Detecting and Explaining Unexpected Image Content

Jake Lee, Jet Propulsion Laboratory, California Institute of Technology
& Columbia University, CS ‘19
Intern under Kiri Wagstaff and Umaa Rebbapragada
Summer 2017

Government sponsorship acknowledged. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA
CL#17-4609

Roadmap

- Introduction & Motivation
- Approach & Implementation
- Experiments
- Limitations & Future work
- Acknowledgements
- Questions

Introduction and Motivation

NASA collects a high volume of images

- Increasing volume and diversity of:
    - Sources
    - Targets
    - Filters
    - etc.

Jet Propulsion Laboratory
California Institute of Technology

PDS Imaging Atlas

Mission

2001 mars odyssey (4166414)
cassini (973529)
chandrayaan-1 (21645)
clementine (1996197)
galileo (20123)
lcross (113044)
lunar reconnaissance orbiter (4116047)
magellan (72818)
mars exploration rover (6799735)
mars global surveyor (243227)
mars pathfinder (17899)
mars reconnaissance orbiter (1591176)
mars science laboratory (9186574)
messenger (979681)
new horizons (15179)
phoenix (256433)
viking lander (6585)
viking orbiter (61693)
voyager (393059)

mars science laboratory (9186574)

- Such high volume is difficult to process
- Certain cases allow for automation
- Any solutions for exploration or discovery?

https://pds-imaging.jpl.nasa.gov/search/?q=*%3A*

A brief introduction to DEMUD

- Developed in 2013 as a prior-free rare category detection algorithm
- Prioritizes interesting data
- Provides explanations for its prioritizations

Wagstaff, Kiri L., et al. "Guiding Scientific Discovery with Explanations Using DEMUD." *AAAI*. 2013.

DEMUD usage

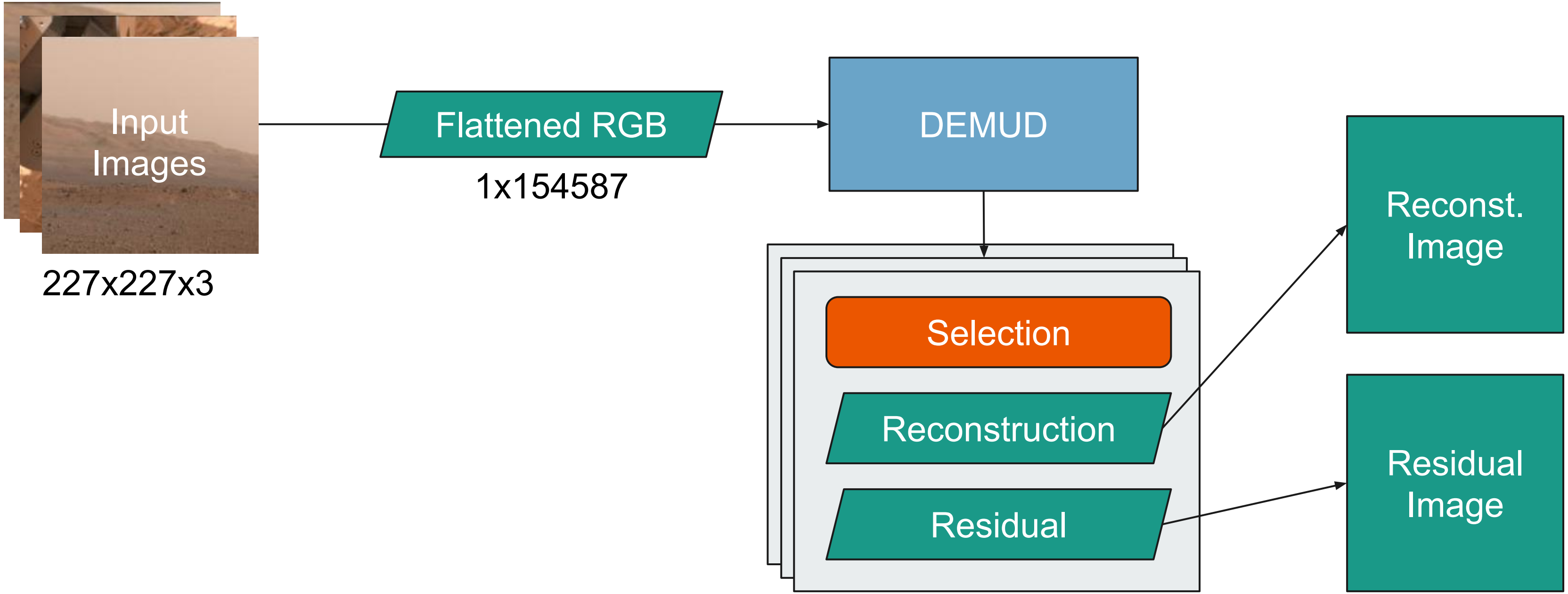
Applied to ChemCam spectral data (2014)

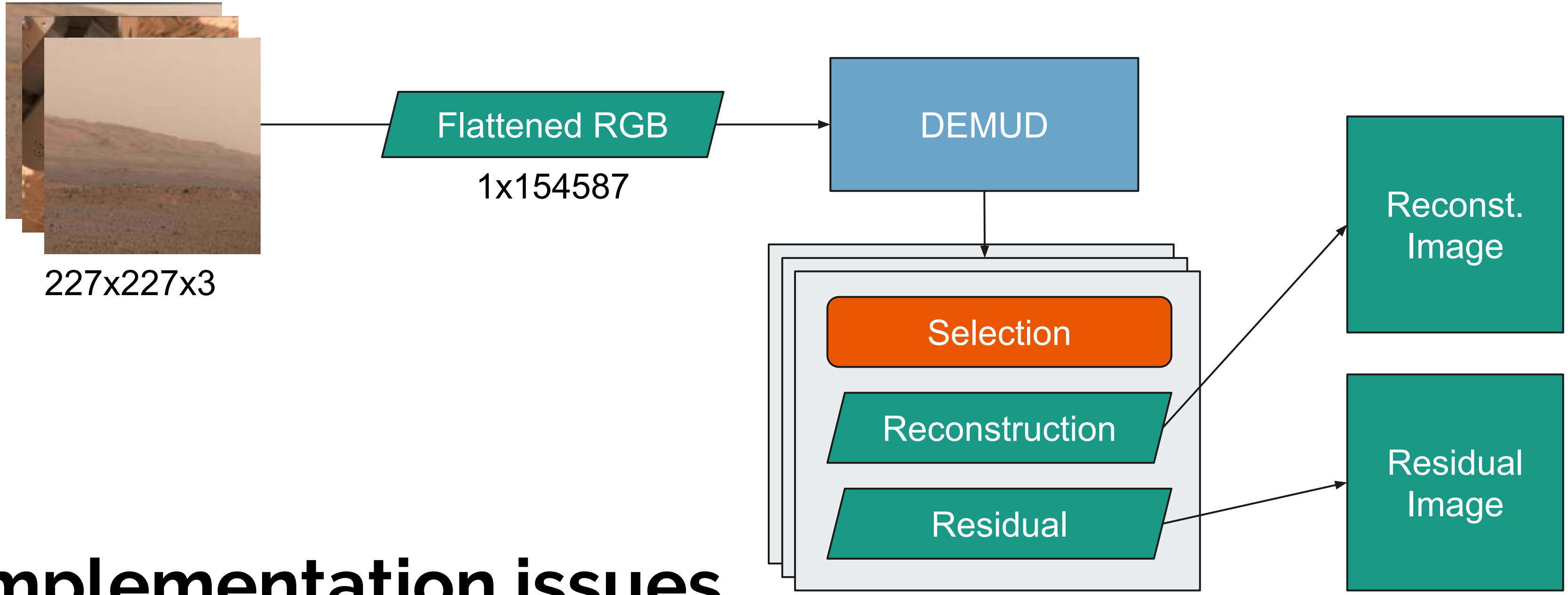
Table 1. First ten ChemCam targets selected by DEMUD from 11,750 spectra collected during MSL's first 90 sols on Mars.

Sel	Target	Sol	Shot	Explanation (automatically generated by DEMUD)
1	Rocknest3_3	88	2	Reduced Ca, Na, and O
2	Epworth	72	8	Elevated Ca; reduced Na
3	Kam	43	18	Elevated C, Si, and Al
4	Rocknest3_3	88	22	Reduced Ca and O
5	Kenyon	82	25	Elevated O; reduced Ca and Na
6	Murky	22	19	Elevated Mg
7	Rocknest3_3	88	28	Elevated Ca
8	Stark	15	48	Elevated C, Si, and O; reduced Al
9	Kilian	72	10	Reduced Ca
10	Thor_Lake	22	34	Elevated Ba; reduced Ca and K

Wagstaff, K. L., N. L. Lanza, and R. C. Wiens. "Unusual ChemCam Targets Discovered Automatically in Curiosity's First Ninety Sols in Gale Crater, Mars." *Lunar and Planetary Science Conference*. Vol. 45. 2014.

DEMUD + images? DEMUD expects 1-dim vectors...





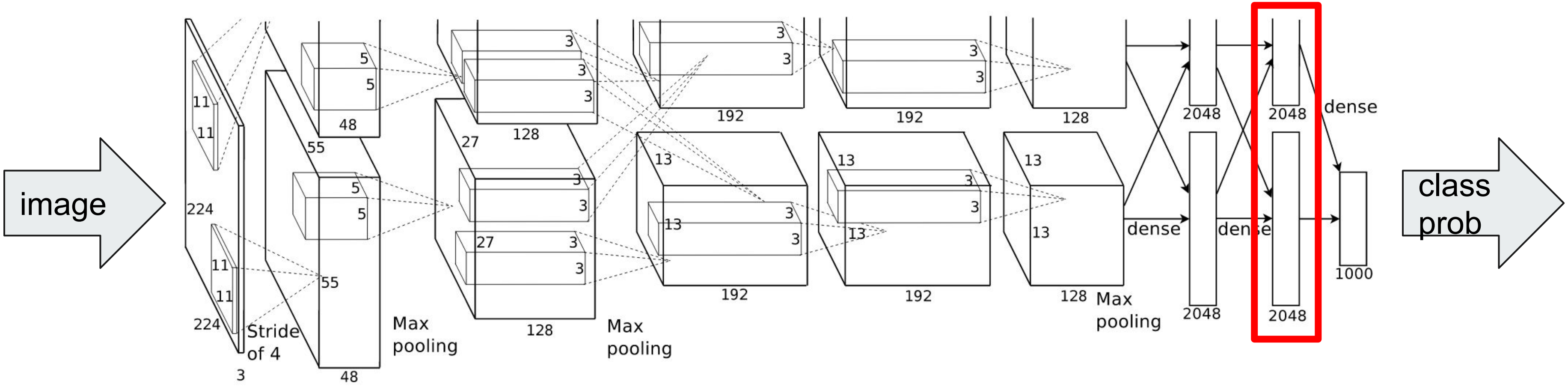
Implementation issues

- We lose a lot of information
  - shapes, texture, edges, etc.
- Seemingly small changes make a big difference
- Residual difficult to interpret - more on that later

Solutions?

- Computer Vision Techniques
  - Histogram of Oriented Gradients (HOG)
  - Scale-Invariant Feature Transform (SIFT)
- Convolutional Neural Networks
  - Popularized in 2012 for image classification
  - Neural network trained on labeled image data
  - Learns abstract features

Using CNNs for feature representation

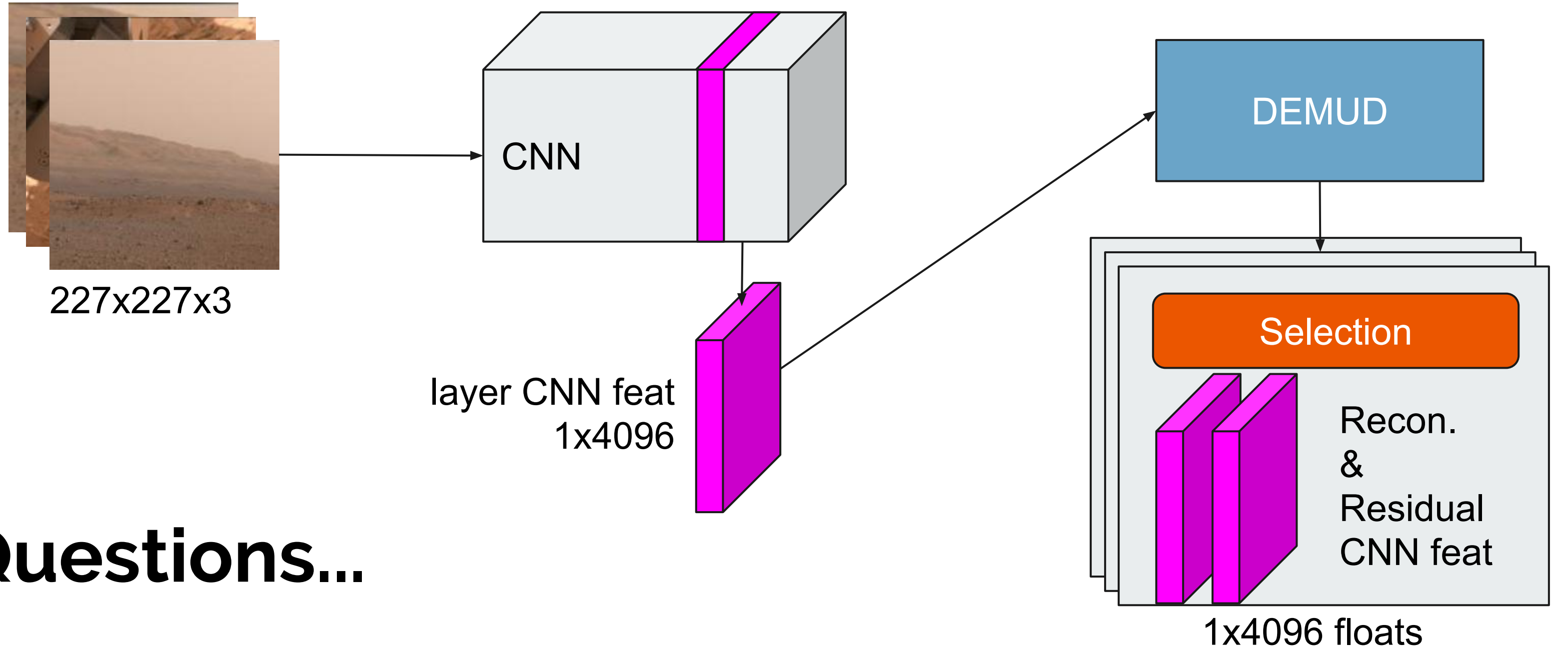


Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

DEMUD implementation is easy!

1. Install PyCaffe
2. Get feature vectors from a layer's activations (say, fc7)
3. Apply DEMUD to CNN feature vectors
4. ...
5. Profit!

Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014.



Questions...

- Which layer do we use?
- DEMUD explanations/residuals?
    - 4096 floats
    - Not very interpretable...

Understanding [Visualizing] deep image representations by inverting them

Mahendran et al., 2015

Mahendran, Aravindh, and Andrea Vedaldi. "Understanding deep image representations by inverting them." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

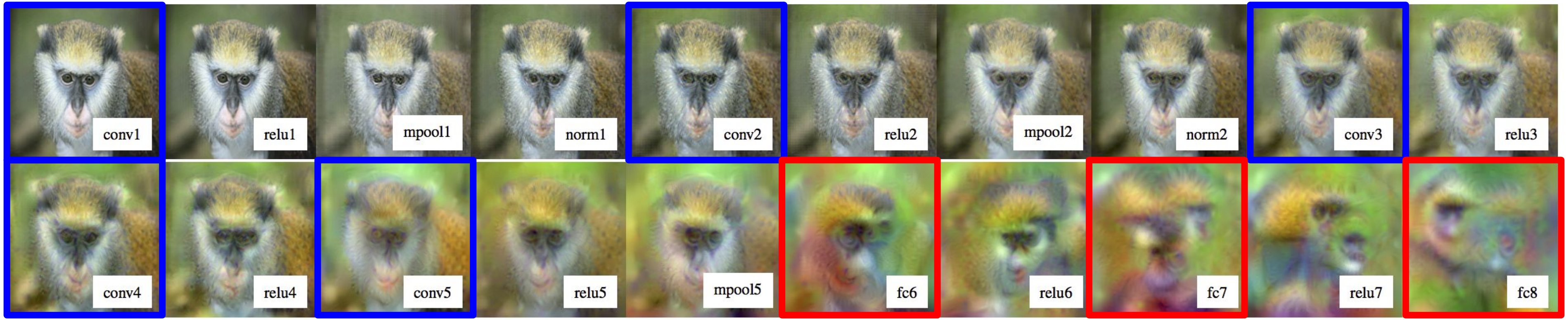
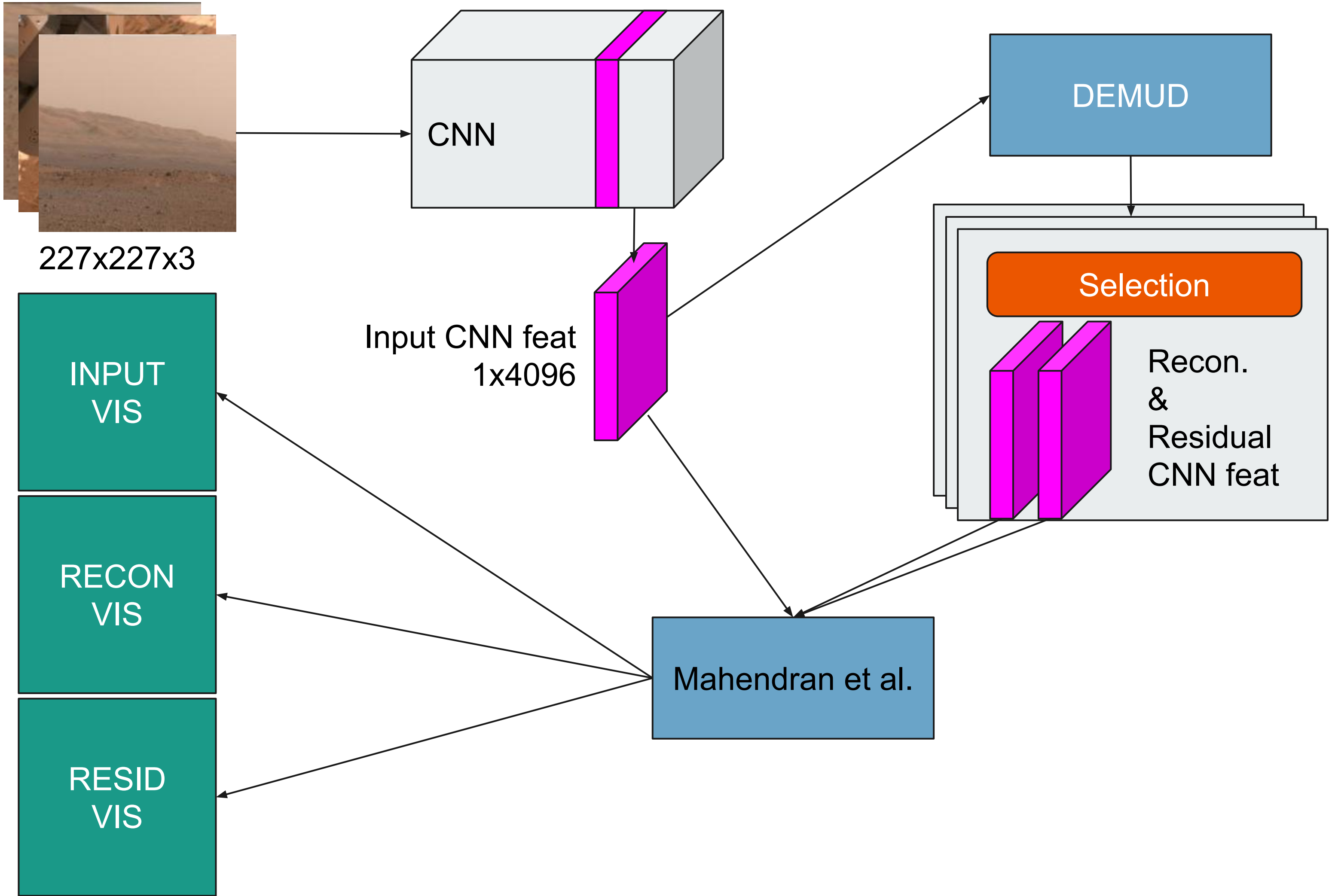


Figure 6. **CNN reconstruction.** Reconstruction of the image of Fig. 5.a from each layer of CNN-A. To generate these results, the regularization coefficient for each layer is chosen to match the highlighted rows in table 3. This figure is best viewed in color/screen.

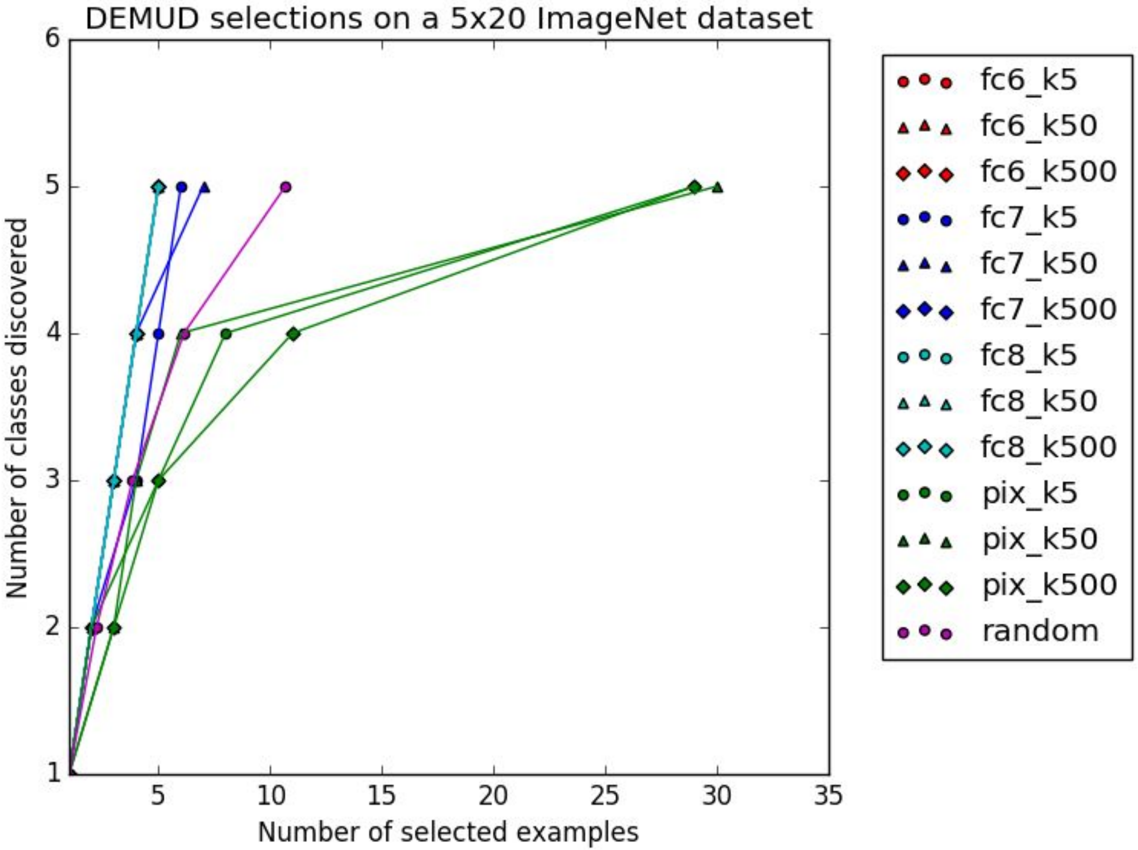

CNN
DEMUD
227x227x3
Selection
Input CNN feat
1x4096
Recon.
&
Residual
CNN feat
INPUT
VIS
RECON
VIS
RESID
VIS
Mahendran et al.

Experiments

An easy problem

- Dataset: 5 pre-labeled classes, 20 images each (from ImageNet)
    - Tiger Cub, Terrier, Mountain Bike, Sports Car, Volcano
- Different variations to compare performance
    - representations: fc6, fc7, fc8, and pixel-based
    - k-vals: 5, 50, 500

Images courtesy of the ImageNet database

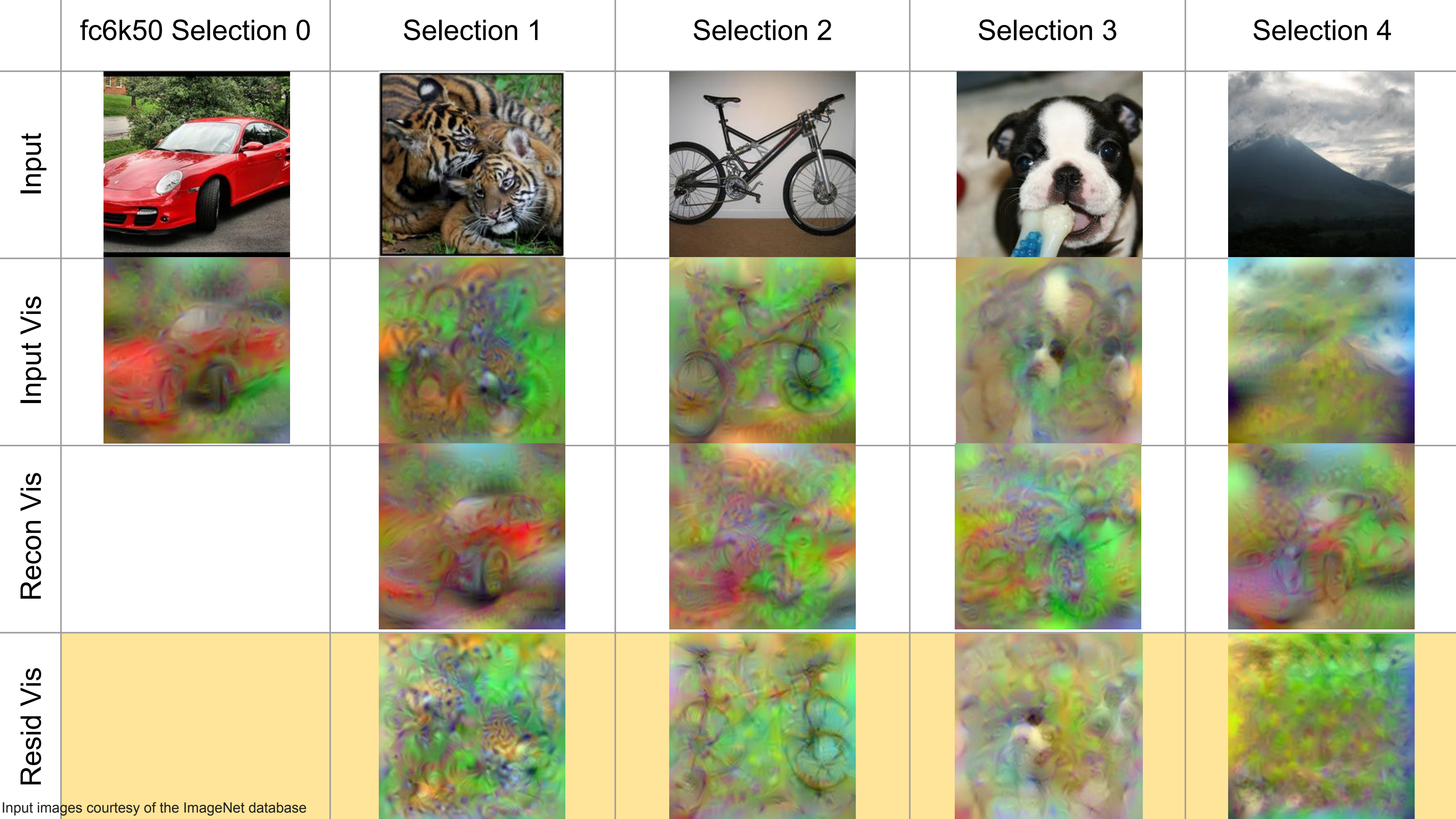


Performance plot - higher the slope, the better

https://math.stackexchange.com/questions/2172876/coupon-collector-without-replacement

	pixel Selection 0	Selection 1	Selection 2	Selection 3	Selection 4
Input					
Input Vis					
Recon Vis					
Resid Vis					

Input images courtesy of the ImageNet database

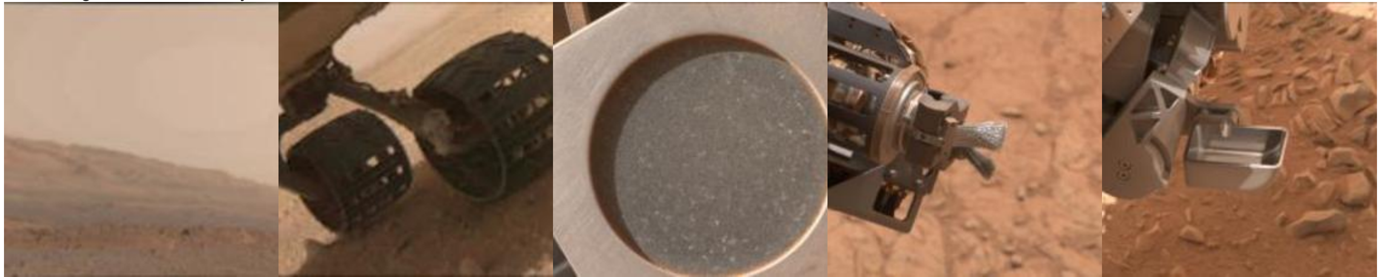


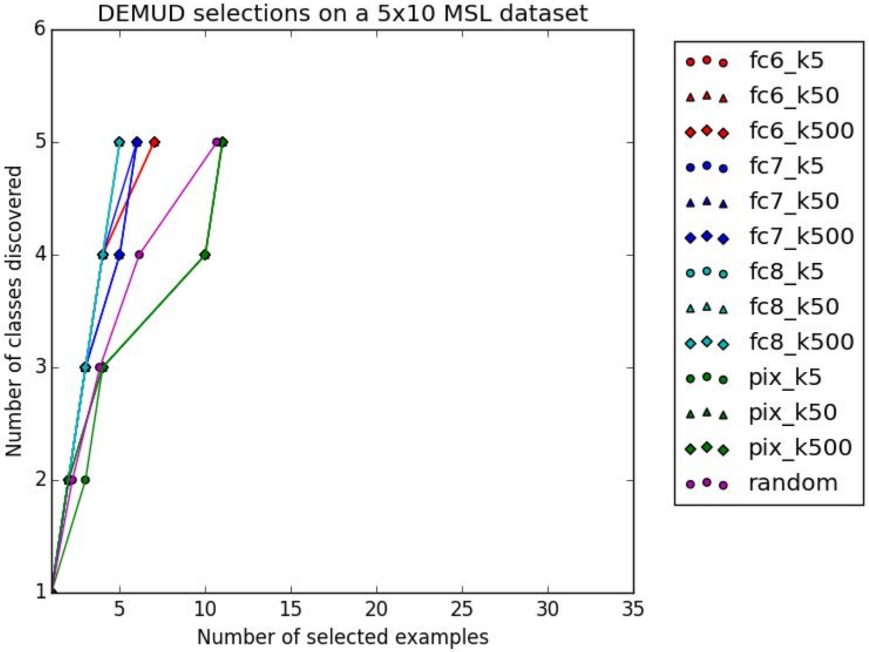
Input images courtesy of the ImageNet database

A harder, more relevant problem

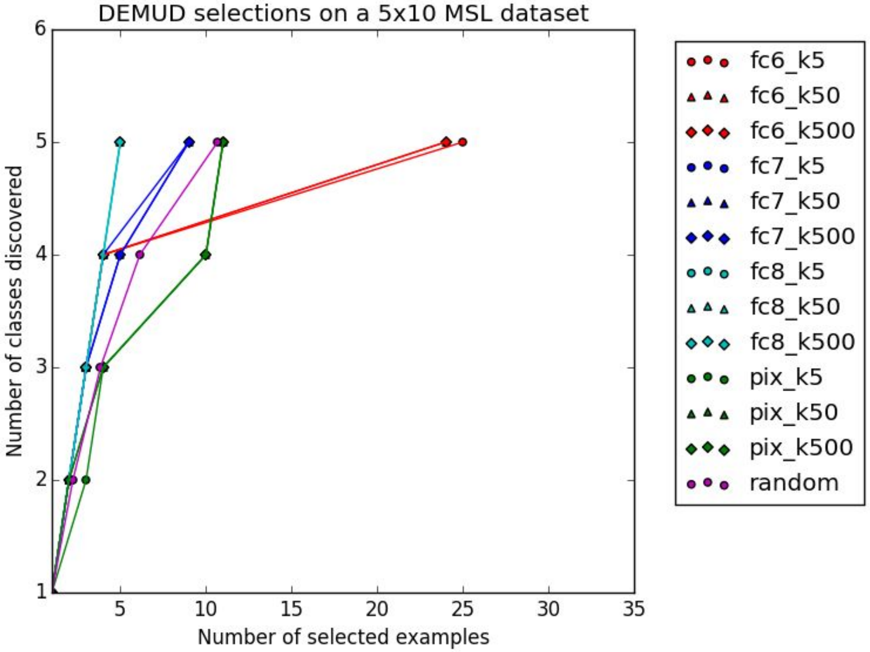
- Dataset: 5 pre-labeled classes, 10 images each (from MSL)
  - horizon ML, wheel MH, apxs_cal_target MH, drt_side MR, scoop ML
- In addition to previous variations...
  - ImageNet trained AlexNet, Fine-Tuned AlexNet
  - 25 classes, imbalanced, mixed sources,

MSL image data set courtesy of NASA/JPL-Caltech/MSSS





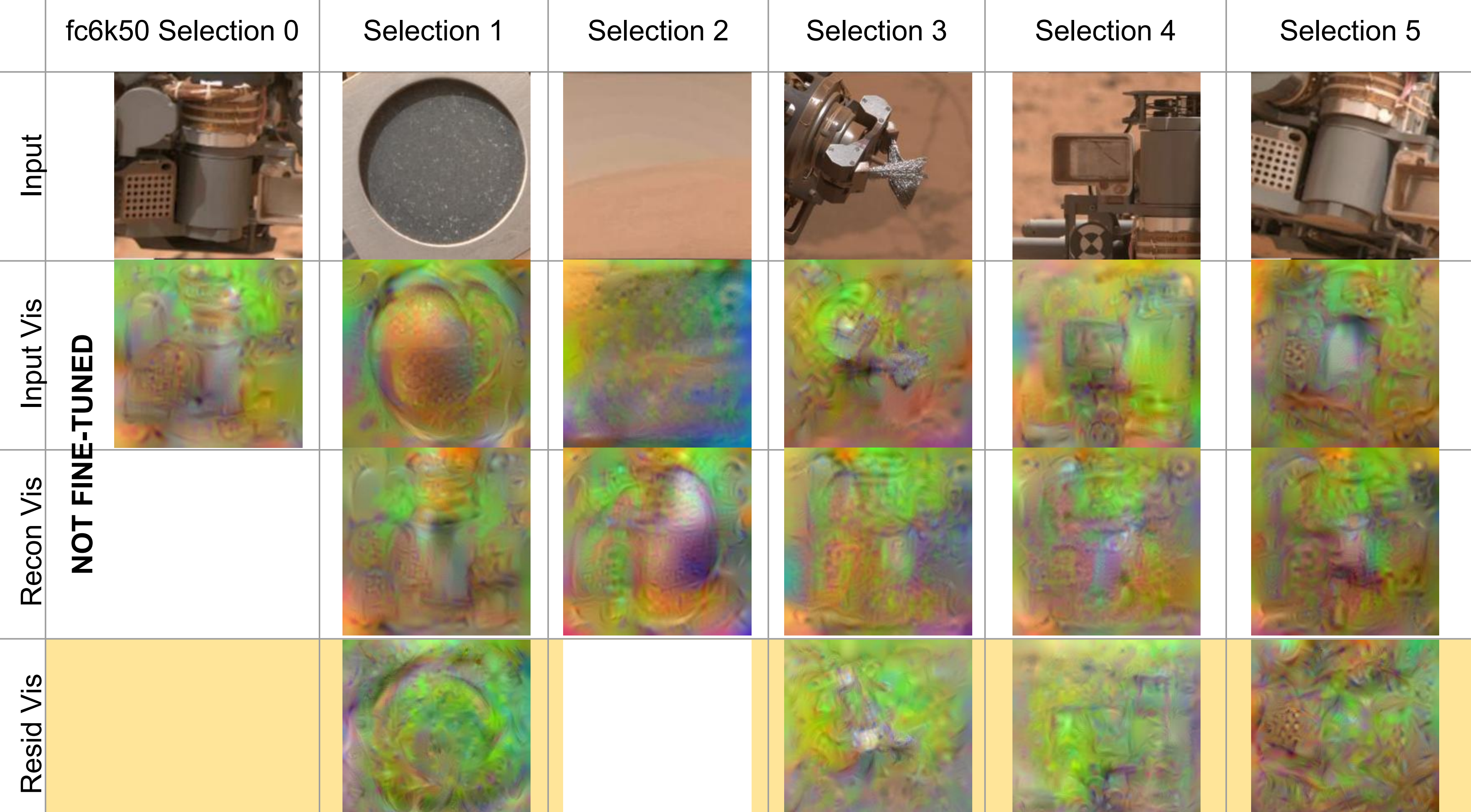
Feature vectors from CNN
only trained on ImageNet



Feature vectors from CNN
fine-tuned with MSL imagery

Performance plot - higher the slope, the better

https://math.stackexchange.com/questions/2172876/coupon-collector-without-replacement

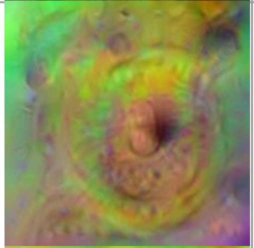


MSL image data set courtesy of NASA/JPL-Caltech/MSSS fc6k50 Selection 6

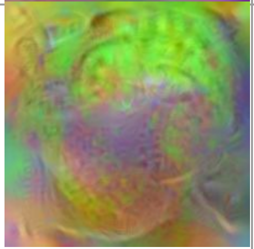
NOT FINE-TUNED

Input

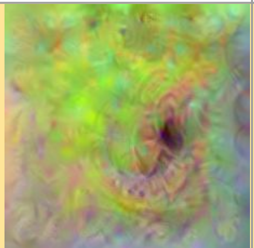
Input Vis



Recon Vis



Resid Vis



Conclusions

- By using CNN feature vector representations to motivate DEMUD selections…
    - Improved DEMUD's selection accuracy
    - Maintained ability to interpret residuals
- Furthermore, discovered
    - CNN feat vect reps have information than expected in its fc layers
    - Images in a different domain than original training could still work
    - CNN feat vect reps are more resilient to manipulation than expected

Limitations and Future Work

More questions

- Experiments:
  - Current experiments seem "too easy"
  - Only loaded datasets
  - Not so much "planetary imagery"
  - Fine-tuned network performance
  - Other layer, k val inversions

Even more Questions

- Broader topics:
  - Different or improved inversion methods
  - Different network architecture
- Wrap and automate the system
- Ultimately, apply to high-volume PDS data and analyze results

Dosovitskiy, Alexey, and Thomas Brox. "Inverting visual representations with convolutional networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

Acknowledgements

- Kiri Wagstaff, Umaa Rebbapragada
- Hannah Kerner, Brian Bue
- Other interns in the “hotel cubicle”
- PDS Imaging Node