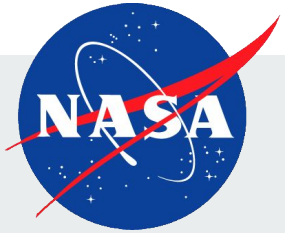


Improving the Interpretability of DEMUD on Image Data Sets

Jake Lee, Jet Propulsion Laboratory, California Institute of Technology
& Columbia University, CS ‘19
Intern under Kiri Wagstaff
Summer 2018

Government sponsorship acknowledged. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.
CL#18-4779

Motivation

- Ever-increasing of volume of image data in all fields
- Analysis of imagery is time-consuming and labor-intensive
- Lots of existing work on supervised image classification
    - Wagstaff, Kiri L., et al. "Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas." Conference on Innovative Applications of Artificial Intelligence. 2018.
- However, scientific discovery relies on unexpected observations

A brief introduction to DEMUD

- A prior-free novelty detection algorithm
- Prioritizes interesting data by attempting to discover all existing classes as quickly as possible
- Provides explanations for its prioritizations

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

- Ongoing work since Summer 2017
- Presented at 2018 ICML Workshop on Human Interpretability in Machine Learning

Wagstaff, Kiri L., and Jake Lee. "Interpretable Discovery in Large Image Data Sets." *2018 ICML Workshop on Human Interpretability in Machine Learning*. 2018. pp. 107-113

Represent

AlexNet
CNN fc6

Input Feature

Discover

DEMUD
Novelty
Detection

Selection

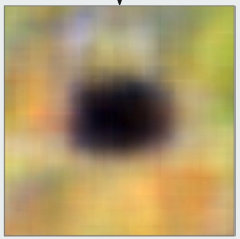
Expected
Feature

Novel
Feature

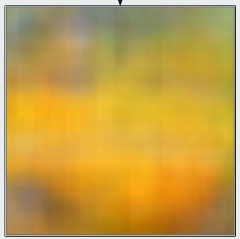
Selection - Expected = Novel

Explain

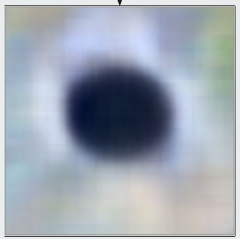
(Dosovitskiy & Brox, 2016) CVPR



Selection

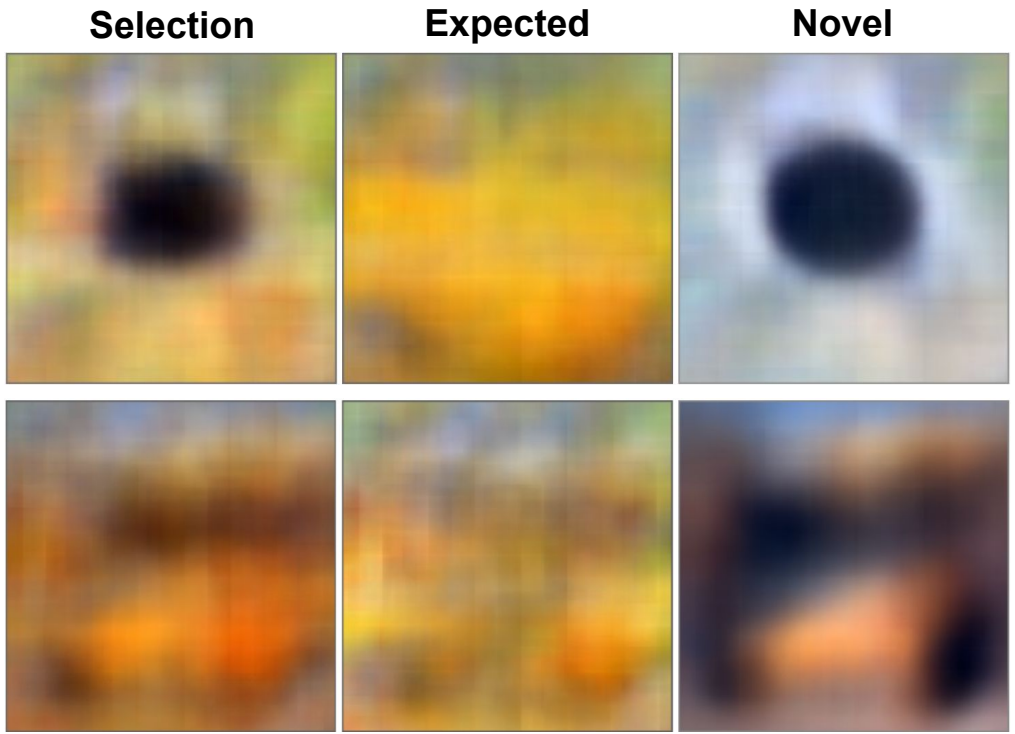


Expected



Novel

Example output ("Yellow" ImageNet dataset, fc6)



Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009.

Example output (MSL Image dataset, fc6)

Kiri L. Wagstaff, You Lu, Alice Stanboli, Kevin Grimes, Thamme Gowda, and Jordan Padams. "Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas." Proceedings of the Thirtieth Annual Conference on Innovative Applications of Artificial Intelligence, 2018. 10.5281/zenodo.1049137

Improve the Visualizations!

Represent

AlexNet CNN

Input Feature

Discover

DEMUD Novelty Detection

Selection

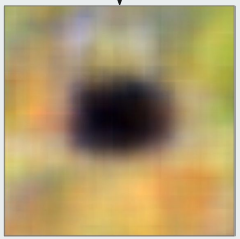
Expected Feature

Novel Feature

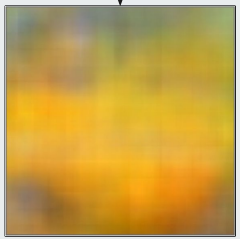
Selection - Expected = Novel

Explain

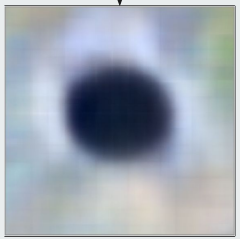
(Dosovitskiy & Brox, 2016) CVPR



Selection



Expected



Novel

A Better Visualization Method

- Dosovitskiy & Brox 2016 **CVPR**
- Dosovitskiy & Brox 2016 **NIPS**

Dosovitskiy, Alexey, and Thomas Brox. "Generating images with perceptual similarity metrics based on deep networks." Advances in Neural Information Processing Systems. 2016.
Dosovitskiy, Alexey, and Thomas Brox. "Inverting visual representations with convolutional networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

Dosovitskiy & Brox 2016 CVPR **"D&B L2"**

AlexNet fc6

Up-conv NN

l2-loss

optimizing

fixed

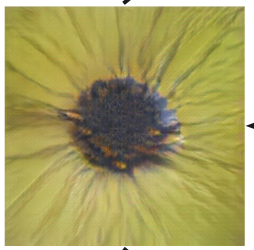
comp-loss

Comparator AlexNet Conv5

Dosovitskiy & Brox 2016 NIPS **"D&B ADV"**

AlexNet fc6

Up-conv NN



l2-loss

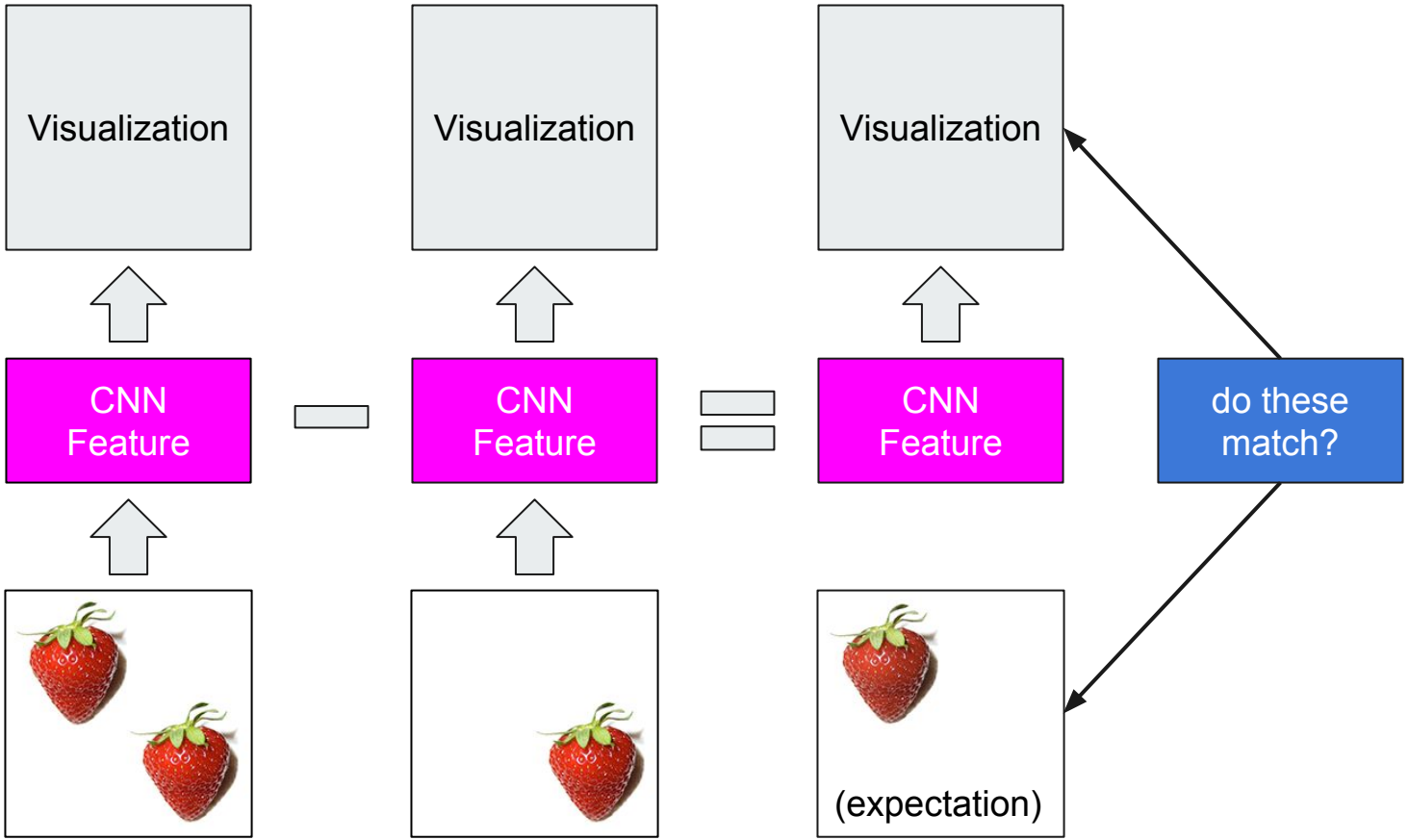
adv-loss

Discriminator

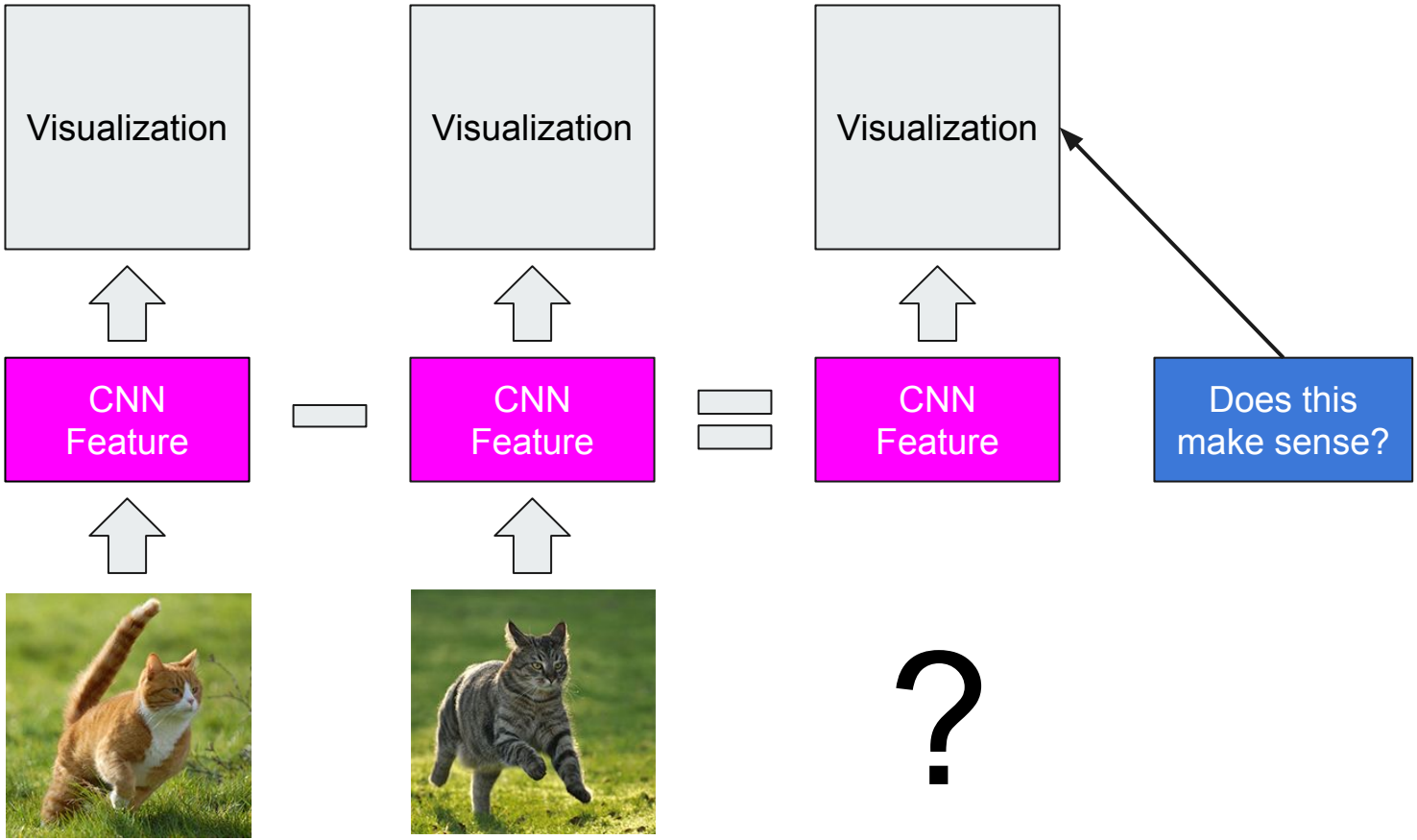
Evaluating Visualization Methods

- These methods were never intended to visualize modified feature vectors
- Specifically, DEMUD performs a subtraction in feature space for its explanations
    - **Selection - Expected = Novel**
- Unclear whether visualizing modified features can be meaningful

Simple Image Arithmetic



Simple Image Arithmetic

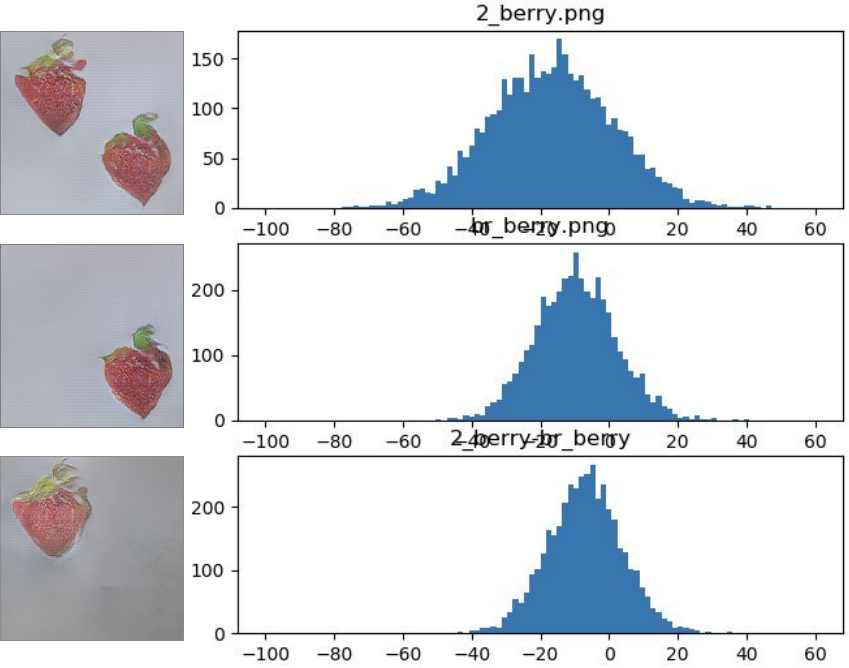


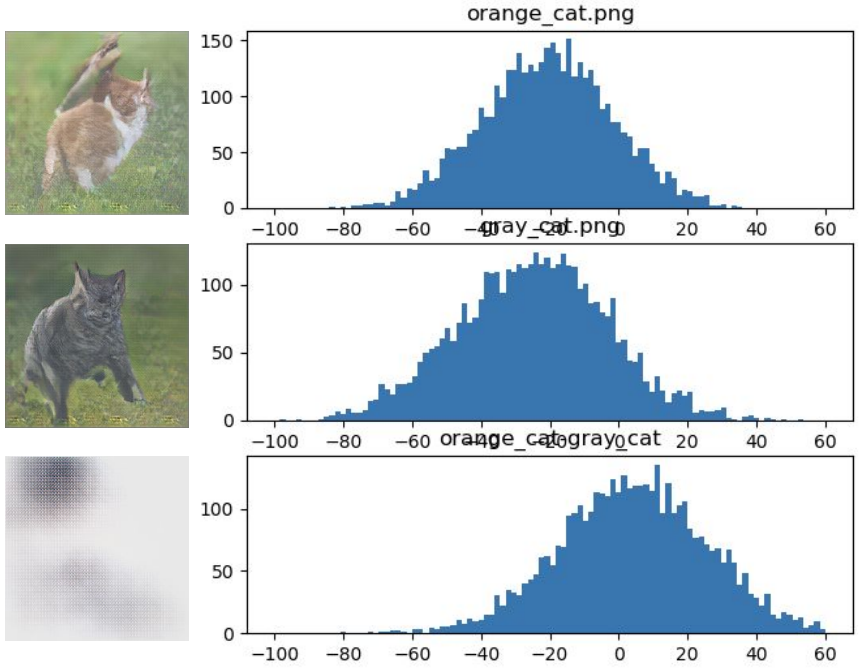
D&B L2 Results

D&B ADV Results

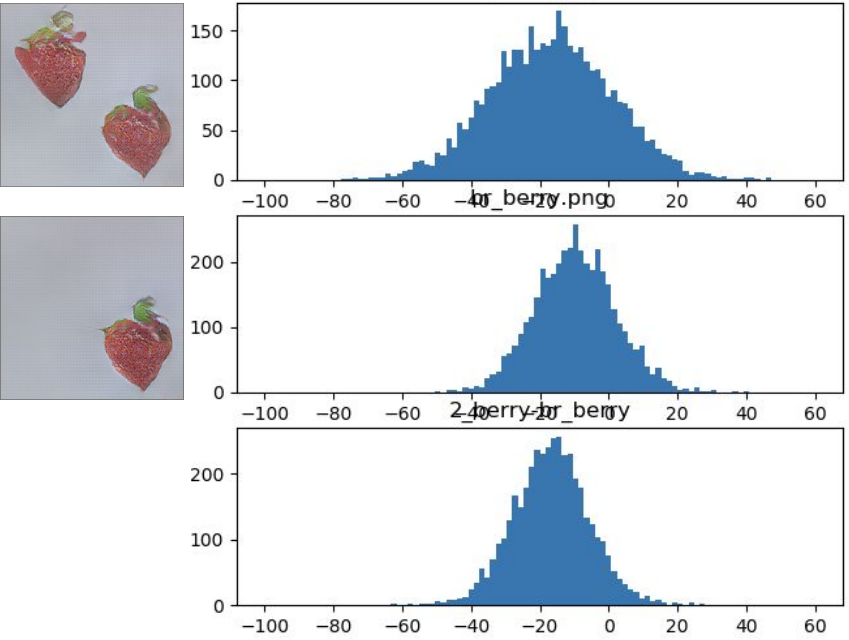
😢

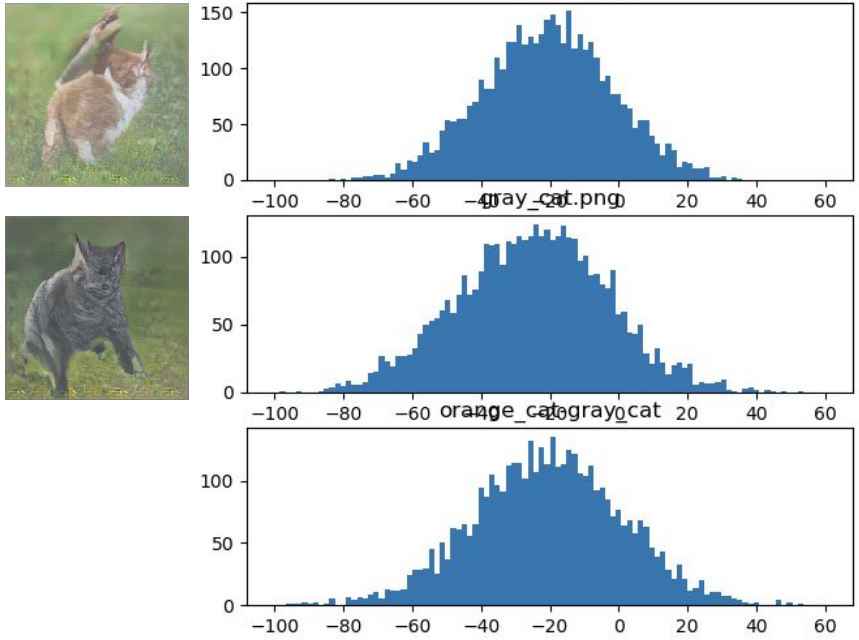
Plotting the feature distribution (fc6, 4096 values)



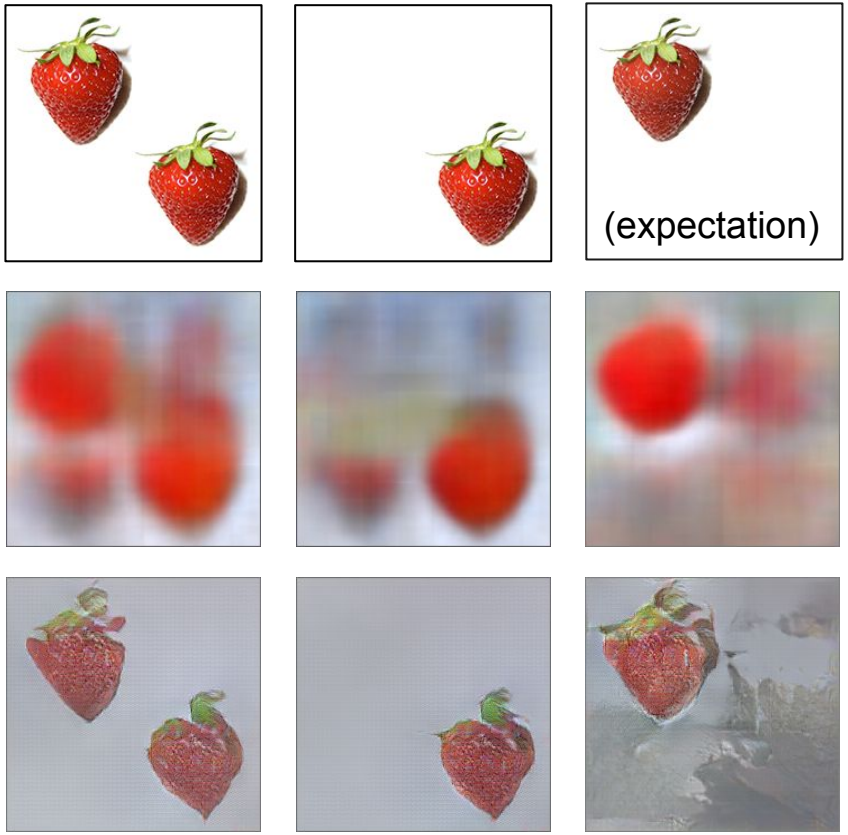


Mean-shift normalization

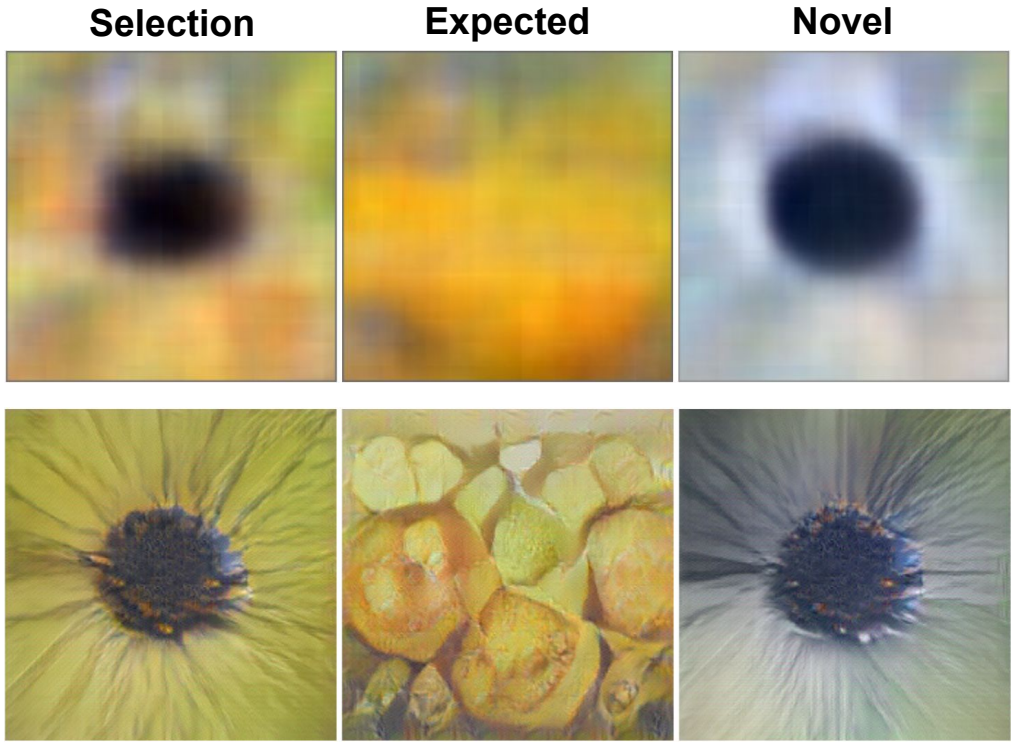




D&B ADV Results (with mean-shift normalization)

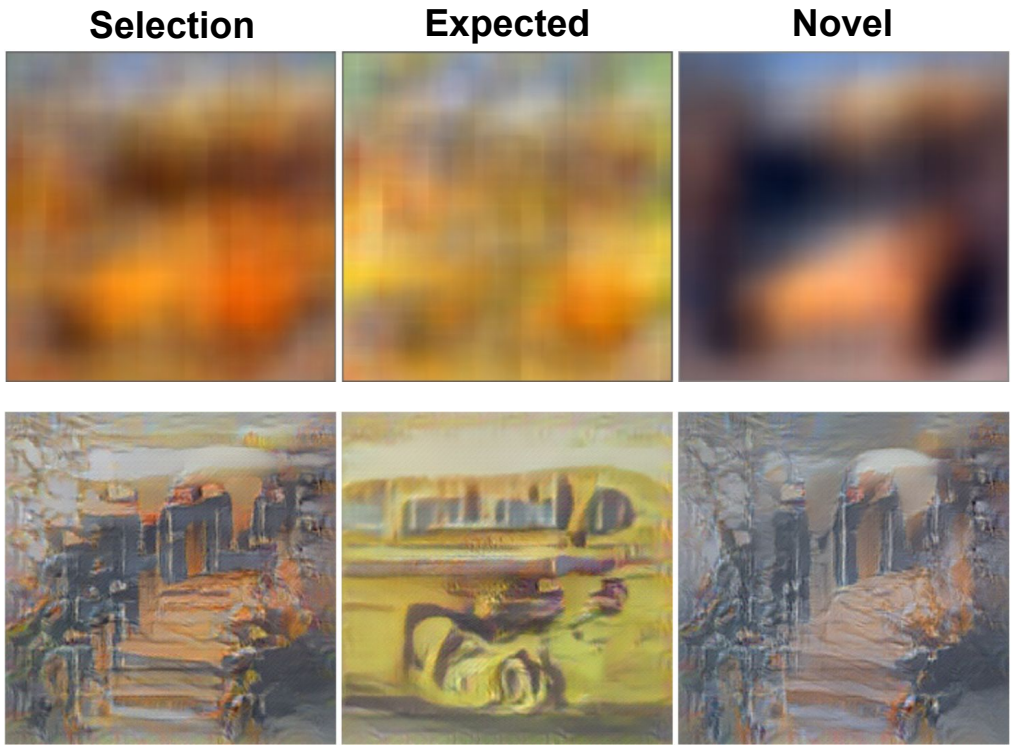


Improved DEMUD visual explanations



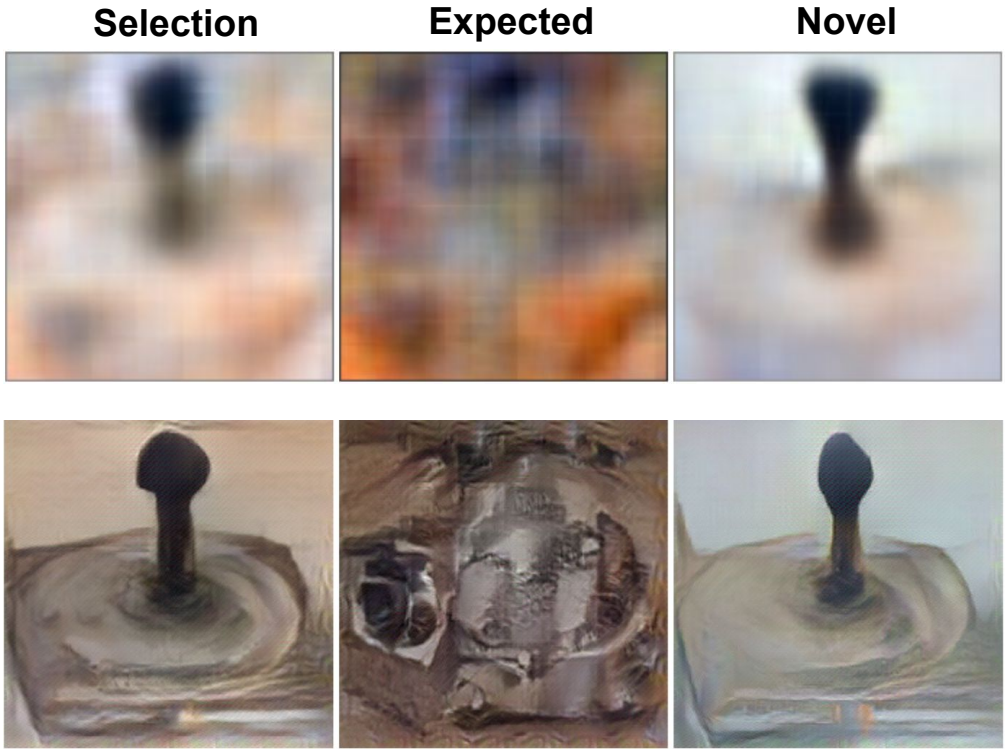
Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009.

Improved DEMUD visual explanations



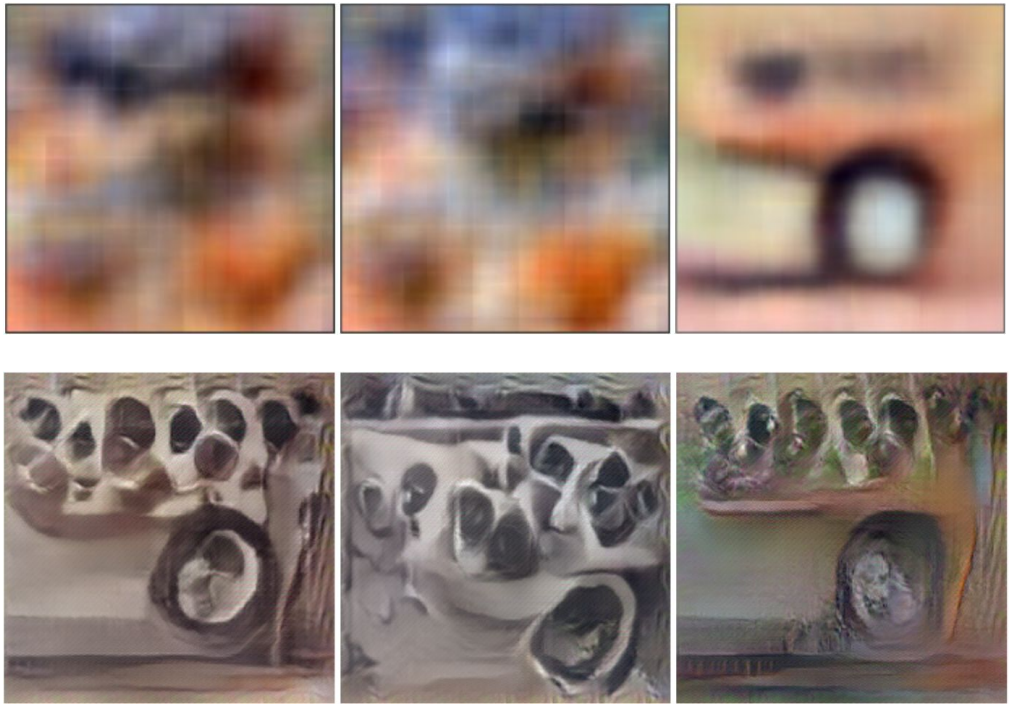
Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009.

Improved DEMUD visual explanations



Kiri L. Wagstaff, You Lu, Alice Stanboli, Kevin Grimes, Thamme Gowda, and Jordan Padams. "Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas." Proceedings of the Thirtieth Annual Conference on Innovative Applications of Artificial Intelligence, 2018. 10.5281/zenodo.1049137

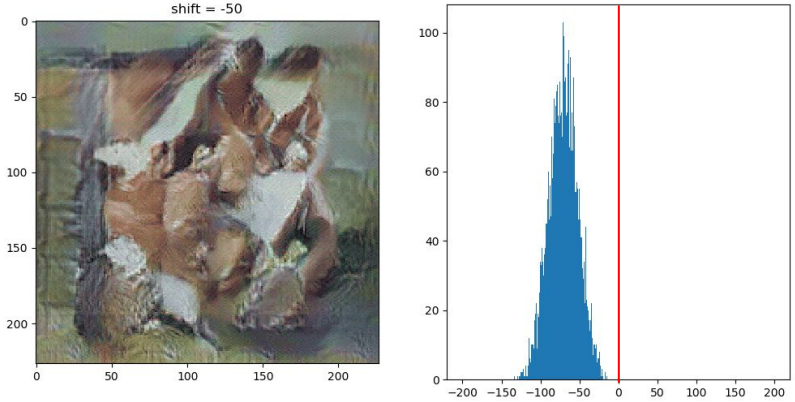
Improved DEMUD visual explanations



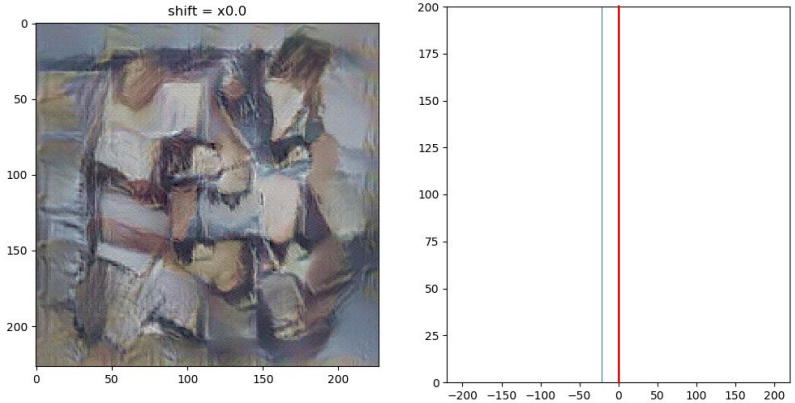
Kiri L. Wagstaff, You Lu, Alice Stanboli, Kevin Grimes, Thamme Gowda, and Jordan Padams. "Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas." Proceedings of the Thirtieth Annual Conference on Innovative Applications of Artificial Intelligence, 2018. 10.5281/zenodo.1049137

Visualization sensitivity analysis

Mean-shift



Dist-scale



Ongoing work

- More investigation into feature-level interactions and operations
- Visualizations for fc6, fc7, fc8
- More Experiments with DEMUD
- User Study

Acknowledgements

- Kiri Wagstaff
- Fellow summer interns
- Alexey Dosovitskiy
- PDS Imaging Node