Human-Machine Collaboration for Fast Land Cover Mapping

Caleb Robinson, Anthony Ortiz, Kolya Malkin, Blake Elias, Andi Peng, Dan Morris, Bistra Dilkina, Nebojsa Jojic

Problem

- ***Semantic segmentation*** of satellite imagery
  - Label each pixel in an image as ***water***, ***forest***, ***low vegetation***, or ***impervious surface***
- ***Applications***: urban planning, ecology, preservation

Challenges

- ***Domain adaptation***: limited training data in few geographic areas. Models generalize poorly in different lighting conditions, seasons, geographies.
- Neural nets are ***data hungry*** when adapting to new domains
- Standard neural network architectures make ***fatal, un-explainable mistakes***; human experts cannot trust purely machine-learned models

Our Approach

Leverage complementary strengths of ***humans*** and ***machines***.

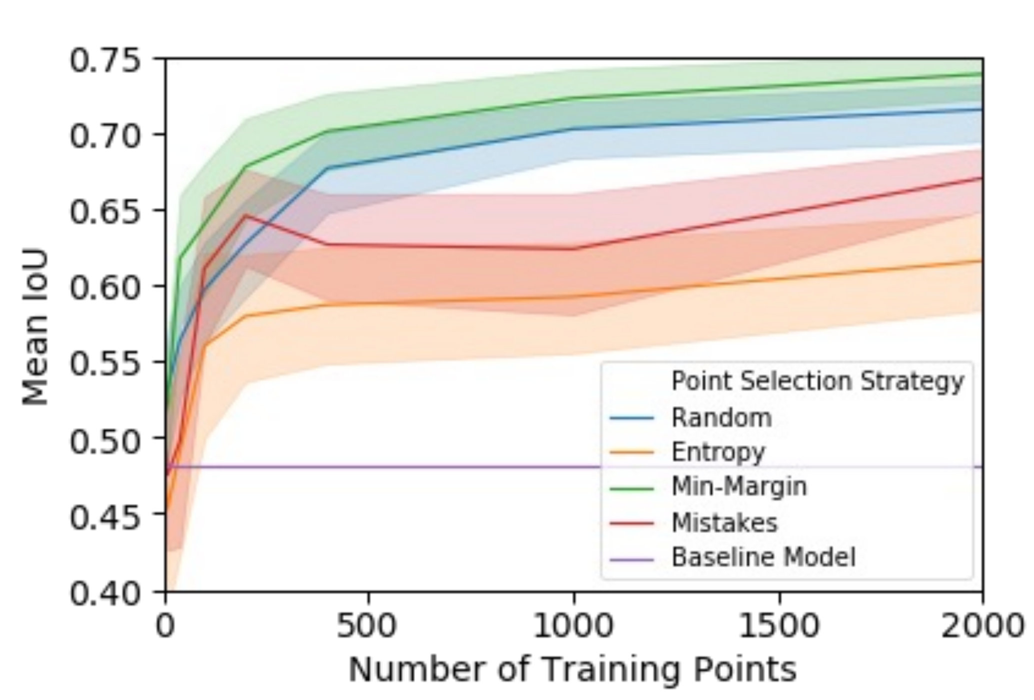
- ***Human***: high-level, near-instant scene understanding
- ***Machine***: ability to learn from data, amplify human work
- Evaluate performance of the ***combined system.***
- ***Labels*** are the final product; ***model is auxiliary.***

Model and Training

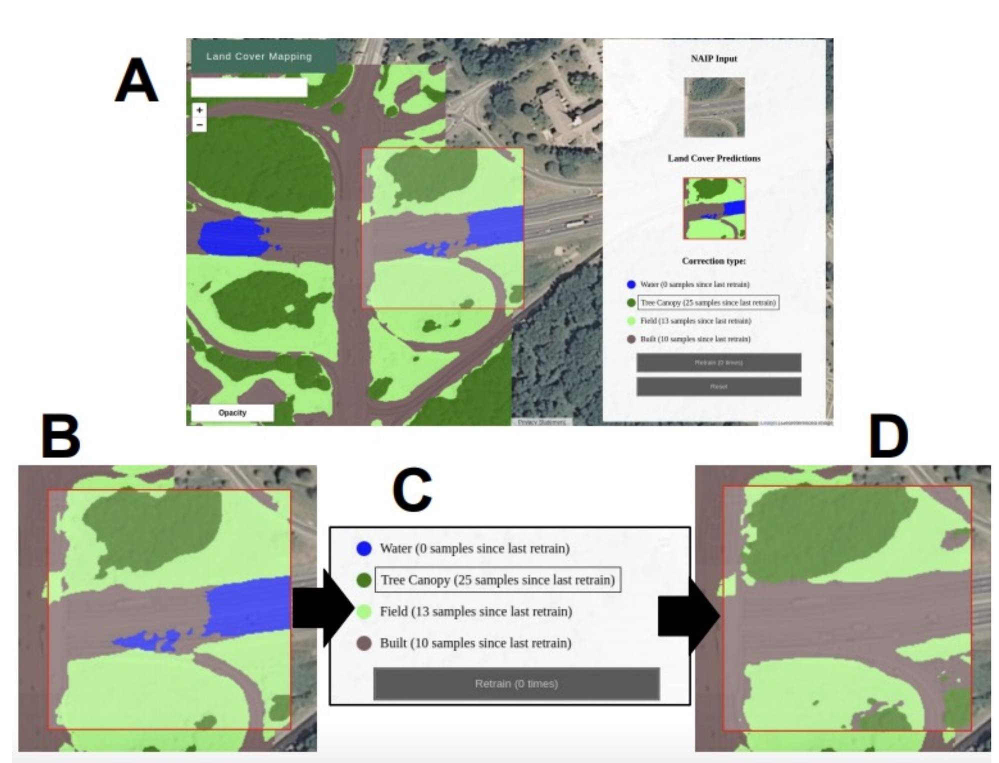
- Train a U-net architecture on 90,000 randomly-selected 240 x 240 px images from the state of Maryland (data set provided by Chesapeake Conservancy).
- Provide a small amount of training data in a new geographic area (10 – 2000 of new labeled pixels)
- Run additional training via:
  - Dropout (search for a set of neurons to remove yielding higher accuracy)
  - Gradient descent on a subset of the weights
- Best results from training either last 2 or last 3 layers

Seeking Sample Efficiency

- Neural networks are ***data hungry***; labeling is expensive.
- Uncertainty-based ***active learning*** (labeling points with most uncertain predictions) is a natural first attempt
- Only slightly better than ***randomly-chosen*** training points
- Training where model makes errors ***performs worse than random.***



Hypothesis: humans may better identify points worth labeling, and achieve higher sample efficency, compared with standard active learning query methods.



Web interface for human labelers.

Experiments

Random Labeling Tool

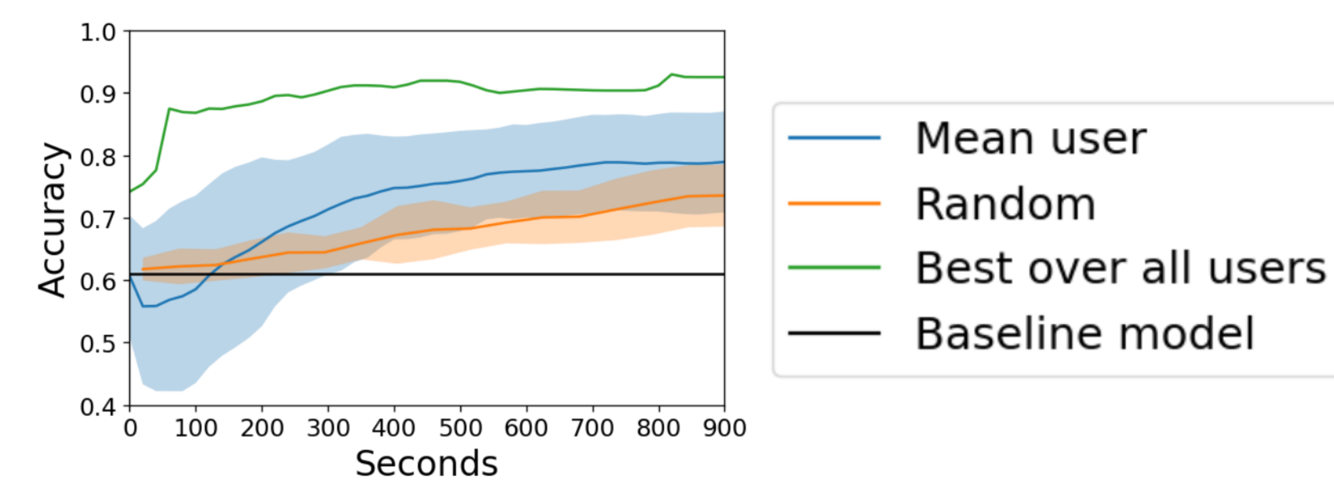
- Crowdsourced collection on mTurk to acquire unbiased ground truth labels in four 85 km^2 areas in New York (6009 labels total, 91.1% in accordance with Chesapeake Bay data)

Hybrid Labeling Tool

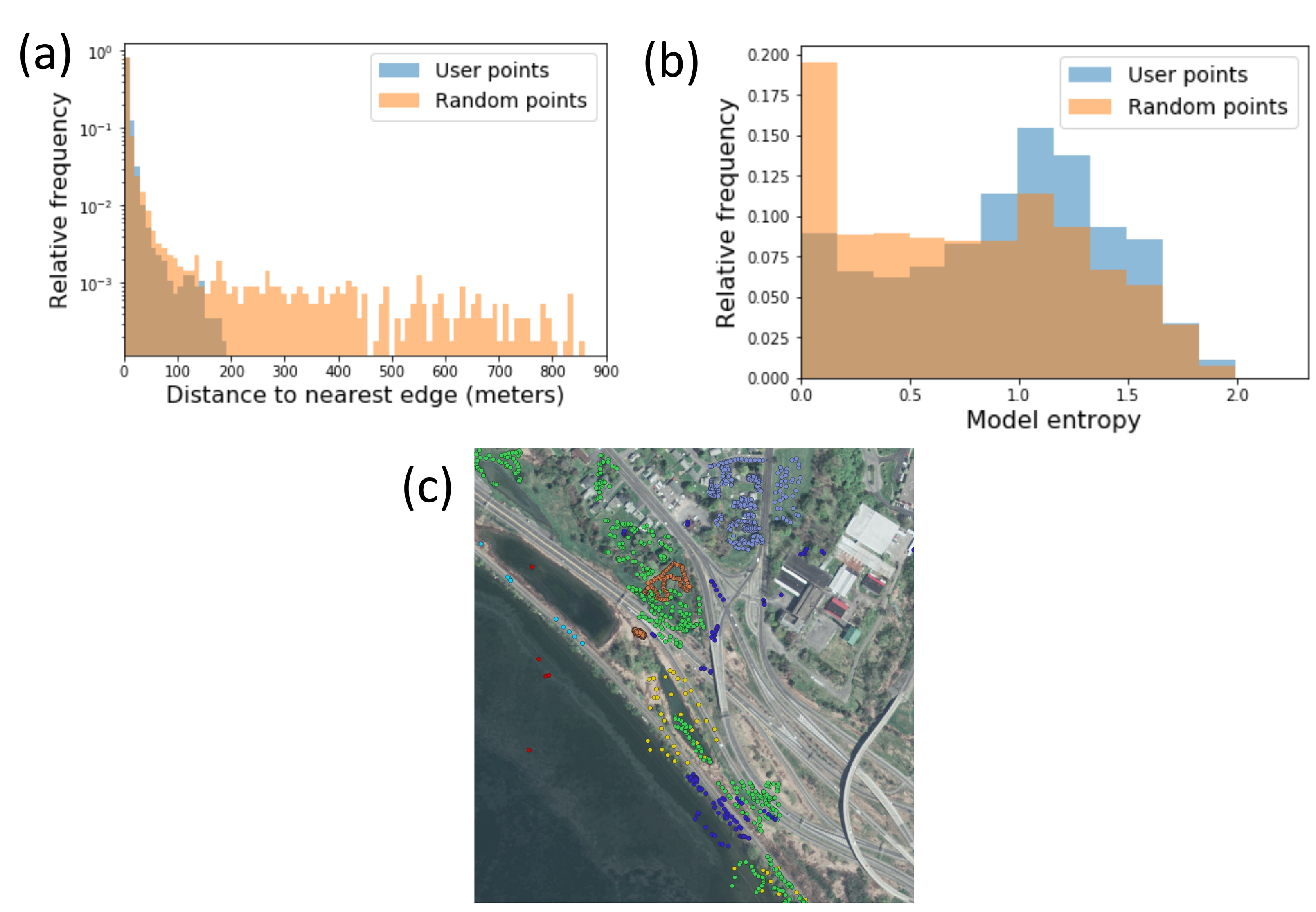
- Built web interface where users ***observe predictions*** of a model and provide ***more labels*** as needed.
- Using our web interface, we had 50 mTurkers fine-tune the pretrained baseline model in four 15-minute sessions on four target areas with two adaptation methods
- We compare the performance with the Random query method using the same ground truth dataset

Results

- Humans out-perform random query method
- Best-performing users could potentially reduce a ***10-month, $1.3m*** manual labeling effort, to ***925 hours*** and ***$18.5k***.
- Better performing users are detectable: performance in trial 2 predicts performance in trial 3 ($p < 0.01$, $\rho = 0.4$).
- Trials 2 and 3 are less predictive ($\rho = 0.1$) of trial 4, when the underlying fine-tuning method is switched – indicating users have adapted to the specific learning algorithm



- Users label points that are:
  - ***Near edge features*** in the image. (a)
  - ***Medium/high entropy*** in model prediction. (b)
  - Concentrated in ***select sub-areas***. (c)
  - ***Balanced*** among the 4 classes. In one trial:
    - User label classes ~ [18.8%, 29.8%, 23.2%, 28.2%]
    - Underlying image ~ [8.5%, 53.9%, 35.6%, 2.0%]
- Various attempts to simulate users ***did not achieve better results*** than random query method -- while real users do.



Contributions

- A human-in-the-loop system for image segmentation.
- Evidence that human judgement enhances sample efficiency, making both ML and human labor more valuable than before.
- A call to incorporate human cognition in the loop, rather than trying to emulate it.