

How can AI and machine learning help answer scientific questions?

Kiri L. Wagstaff

Special Advisor on AI
OSU Libraries



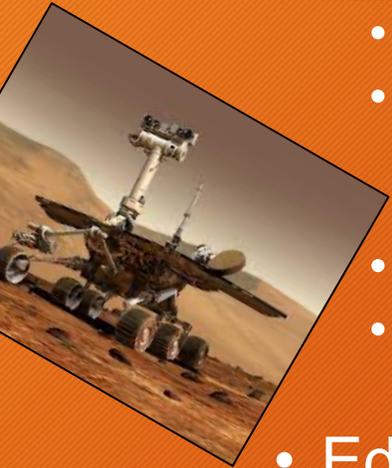
February 27, 2026

OSU College of Forestry



Who am I?

- Researcher – Ph.D. Computer Science + M.S. Geology
 - NASA Jet Propulsion Laboratory (2003-2022)
 - AI and machine learning for Earth and space exploration
 - Collaborations with planetary scientists, geologists, astronomers, cosmologists, agricultural scientists
 - Mars rover operations (Opportunity rover)
 - Fellow of the Association for the Advancement of Artificial Intelligence (AAAI)
- Educator – Oregon State University since 2019
 - “Machine Learning Challenges in the Real World”
 - “How to Be Less Wrong: A Study in Common Misconceptions”
 - With Andy Olstad (College of Business)
 - AI Advisor to U.S. Senator Mark Kelly (2023-2024)
 - Special Advisor on AI for the OSU Libraries



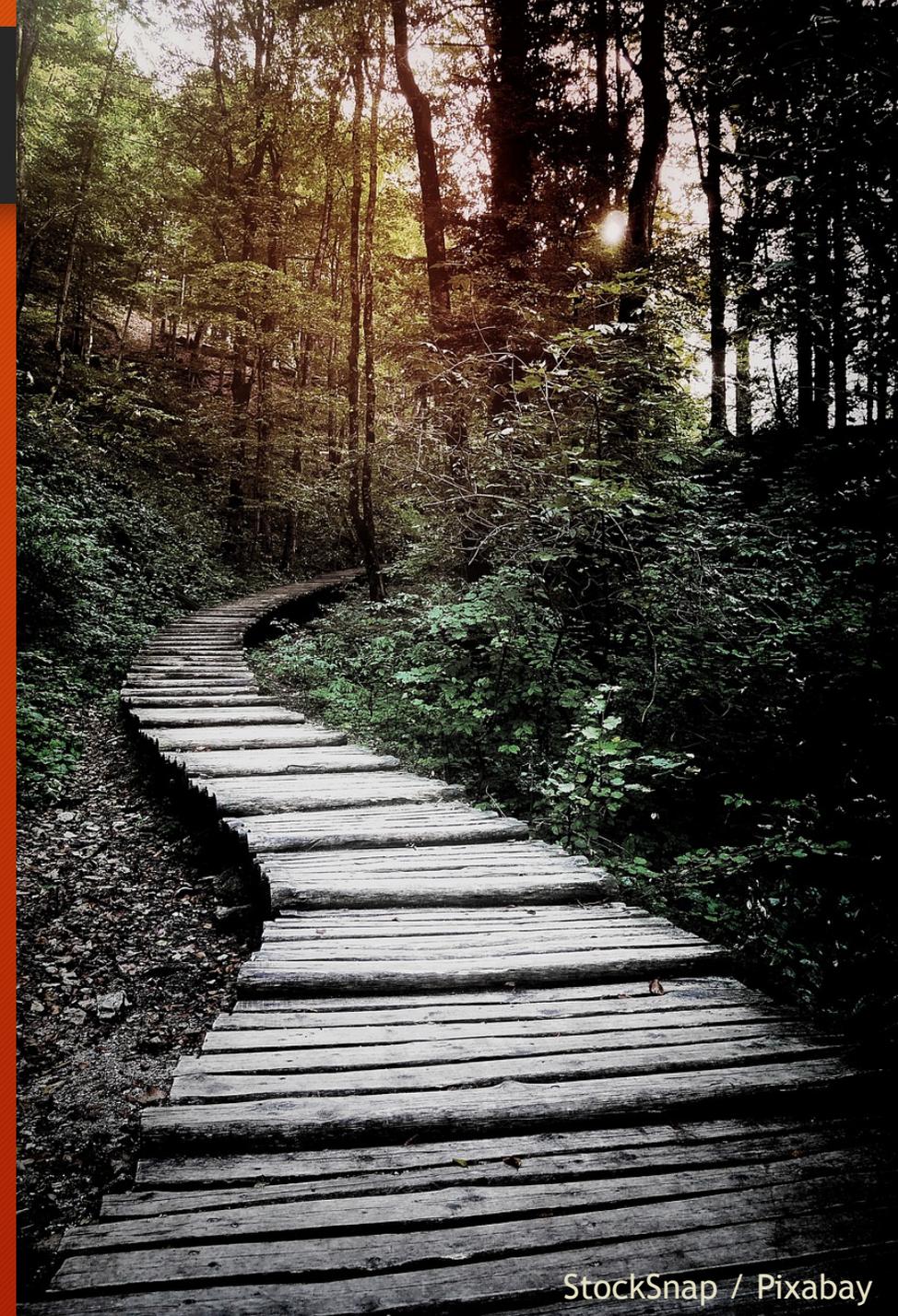
Mars Desert Research Station



U.S. Congress

How can AI/ML help with science?

1. Understand large data archives
 - Content-based search in Mars images
2. Search for features of interest, globally
 - Fresh impacts on Mars
3. Make new discoveries
 - Outliers in the Dark Energy Survey
4. What about generative AI?



Tools in the quest for scientific discoveries



Artificial Intelligence
(Machine Learning)

“Equipped with our five senses – along with telescopes and microscopes and mass spectrometers and seismographs and magnetometers and particle accelerators and detectors sensitive to the entire electromagnetic spectrum – we explore the universe around us and call the adventure science.” – Edwin Hubble

1. Use AI to understand what's in our data archives

- You can search NASA images by specifying *prior* info (prior = known before image is collected)
 - Mission
 - Instrument
 - Target
 - Location
 - Time / orbit number
- You want to search by *content* (only known after image is collected)
 - Crater, dune, impact, streak, ...
 - Layers, sand, drill hole, sky, ...



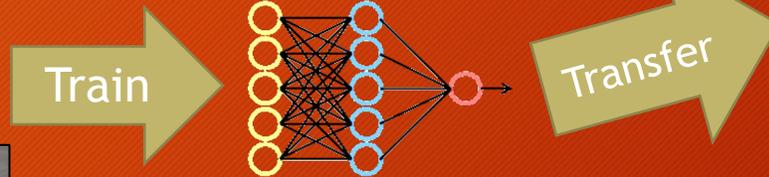
A screenshot of the NASA PDS Image Atlas search interface. The page header includes the NASA logo, Jet Propulsion Laboratory, and California Institute of Technology. The main title is "PDS Image Atlas". A search bar at the top contains the text "Perform a text search like 'mars crater' or 'cassini rings', or a more advanced search like 'TARGET_NAME:enceladus'". Below the search bar, there are options to "Show results for" and "Narrow your search by selecting a facet below". A list of facets is shown on the left, with "Mission" through "PDS Archive Constraints" highlighted in green and "Advanced Constraints" through "MER Image Content" highlighted in blue. The main content area displays a grid of image thumbnails, each with a filename and download options. The interface also includes a "Share" button, "Results: 24", "Page: 1", and a "Sort by: START_TIME" dropdown menu.

<http://pds-imaging.jpl.nasa.gov/search/>

Machine learning for image content annotation

- Transfer learning: “Fine-tune” a neural network trained on Earth images to work on Mars images
- Deployed on >1M Mars **rover** images

ImageNet data set
[Russakovsky et al., 2015]



PDS image search for “layers”

<http://pds-imaging.jpl.nasa.gov/search/>

(x) FILE_NAME:***E01*DRCL*
(x) MSL_CONFIDENCE:[0.99 TO 1]

▼ Mission
mars science laboratory (11)

▶ Spacecraft
▶ Instrument
▶ Target
▶ Product Type
▶ Lighting Geometry
▶ Filters
▶ Lat/Lon Bounding Box
▶ Time Constraints
▶ Orbital Mission Constraints
▶ Landed Mission Constraints
▶ PDS Archive Constraints
▶ Advanced Constraints
▶ MRO HiRISE Image Landmarks Beta
▶ MSL Image Content Beta
▶ Cassini ISS Image Content Beta
▶ Galileo SSI Europa Image Content Beta
▶ MER Image Content Beta
▶ Maps
▶ Reports
▶ Bulk File Download

Feedback

0976MR0043100100502754E01_DRCL
0580MR0024070780400073E01_DRCL
1033ML0045210280305633E01_DRCL
1686MR0087700000802854E01_DRCL
1148MR0051940160601762E01_DRCL
0631ML0026080100302431E01_DRCL
0631MR0026080150401645E01_DRCL
1296MR0061550070700213E01_DRCL
0631MR0026080100401640E01_DRCL



Kiri Wagstaff



Steven Lu



Kevin Grimes



Gary Doran



Thamme Gowda



Emily Dunkel

PDS image search for “wheel”

<http://pds-imaging.jpl.nasa.gov/search/>

Navigation menu:

- ▶ Landed Mission Constraints
- ▶ PDS Archive Constraints
- ▶ Advanced Constraints
- ▶ MRO HiRISE Image Landmarks Beta
- ▼ MSL Image Content Beta
- ▶ Cassini ISS Image Content Beta
- ▶ Galileo SSI Europa Image Content Beta
- ▶ MER Image Content Beta
- ▶ Maps
- ▶ Reports
- ▶ Bulk File Download

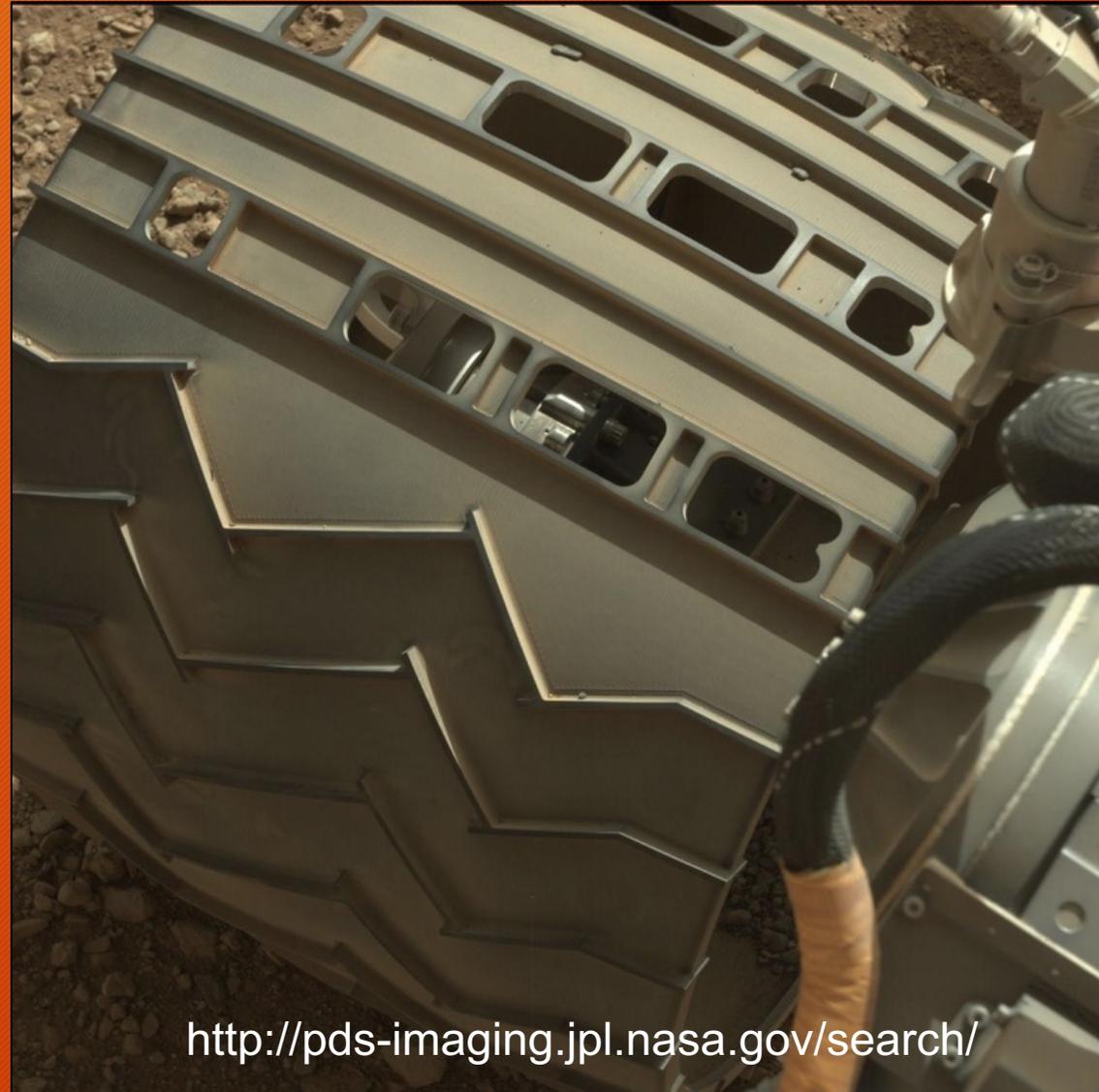
Image Class
wheel (2320)

Confidence Level
0.9 to 1

Grid of 9 image thumbnails:

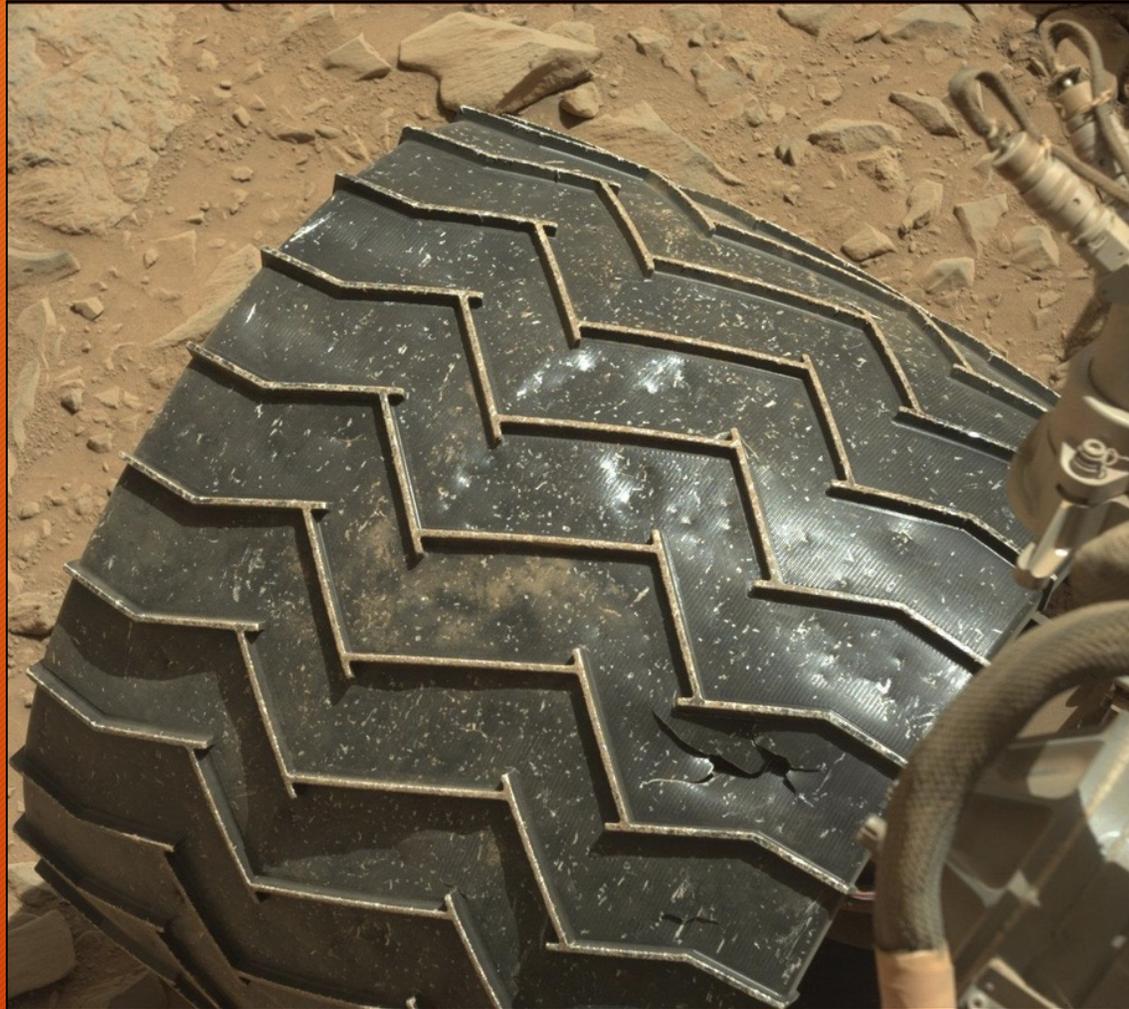
- 0549ML0022140030204115E01_DRCX
- 0549ML0022140030204115E01_DRLX
- 0603ML0025430020301345E01_DRCX
- 0940ML0041300020403055E01_DRXX
- 0940ML0041300020403055E01_DRCX
- 0940ML0041300020403055E01_DRCL
- 0603ML0025430020301345E01_DRXX
- 0603ML0025430020301345E01_DRLX
- 0603ML0025430020301345E01_DRCL

August 2012



<http://pds-imaging.jpl.nasa.gov/search/>

December 2013



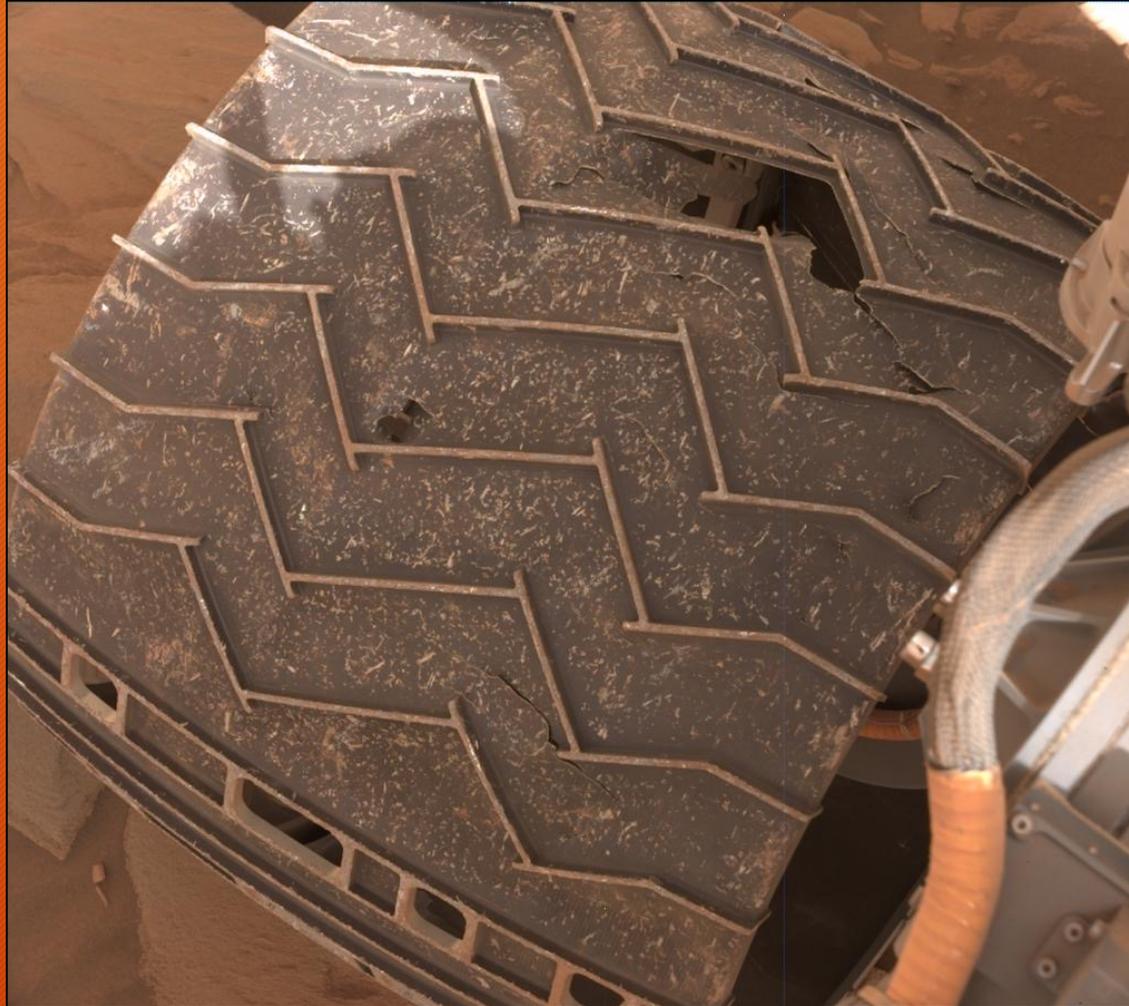
<http://pds-imaging.jpl.nasa.gov/search/>

November 2016



<http://pds-imaging.jpl.nasa.gov/search/>

March 2020



<http://pds-imaging.jpl.nasa.gov/search/>

It works for orbital images, too

- E.g., search for images containing craters
- Shows landmarks with ≥ 0.9 probability
- Landmarks are localized

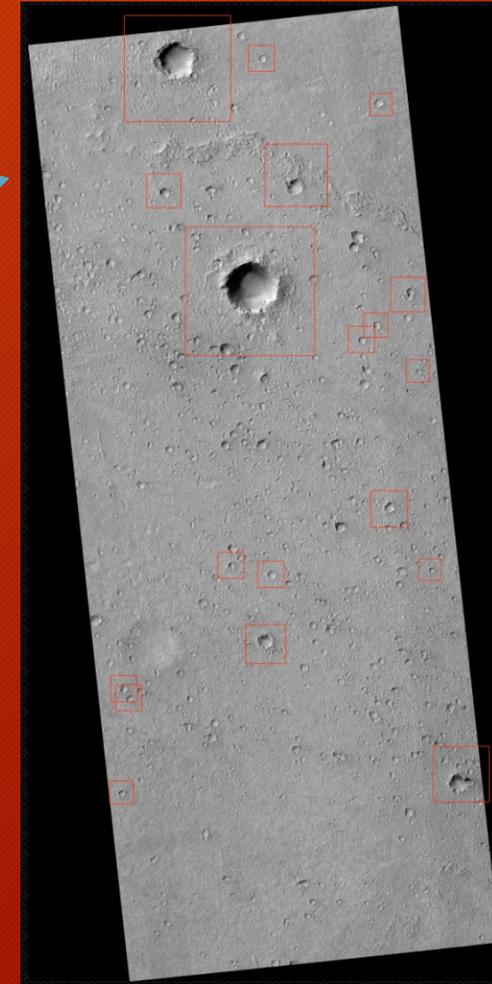
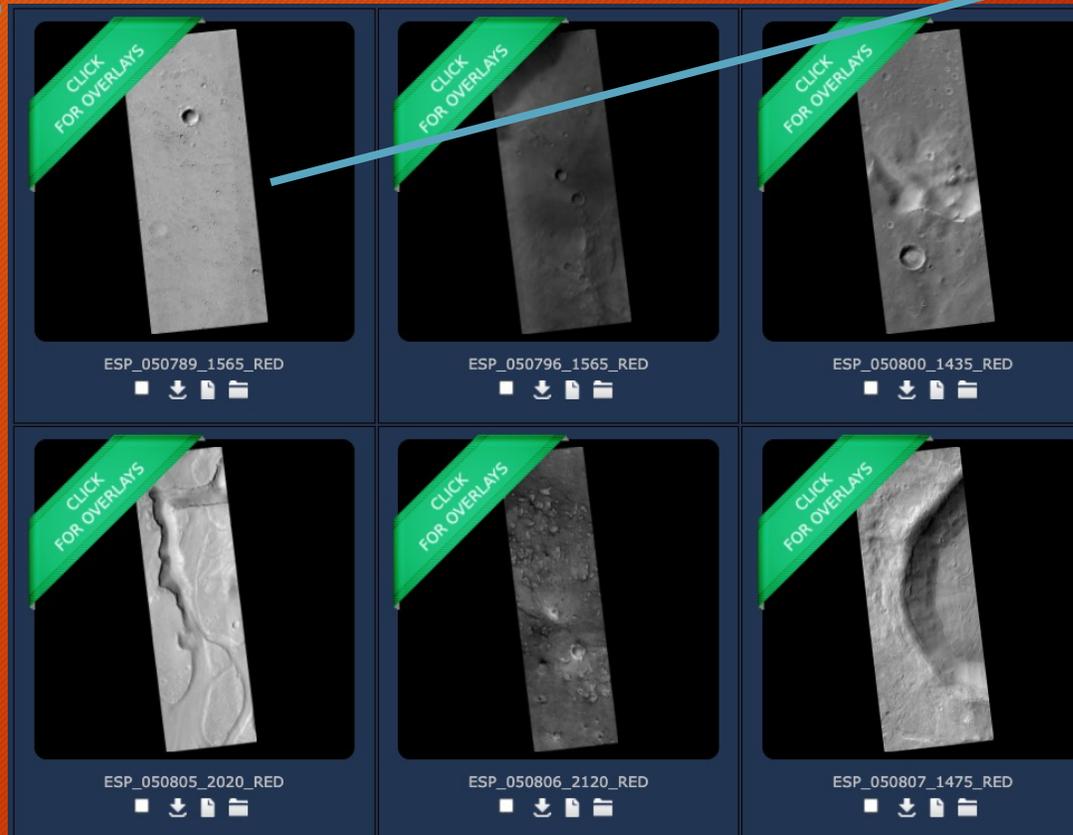
First 6 results from "crater" search

PDS Image Atlas

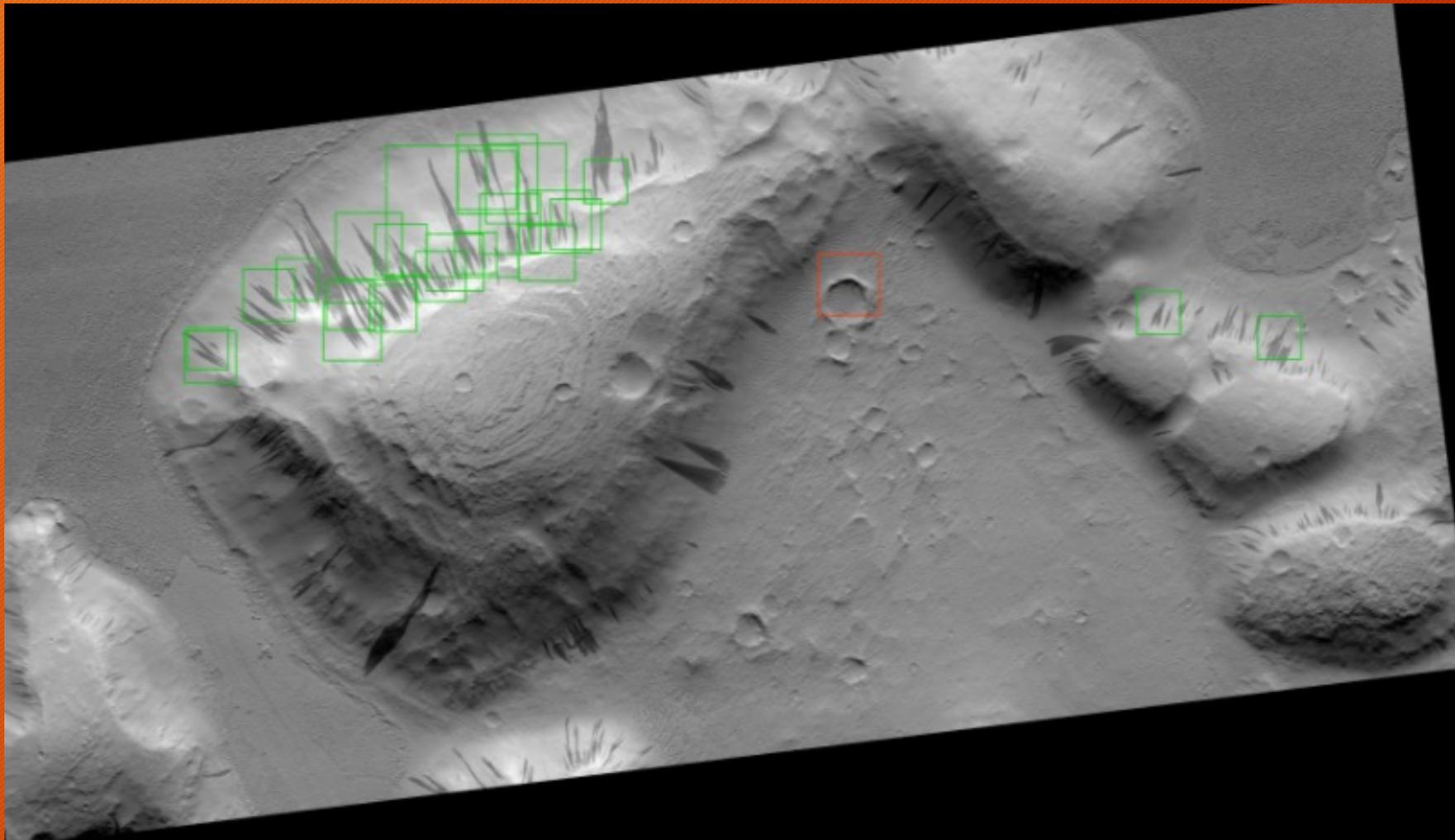
▼ MRO HiRISE Image Landmarks Beta

Image Class
(MRO HiRISE RDRs Only)

- bright dune (527)
- crater (25071)
- dark dune (2502)
- impact ejecta (2480)
- slope streak (2153)
- spider (228)
- swiss cheese (1058)

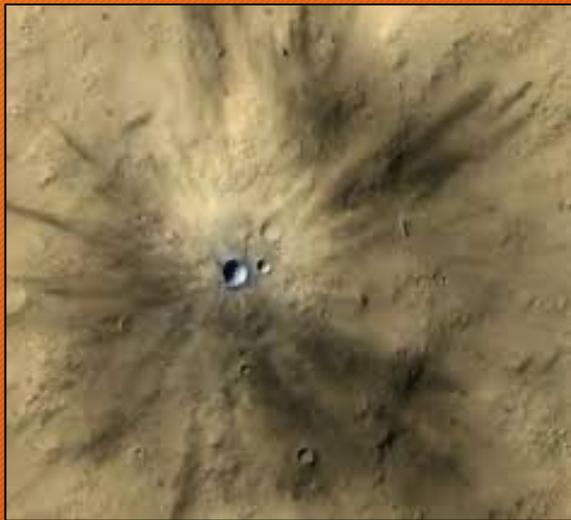


Dark slope streaks



2. Use AI to search an entire planet

- Mars orbital images show evidence of recent meteorite impacts on Mars



Credit: NASA/JPL-Caltech/MSSS/UA



Credit: NASA/JPL-Caltech/UA

- ~1000 have been manually catalogued
- Previous orbital image classifier uses HiRISE data, which covers only ~3% of the planet
- To find more, we need an image classifier for CTX (global) data



Kiri Wagstaff



Ingrid Daubar



Michael Munje



Gary Doran



Valentin Bickel



Joe Pate



Annabelle Gao

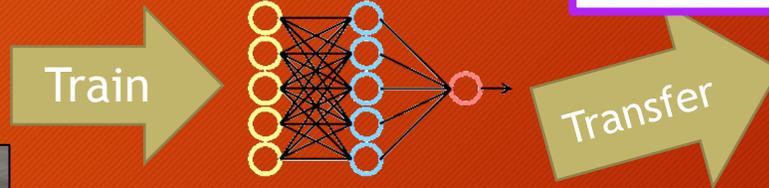


Daniel Wexler

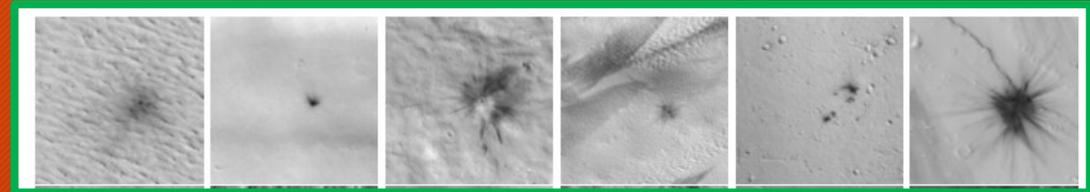
Fresh impact image classifier

- Transfer learning: “Fine-tune” a neural network trained on Earth images to work on Mars images
- Deployed on 12 TB of Mars **orbiter** images

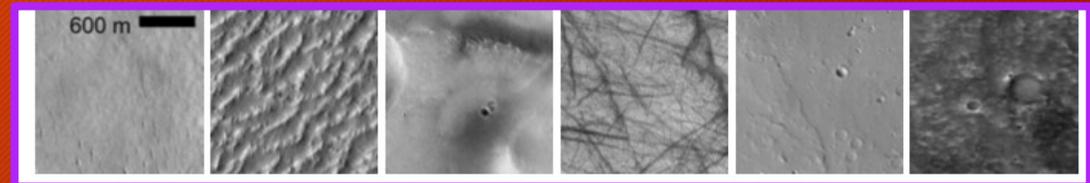
ImageNet data set
[Russakovsky et al., 2015]



Fresh impacts (n=1856)

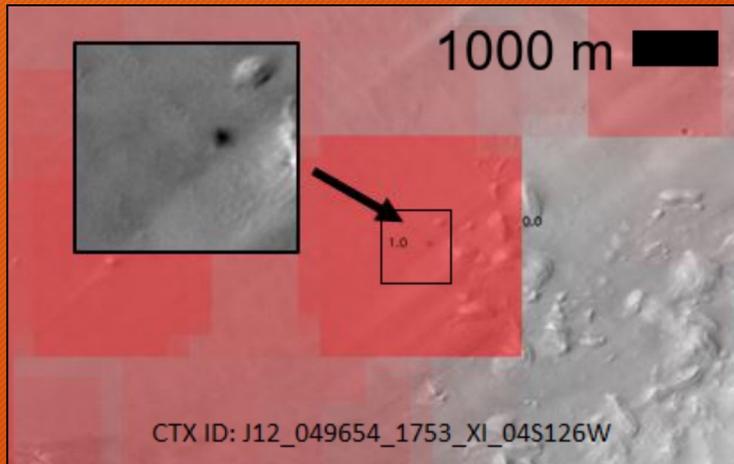


Not impacts (n=4793)



Hunting for new Mars impacts

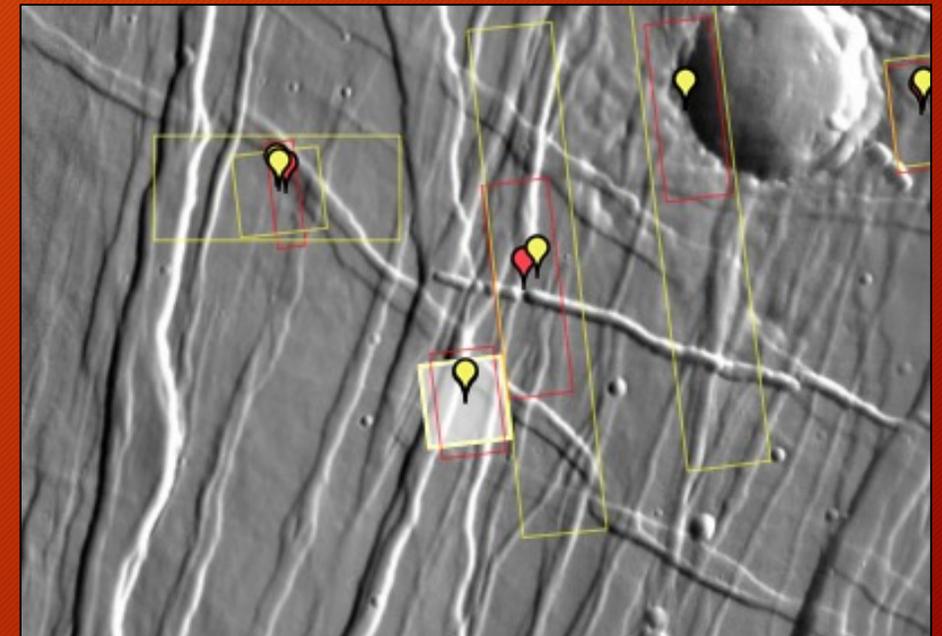
- Slide classifier window over global Mars images to find matches
 - 2 billion classifications (CTX: 5000 x 25,000+ pixels)
 - A cluster of ~500 CPUs took a week (~5 sec/image)
 - Manual hunt (40 min/image) would require ~8.5 person-years (no breaks!)



[Munje et al., 2020]



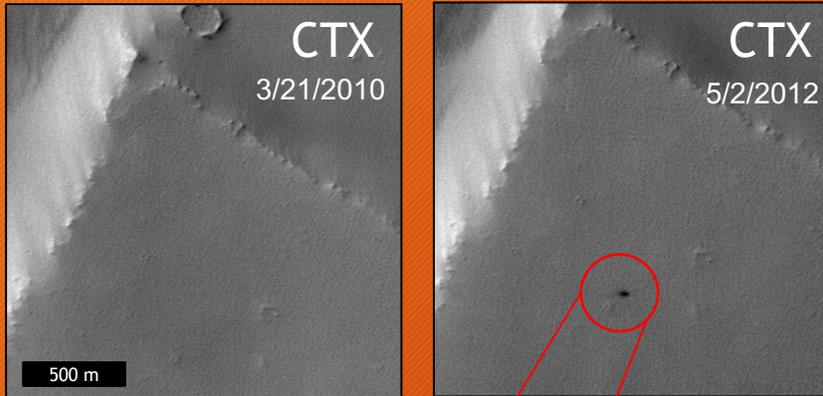
<https://www.uahirise.org/hiwish/>



- Possible new impact? Request high-resolution image with HiWish!

New Mars impact discoveries

Noctis Fossae

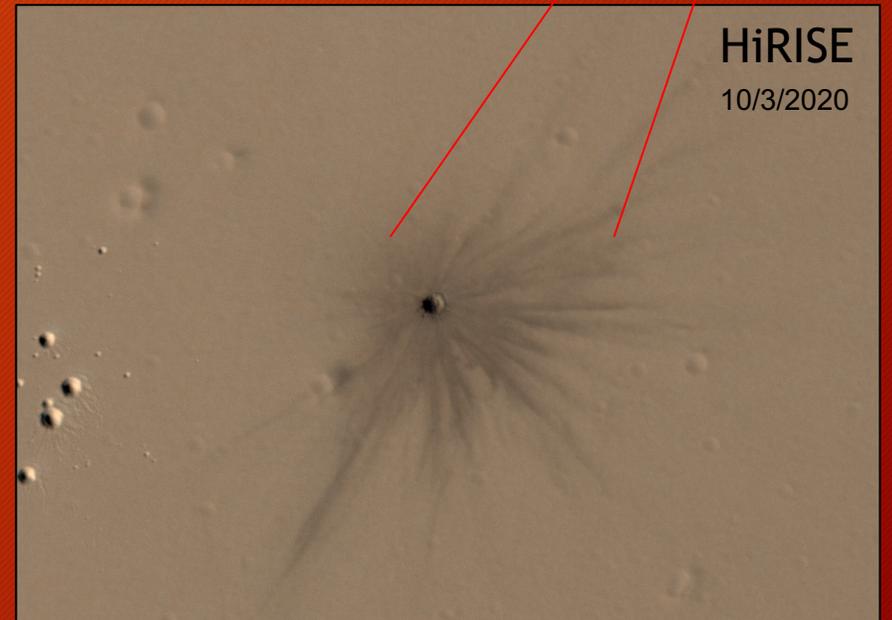
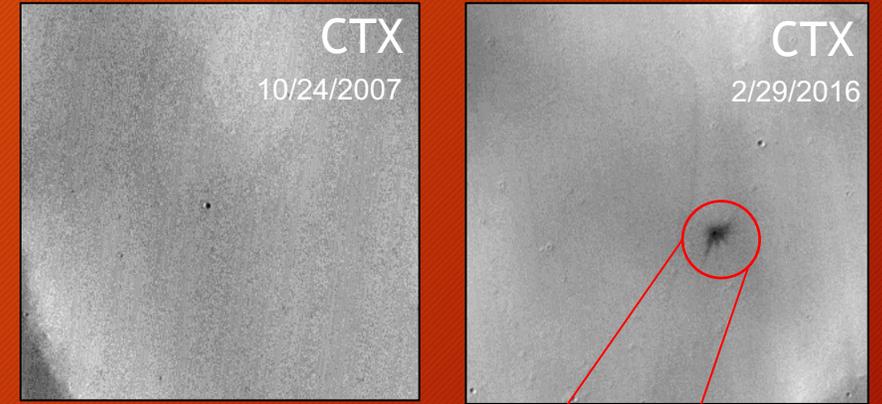


CTX: 6 m/pixel



HiRISE: 0.3 m/pixel

Noctis Labyrinthus

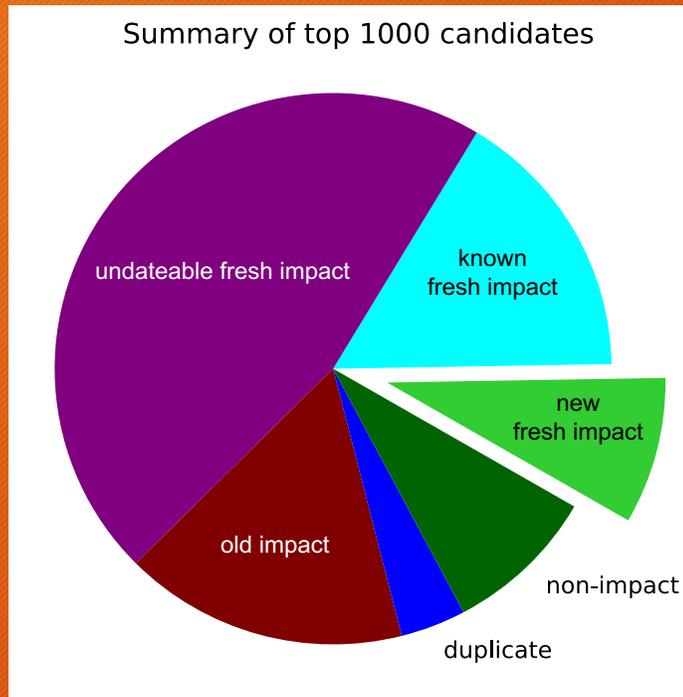


2/27/2026

20

70+ new Mars impact discoveries

- >1,000,000 candidates with $P(\text{impact}) \geq 0.99$
- Group by location and time, then manually review the top 1000 (by probability)
 - New impact? Request high-resolution image with HiWish

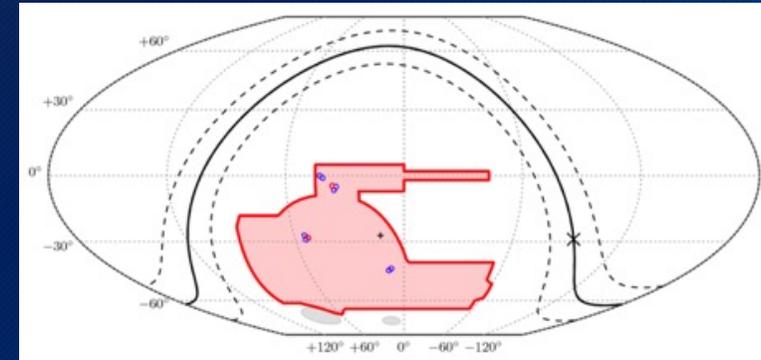


- 99 already known (confirmation)
- 69 new discoveries
 - (+3 also independently found by humans during our study)
- 465 are impacts without a “before” image (undateable)
- 166 old impacts
- Only 90 (9%) non-impacts

[Wagstaff et al., *Icarus* 2022; Daubar et al, *JGR* 2022]

3. Use AI to make new discoveries

- What lies within the Dark Energy Survey?
- DES surveyed 5000 square degrees of southern sky over 6 years (Aug. 2013 - Jan. 2019)
 - 4-m telescope in Chile
- 400M objects (~310M galaxies)
 - Goal: estimate structure, geometry, and expansion of the universe
 - Many quality filters, but **artifacts** remain
- Our goal: Identify anomalies in DES catalog to
 - Find, filter, and understand **artifacts**, which can
 - Inform improvements to the DES processing pipeline and
 - Yield better estimates of cosmological model parameters
 - Discover **new scientific phenomena** (what's out there?)



[Abbott et al., 2018]



Kiri Wagstaff



Umaa Rebbapragada

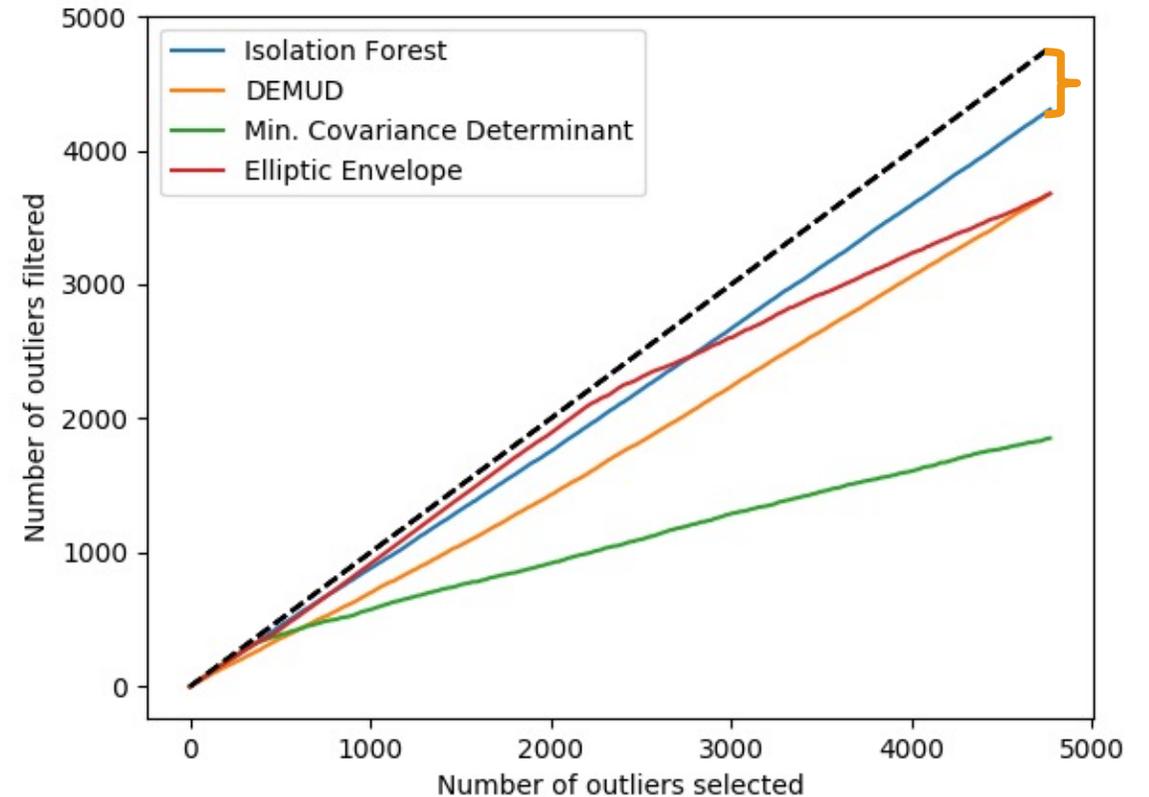


Eric Huff

Anomaly score validation

- Identify anomalies from 12M objects
- Independent DES catalog cleaning effort removed millions of objects
 - 31 months x 10 people
- Isolation Forest had highest agreement with human filtering
 - IF identified 4,768 total anomalies

Number of true anomalies found in top 4,768 of 12M

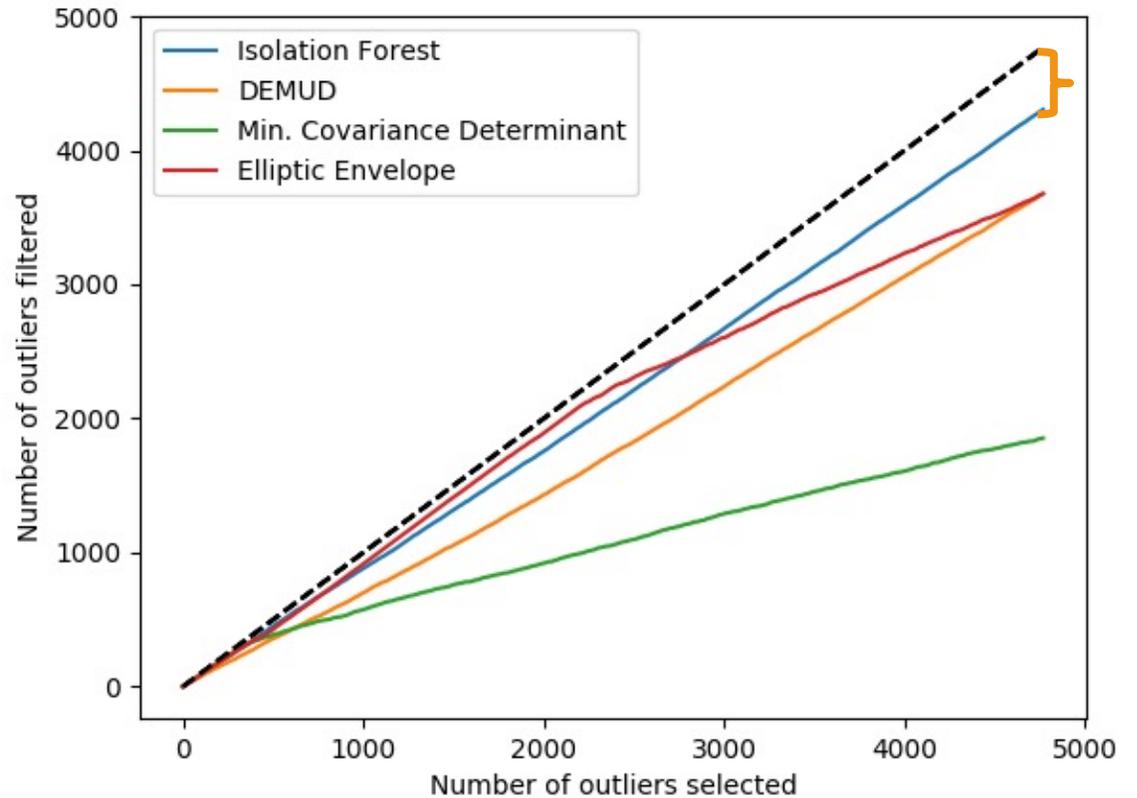


Anomaly score validation

- Identify anomalies from 12M objects
- Independent DES catalog cleaning effort removed millions of objects
 - 31 months x 10 people
- Isolation Forest had highest agreement with human filtering
 - IF identified 4,768 total anomalies
- **538 objects that IsoForest flagged but humans did not**
 - 38% due to modeling errors
 - 33% due to data corruption
 - 9% normal objects
 - 20% of potential scientific interest

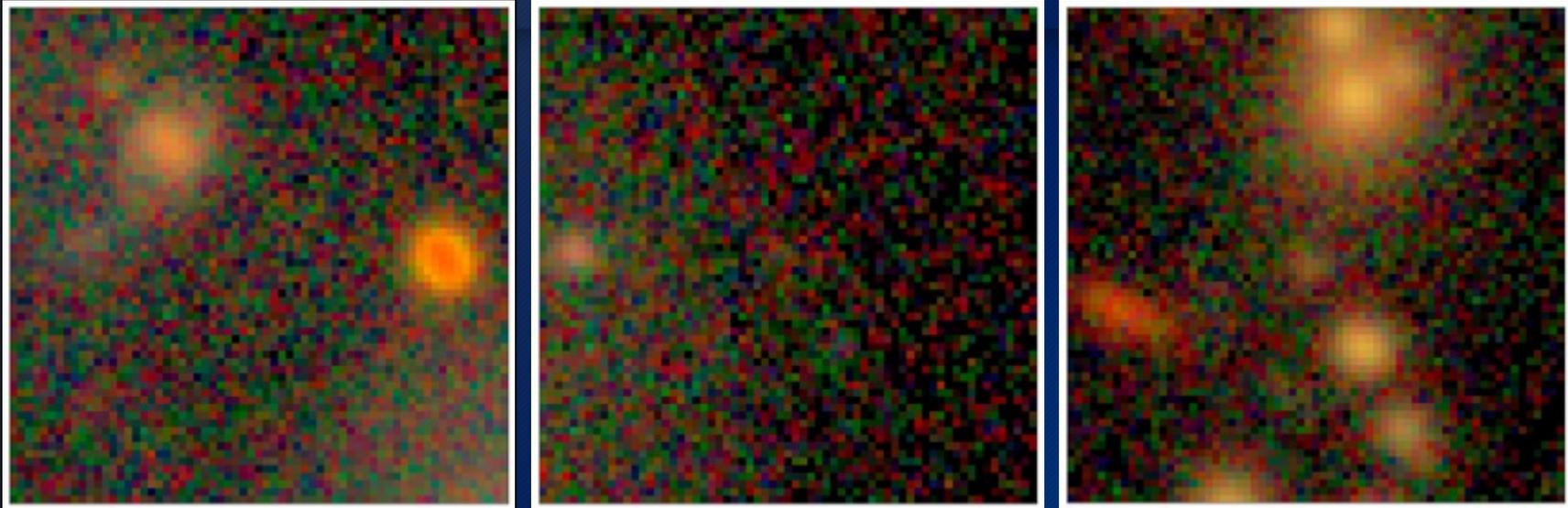


Number of true anomalies found in top 4,768 of 12M

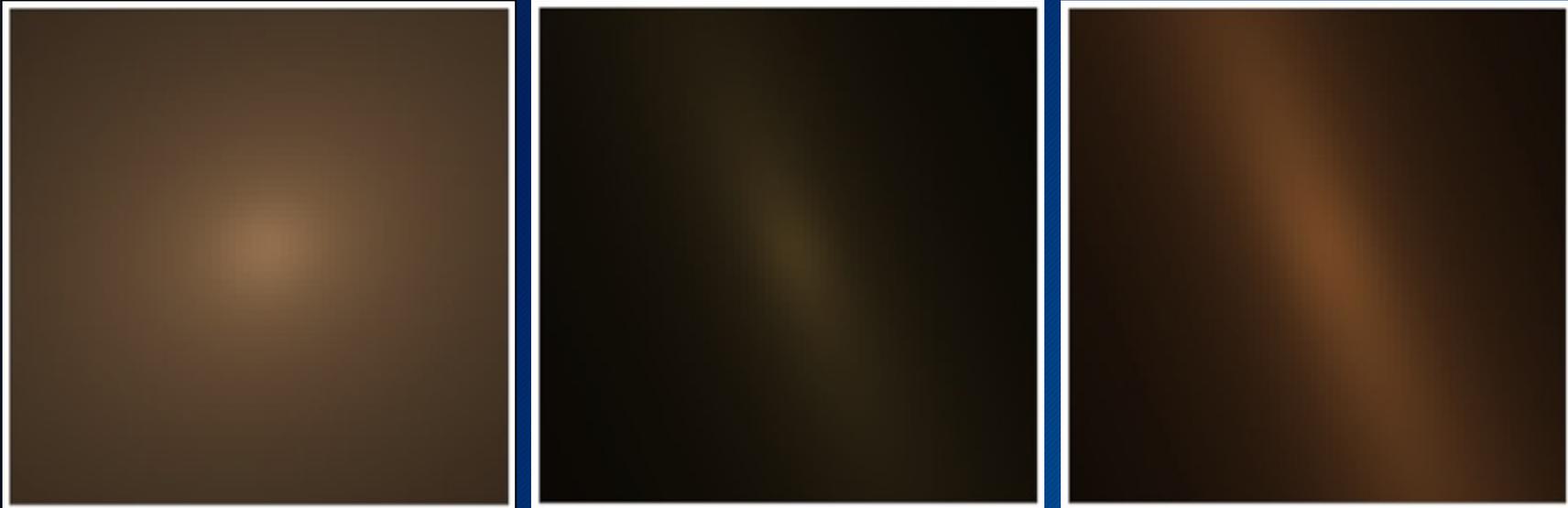


DES Anomalies: Modeling Errors

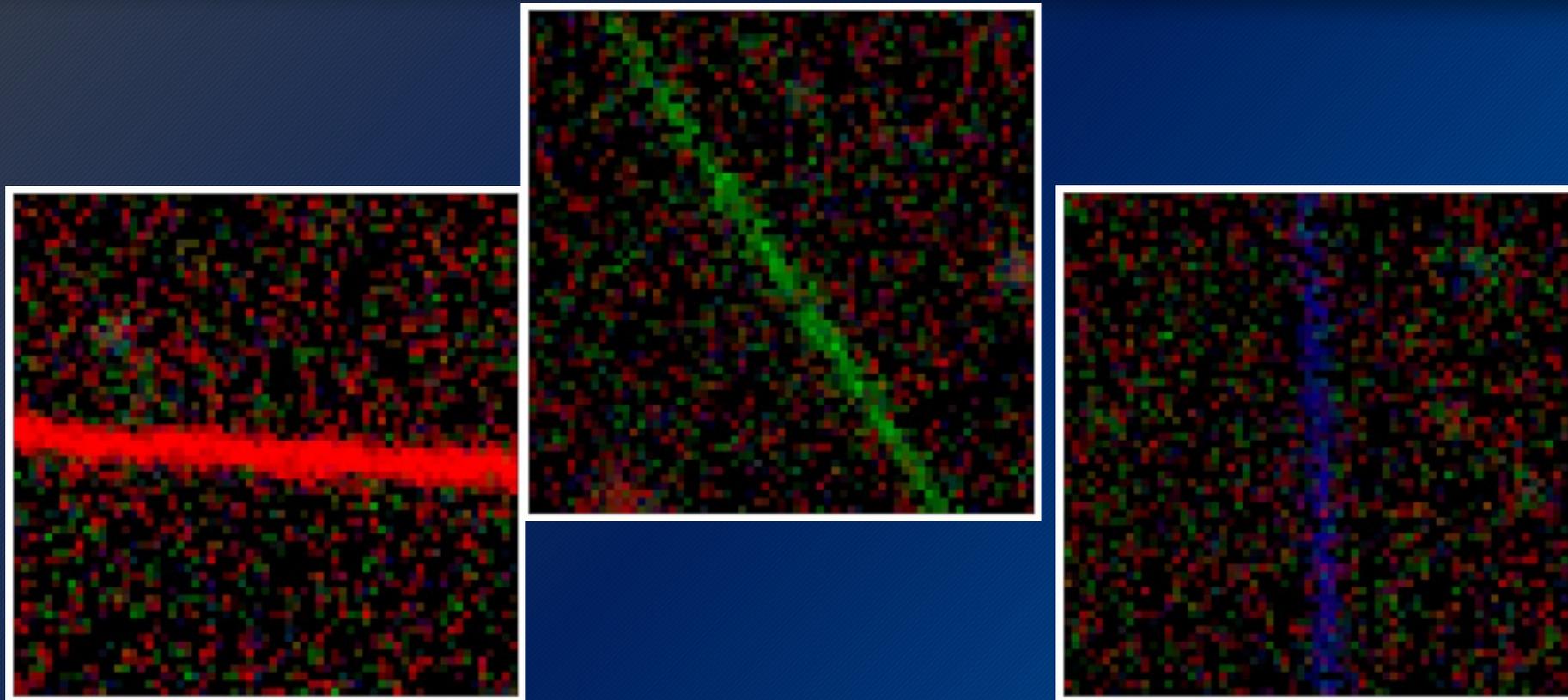
Data



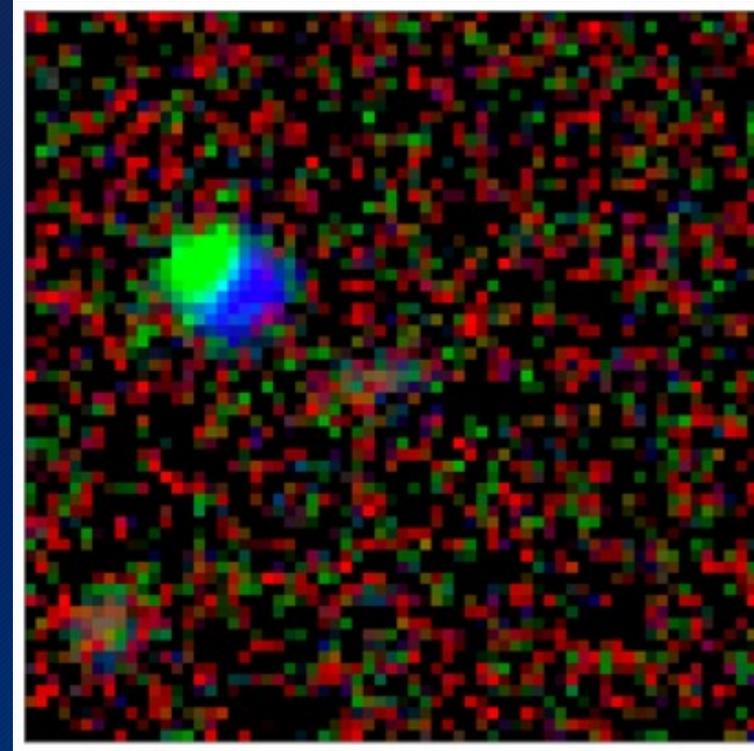
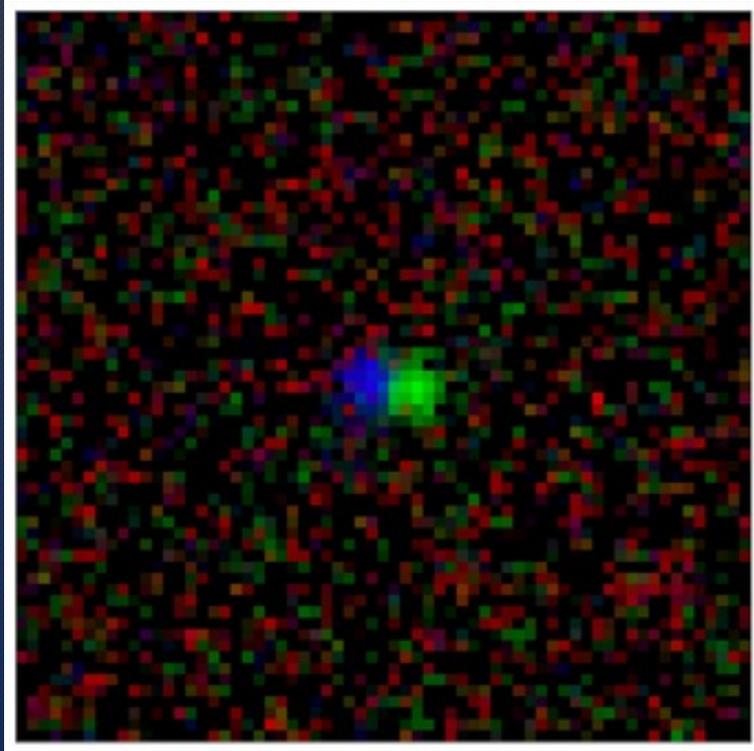
Model



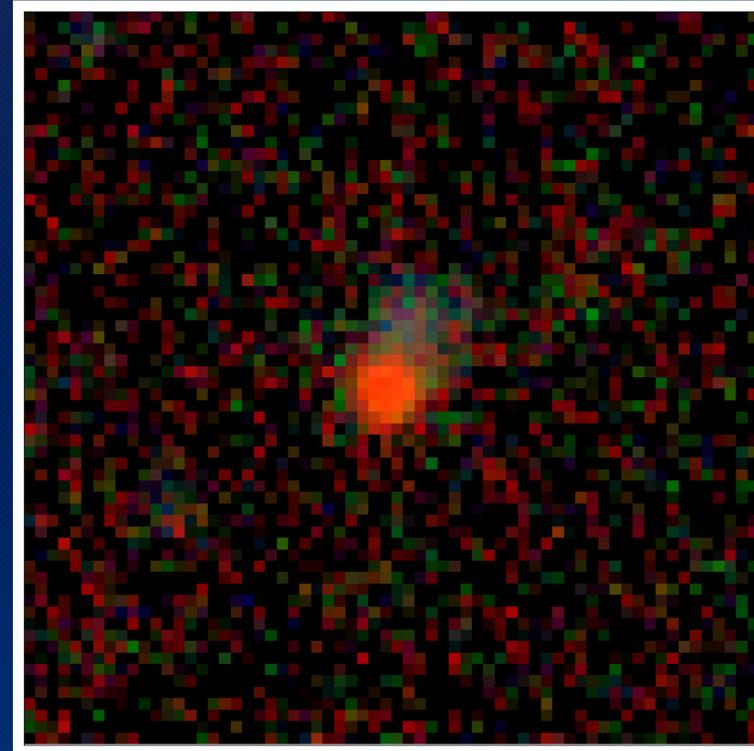
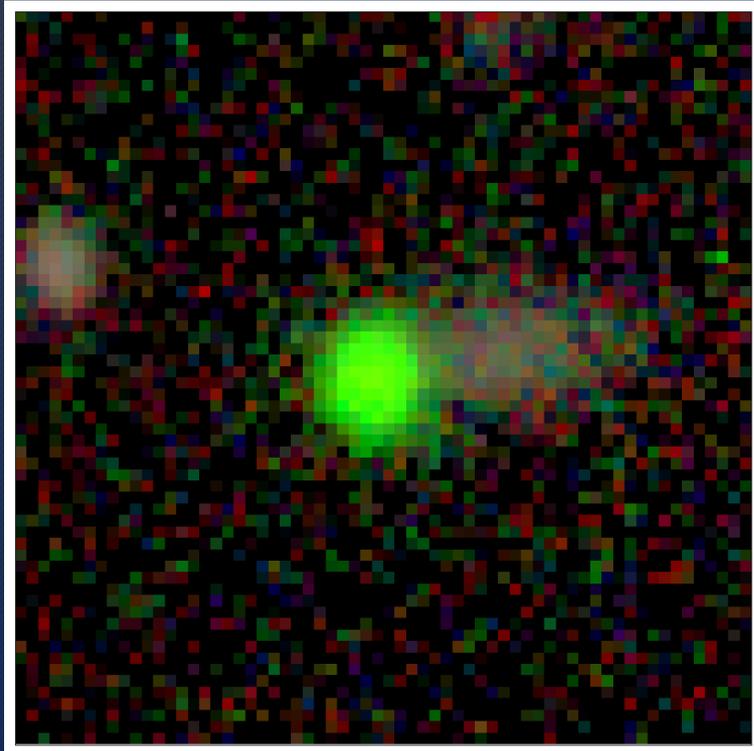
DES Anomalies: Satellite Trails



DES Anomalies: Moving Objects (Asteroids)

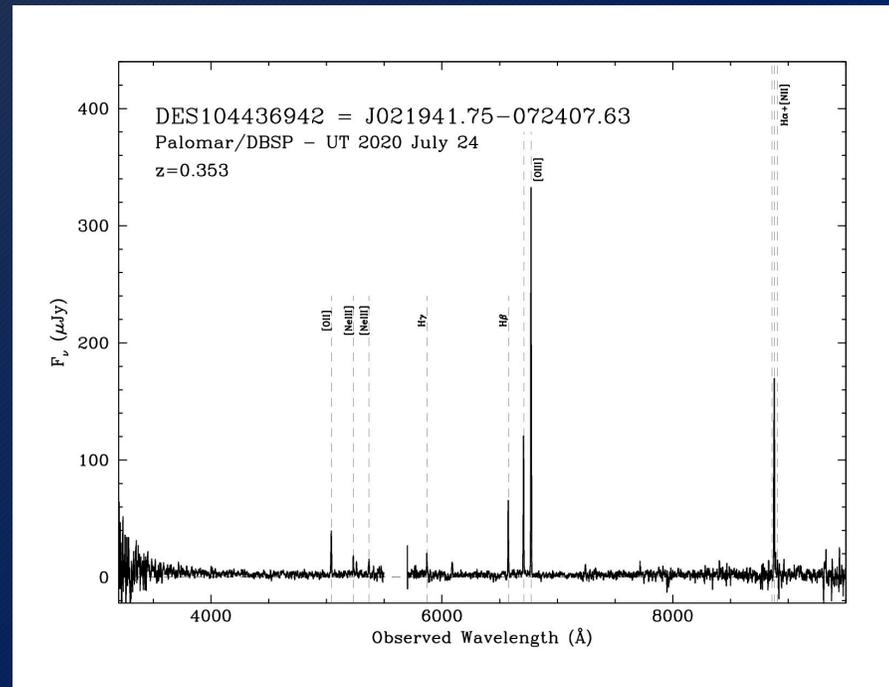
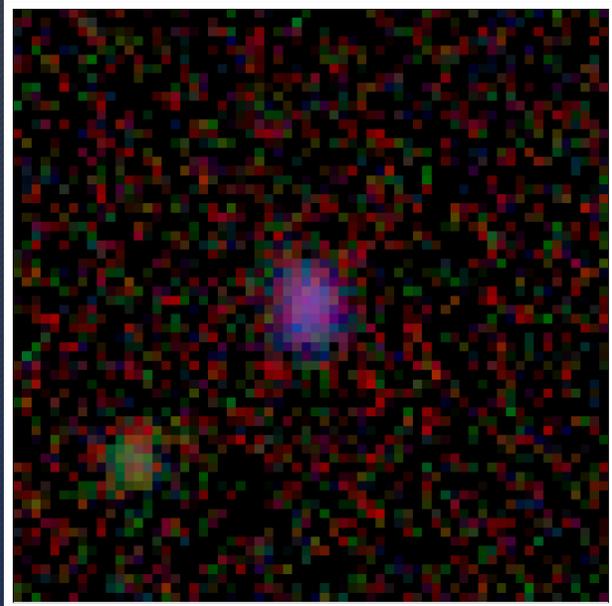


DES Anomalies: Supernovae?



Mystery galaxy – new discovery?

Follow-up observation from Palomar

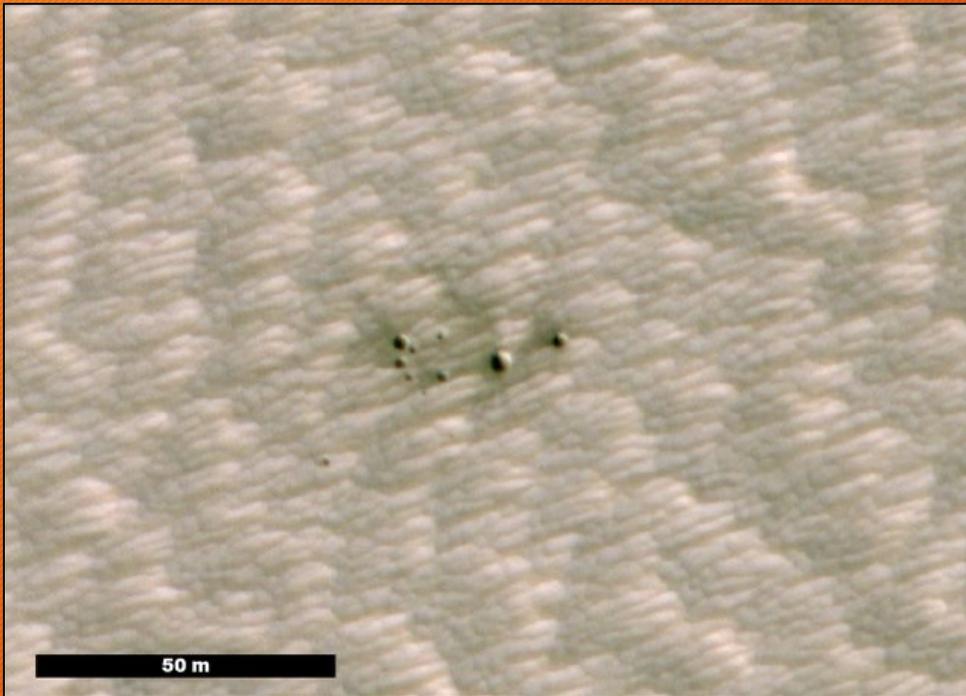


[Wagstaff et al., *ADASS* 2020]

Galaxy with
high star formation rate
yet low total mass

Where's the generative AI?

1. Can it annotate image content? (E.g. Image caption generation)



Please generate a caption for this image of the Martian surface.

ChatGPT:

Caption:

High-resolution image of the Martian surface captured by NASA's HiRISE camera aboard the Mars Reconnaissance Orbiter, showing the landing site of the Curiosity rover in Gale Crater. The dark spots and streaks mark the descent stage impact and the rover's touchdown area on the patterned surface of Mars. Scale bar: 50 meters.

We want to know what's out there, not what *could* be out there (great for sci-fi stories, but not for science)

- Generative AI creates a common / likely caption, not an accurate one

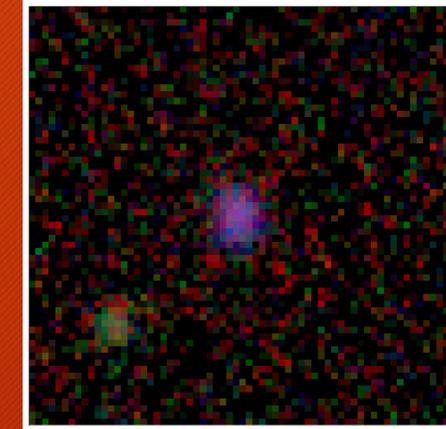
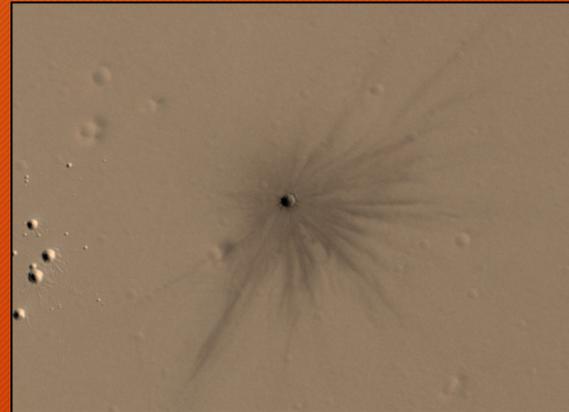
So: How can AI and ML help answer scientific questions?

- Classification and discovery methods help accelerate science



Classify and annotate large data archives to enable content-based search

Quickly scan global data sets for features of interest (fresh impacts)



Identify outliers or anomalies that could lead to new discoveries

- ML leverages, but does not replace, human expertise
 - Human review is a key part of this process (and it takes time and effort!)
 - ML/AI can focus attention where it is most likely to pay off



Artificial Intelligence

Contact: wkiri@wkiri.com