products are delivered to the PDS Imaging Node in PDS4 bundles. The PDS4 bundles are described using PDS4 Information Model[1] and Machine Learning Analysis Dictionary[2]. For more information, please see iPoster 1196514.

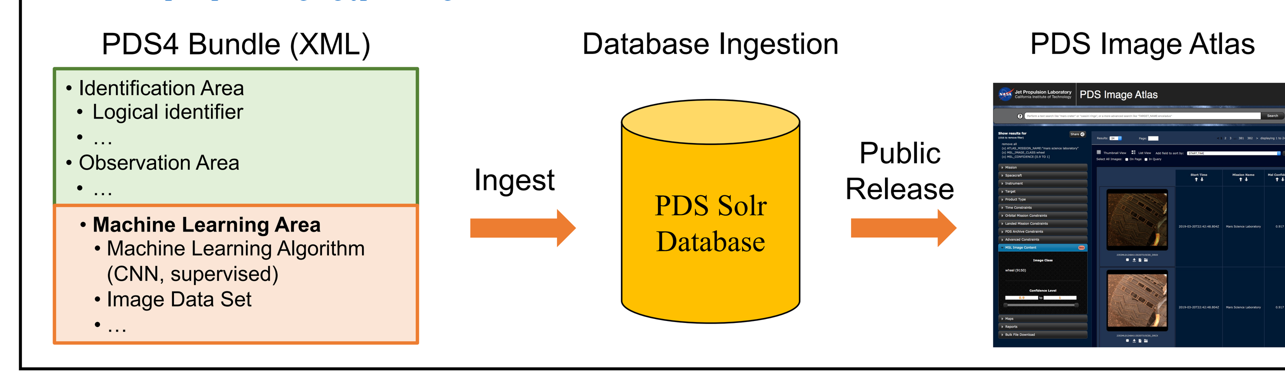
[1]https://pds.nasa.gov/datastandards/documents/im/v1/index_1I00.html
[2]https://pds.nasa.gov/datastandards/dictionaries/index-1.18.0.0.shtml#ml



Step 2: Machine Learning Products Deployment

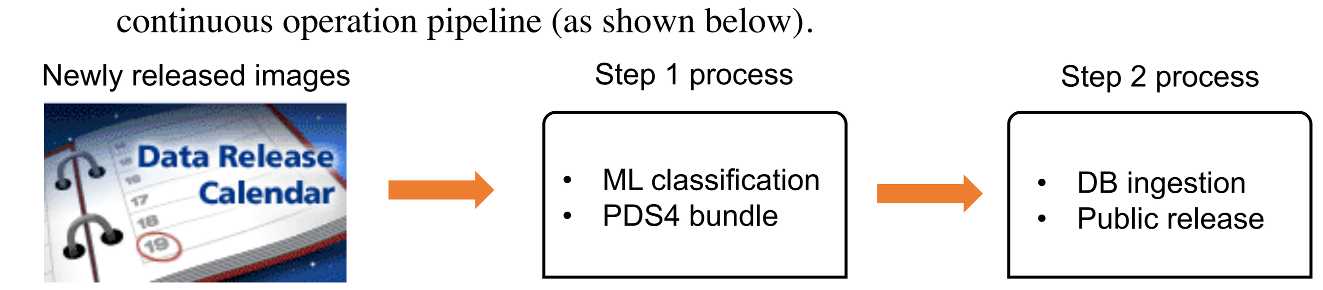
Machine learning PDS4 bundles are first ingested into PDS Solr database, and then made publicly accessible via PDS Image Atlas website[3].

[3]https://pds-imaging.jpl.nasa.gov/search



Step 3: Continuous Operation

Active planetary science missions periodically deliver image observations to the PDS Imaging Node. The newly delivered images will be processed using the continuous operation pipeline (as shown below).



Next step

Fully automate ML operation on AWS cloud

References

- C. Chen, O. Li, D. Tao, A. Barnett, C. Rudin, and J. Su. This Looks Like That: Deep Learning for Interpretable Image Recognition, NIPS, 2019.
- S. Lu, K. L. Wagstaff, K. Grimes, and M. Plotkin. Deep Ensemble MERNet: Recent Advances for the Content-Based Classification of MER Pancam Images. Planetary Science Informatics and Data Analytics Conference, 2022.
- A. Stanboli, B. Bue, K. L. Wagstaff, and A. Altinok. Automated Content Detection for Cassini Images. 3rd Planetary Data Workshop, 2017.
- K. L. Wagstaff, S. Lu, E. Dunkel, K. Grimes, B. Zhao, J. Cai, S. B. Cole, G. Doran, R. Francis, J. Lee, and L. Mandrake. Mars Image Content Classification: Three Years of NASA Deployment and Recent Advances. Proceedings of the Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence, 2021.
- K. L. Wagstaff, J. Panetta, A. Ansar, R. Greeley, M. P. Hoffer, M. Bunte, and N. Schorghofer. Dynamic Landmarking for Surface Feature Identification and Change Detection. ACM Transactions on Intelligent Systems and Technology, 2012.
- K. L. Wagstaff and S. Lu. Efficient Active Learning in New Domains. ICML Workshop on Real World Experiment Design and Active Learning, 2020.
- K. L. Wagstaff, N. L. Lanza, D. R. Thompson, T. G. Dietterich, and M. S. Gilmore. Guiding Scientific Discovery with Explanations using DEMUD. 27th AAAI, 2013.

Acknowledgements

We thank the Planetary Data System for funding and enthusiastically supporting this work. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.