Efficient Active Learning in New Domains
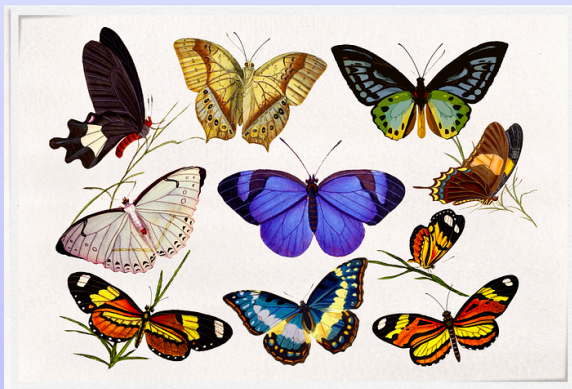
Kiri Wagstaff and Steven Lu

Jet Propulsion Laboratory, California Institute of Technology

Despite benefits, active learning is not used as often as we might predict. Why?

In real-world active learning (unlike artificial experiments), we cannot assume:


Assumption	Challenge
All classes are known in advance	Class discovery
Examples of all classes exist	Cold start problem
Optimal classifier hyperparameters are known	Hyperparameter search
Labelers can wait for next selection	Fast active selection

Questions to enable real-world use of active learning:
1) How best to discover all classes (during learning)?
2) How to reduce classifier cost (and hyperparameter search cost)?

Active learning for Mars rover images

- Real setting: 54,850 images from MSL rover + more each day
  - We don't know the full list of classes: we are exploring the unknown
  - Classes are severely imbalanced

- Goal: label Mars rover images (to train a CNN) efficiently
  - Perfect setting for active learning (and class discovery)!

- Best class discovery: DEMUD; most efficient learning: random sel.
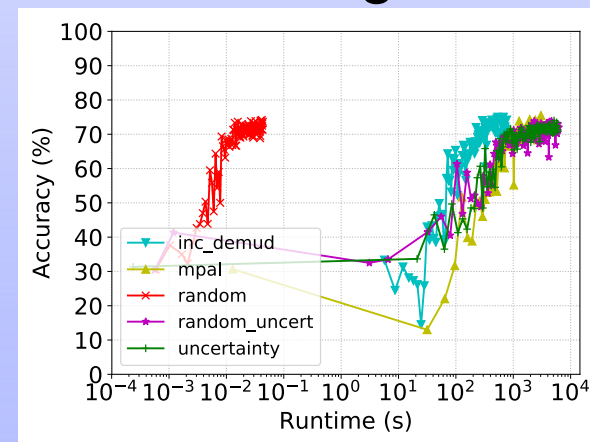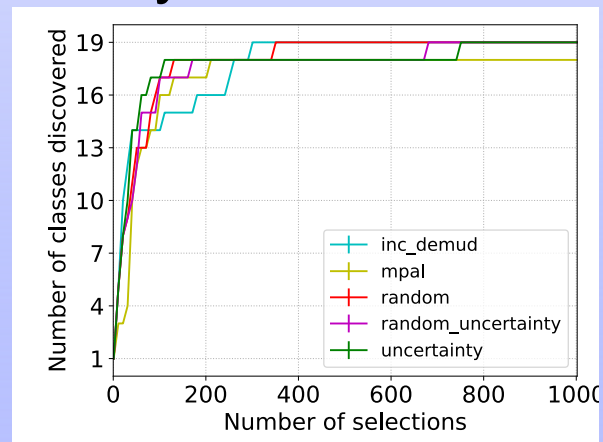
Contact: kiri.wagstaff@jpl.nasa.gov and you.lu@jpl.nasa.gov
Thank you: NASA Planetary Data System Imaging Node