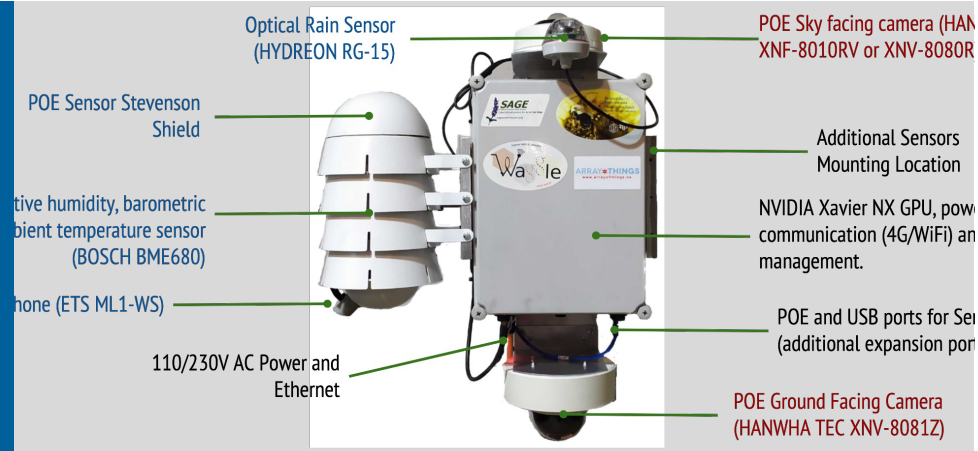


AI @ EDGE



Nicola Ferrier
Senior Computer Scientist
Mathematics and Computer Science Division

and Pete Beckman, Charlie Catlett, Scott Collis,
Neal Conrad, *Dario DeMatties Reyes*, Yongho
Kim, *Seongha Park*, *Bhupendra Raut*, Rajesh
Sankaran, Sean Shahkarami

ALCF AI for Science Student Training Series, October 22, 2024

What is Waggle?

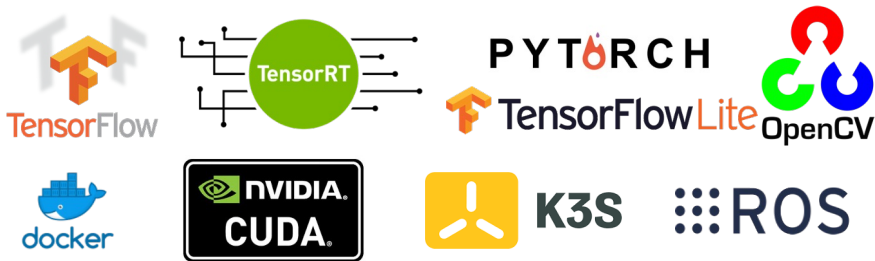
An Architecture and Software Stack for Scientific AI@Edge Computing

Waggle AI@Edge Software Stack

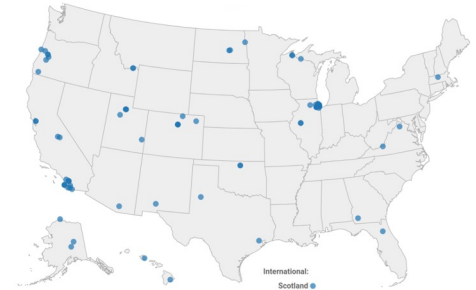
- AI toolchain for the edge
- Goal-based scheduler & resource manager
- Fully containerized AI@Edge applications
- Support for multi-tenancy
- Extreme cybersecurity
- Resilient data movement
- Cloud-based data store and management



Standardized Open Source AI tools for the Edge



Deployed at
Instrument/Sensor
outside or inside



Purpose-built
Waggle node for
remote
instrumentation



Commodity server
deployed with
Waggle Software



AI-Based Measurement & Anomaly Detection, & Control

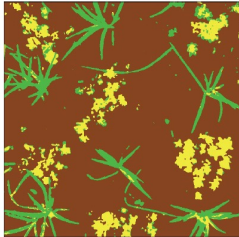
“SOFTWARE DEFINED SENSOR”



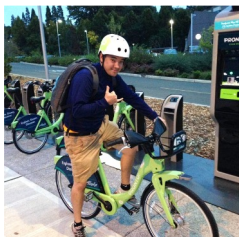
Your software container running here

Analysis produces live results

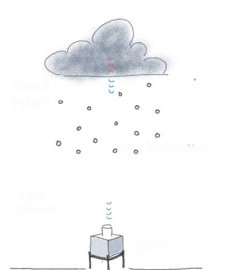
Plant Species




Helmet usage



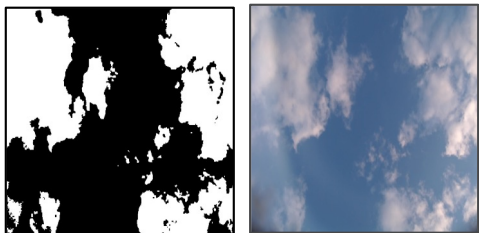
Weather classification to facilitate LiDAR interpretation




Traffic




Cloud type and coverage



Wildfires: detecting smoke



Flooding / surface water

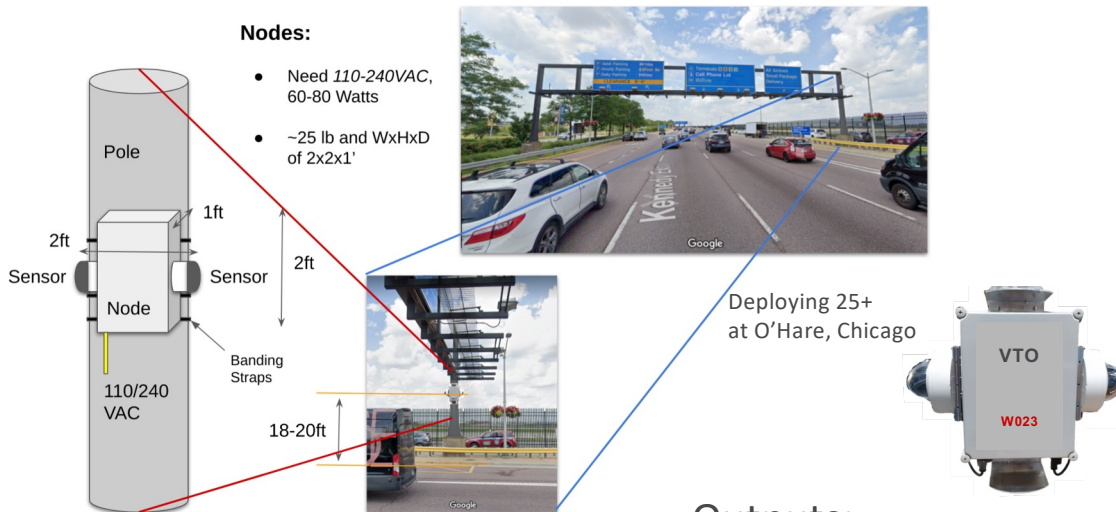


Many measurements cannot be “sensed” directly but can be computed from images, microphones or other datastreams/devices

portal.sagecontinuum.org/apps/explore

MEASURING TRAFFIC STATE USING VEHICLE TRACKING

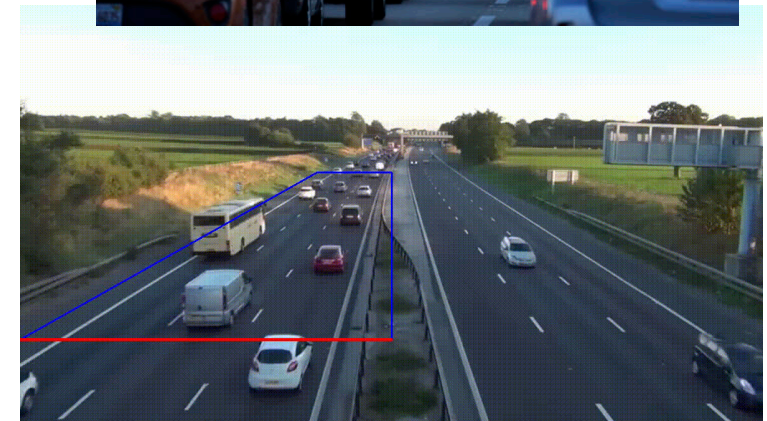
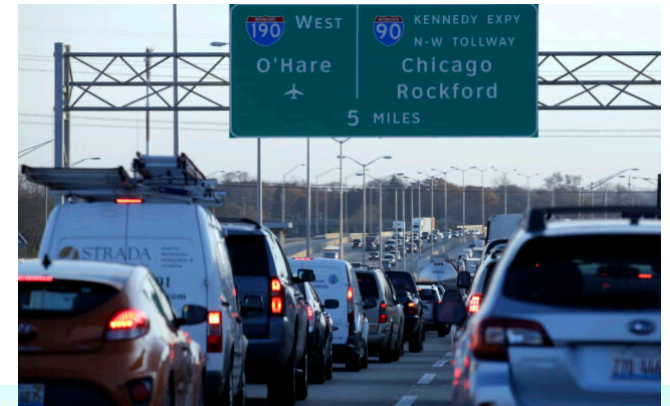
Chicago O'Hare Improvement Project: Understanding Traffic Type, Flow, and Density DOE VTO (EERE)



Dual cameras to capture approaching and leaving traffic for in-situ analysis.

Outputs:

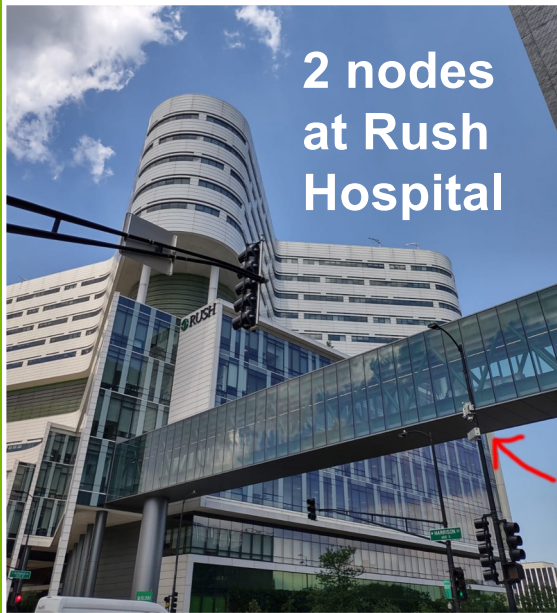
- Flow: # of vehicles / sec
- Speed: averaged speed in km/h
- Density: ($\#$ of vehicles \times average length based on vehicle type) / (length of road in ROI)



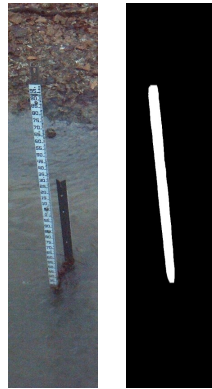
Advanced Radiation Sensors for Safer Cities

NNSA: Domain Awareness Waggle Network (DAWN - ANL) and PANDA (LBNL)

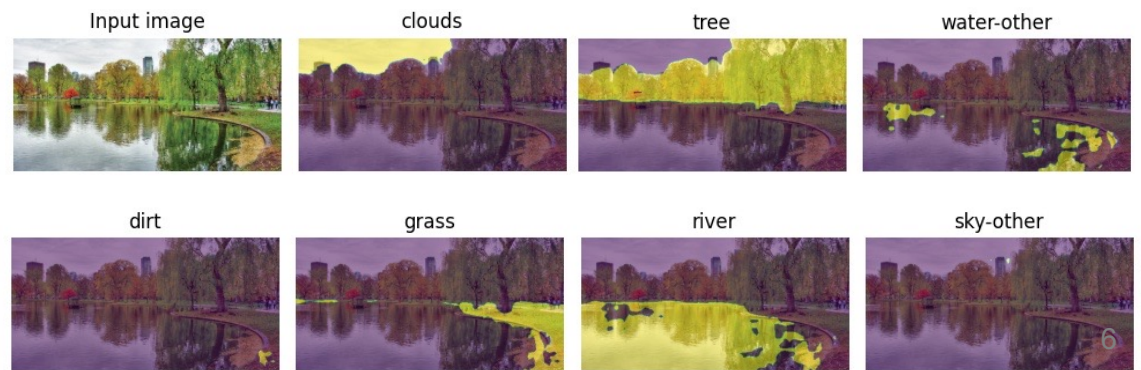
New data streams: Cameras + LIDAR + radiation sensors



Surface Water Detection



Linked with HPC, can be used to build hydrology models and predictive capabilities



Wildfire Detection and Prediction

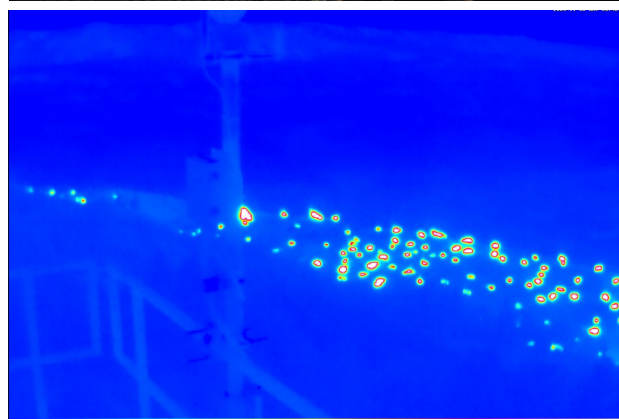
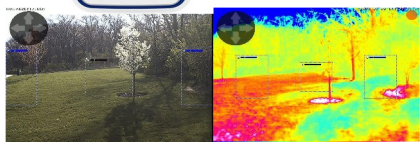
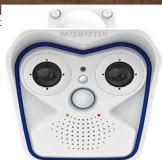
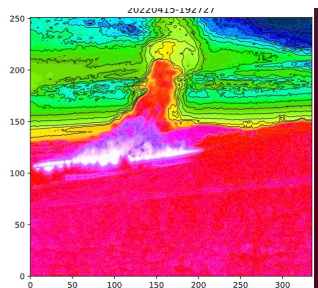
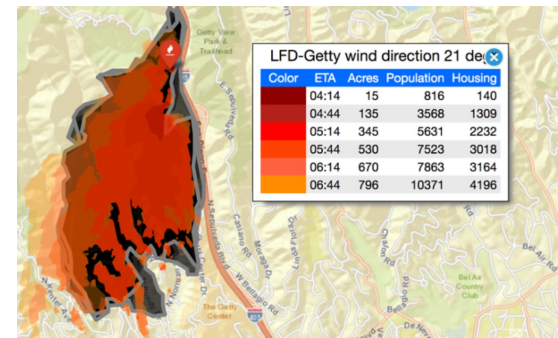
AI@Edge for wildfire detection linked to HPC simulations



Ilkay Altintas, UCSD, Co-PI for SAGE



Konza Prairie controlled burn: April 2022.



Oregon Sage Node, ALERTWildfire
Nov 2023

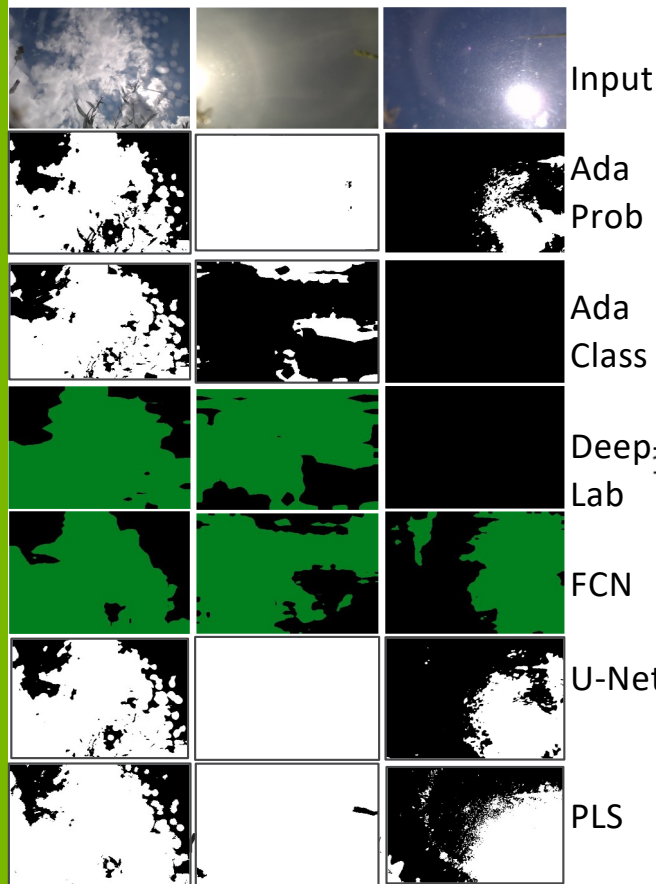
Sage project will move Pan-Tilt-Zoom cameras toward suspected outbreaks, and use infrared cameras to build self-supervised AI training



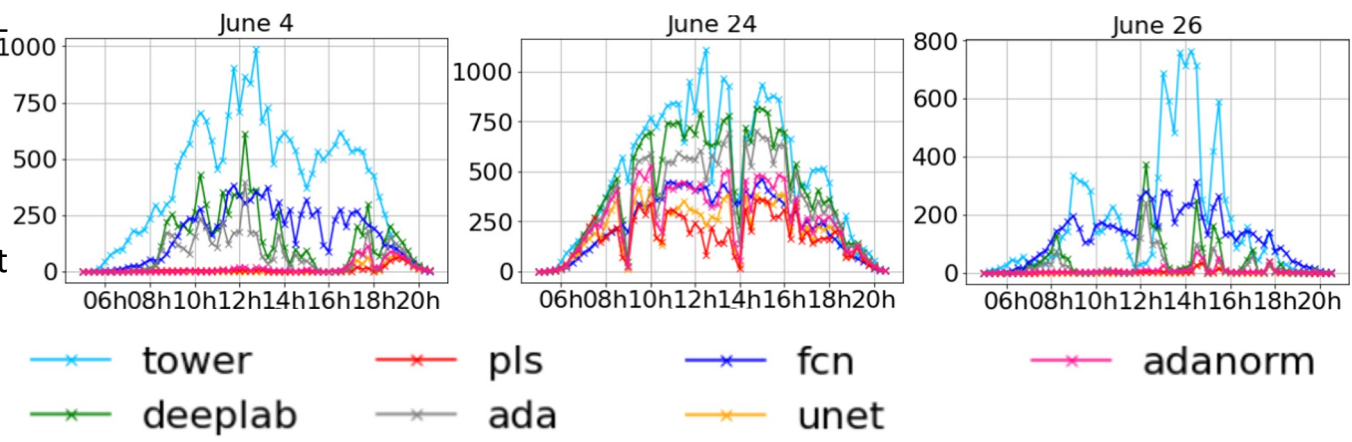
Exelon: Measuring Cloud Coverage for Estimating Solar Irradiance

- Cloud cover using FCN, U-Net, DeepLab v3, PLS, and AdaBoost models
- Solar irradiance = (1-cloud cover(ratio)) x max solar irradiance

Park, Seongha, et al. "Prediction of solar irradiance and photovoltaic solar energy product based on cloud coverage estimation using machine learning methods." *Atmosphere* 12, no. 3 (2021): 395.

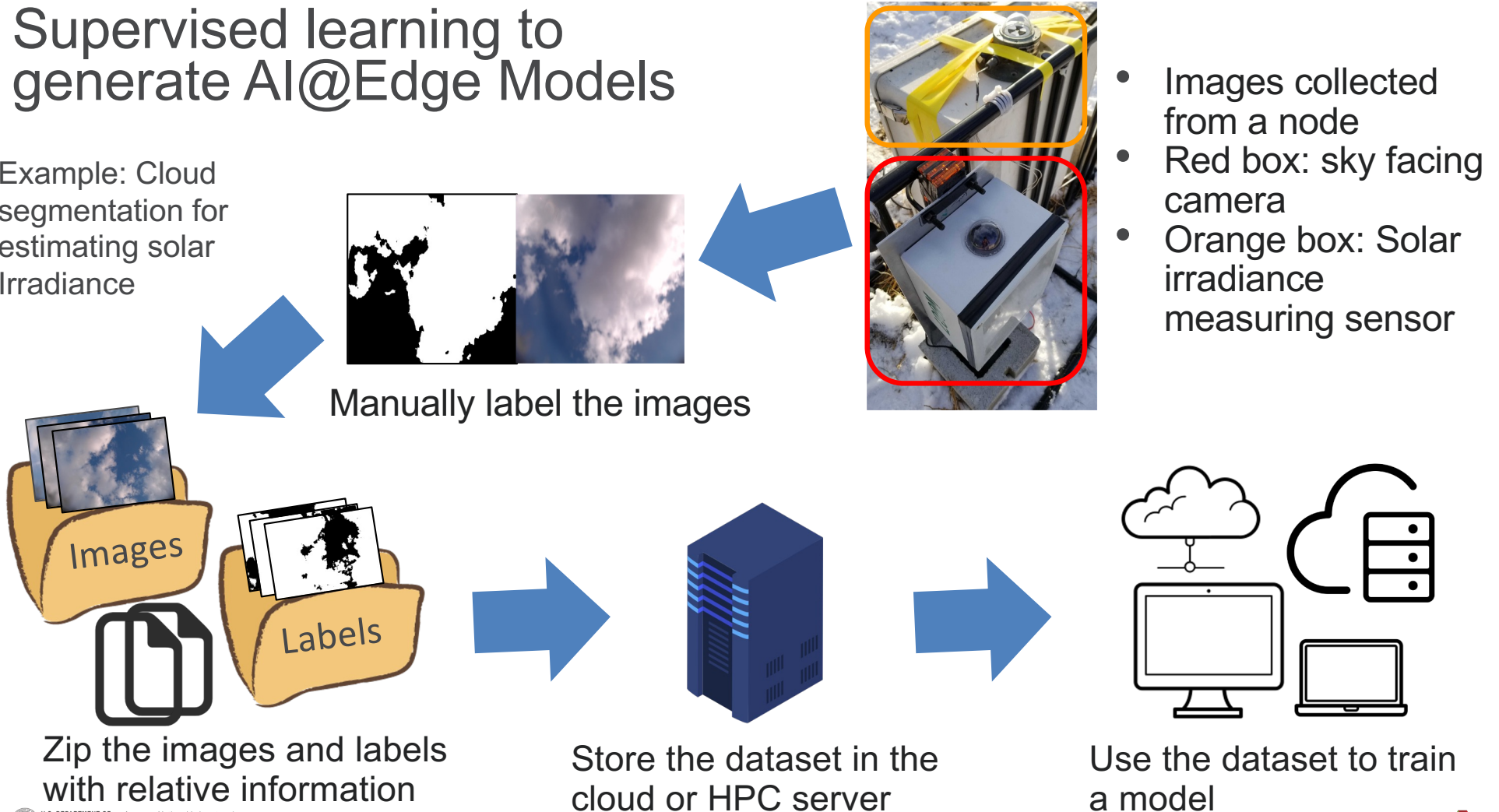


Model	mIoU	mAP	mAR
PLS	0.6467	0.8961	0.6991
FCN	0.5649	0.8974	0.6040
U-Net	0.7626	0.9869	0.7703
DeepLab	0.5335	0.9234	0.5582
AdaBoost (class)	0.6128	0.8494	0.6875
AdaBoost (norm)	0.5856	0.8646	0.6448

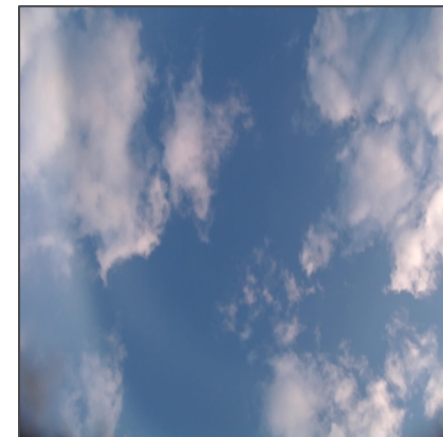
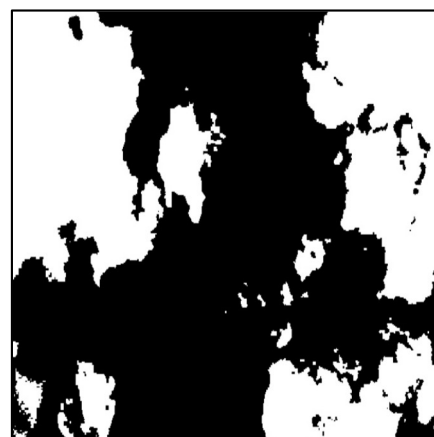
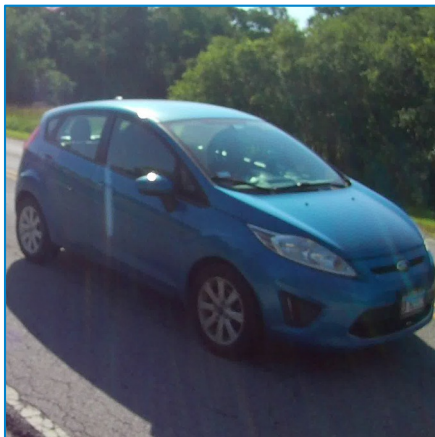
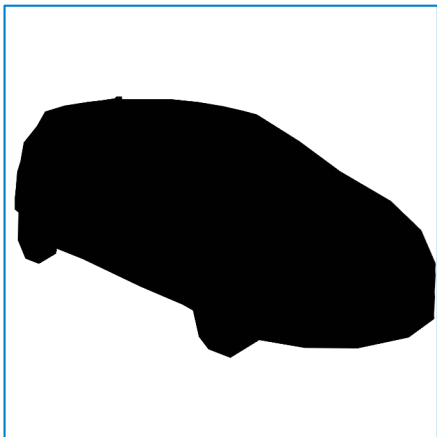


Supervised learning to generate AI@Edge Models

Example: Cloud segmentation for estimating solar irradiance



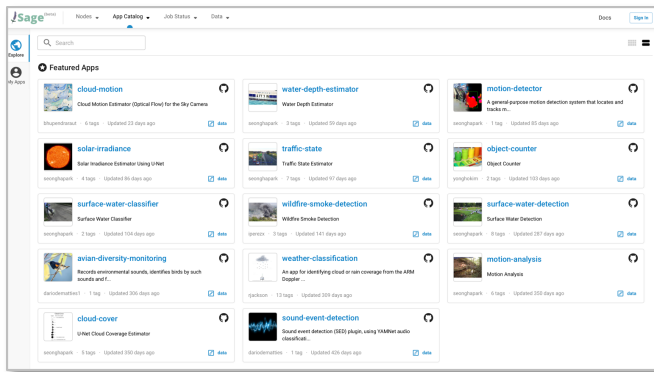
Labeling Examples



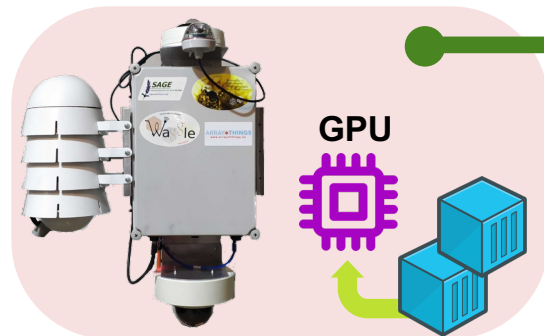
Need large labeled data sets for science including:

- Flooding detection
- Snow depth / coverage
- Smoke and Fire detection
- Cloud coverage
- Rime ice detection
- Urban heat island detection
- Predicting photovoltaics performance
- Seismic activity

AI SCALING CHALLENGE: BESPOKE AI CODE IS DIFFICULT!



Expert Mode:
Purpose-built AI



AI-Enabled Instruments

AI Analyzed Data

Edge Code Repository



Data

Model Tuning &
Data Analysis



End-to-end AI
Workflows

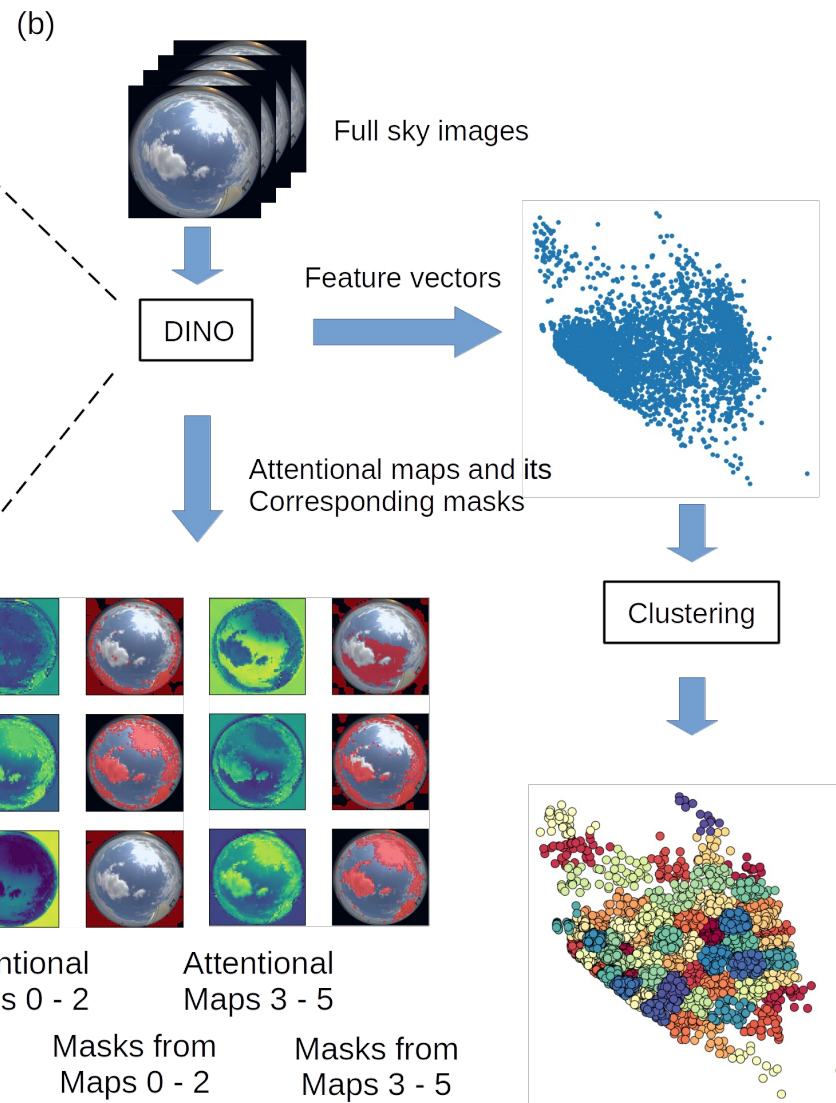
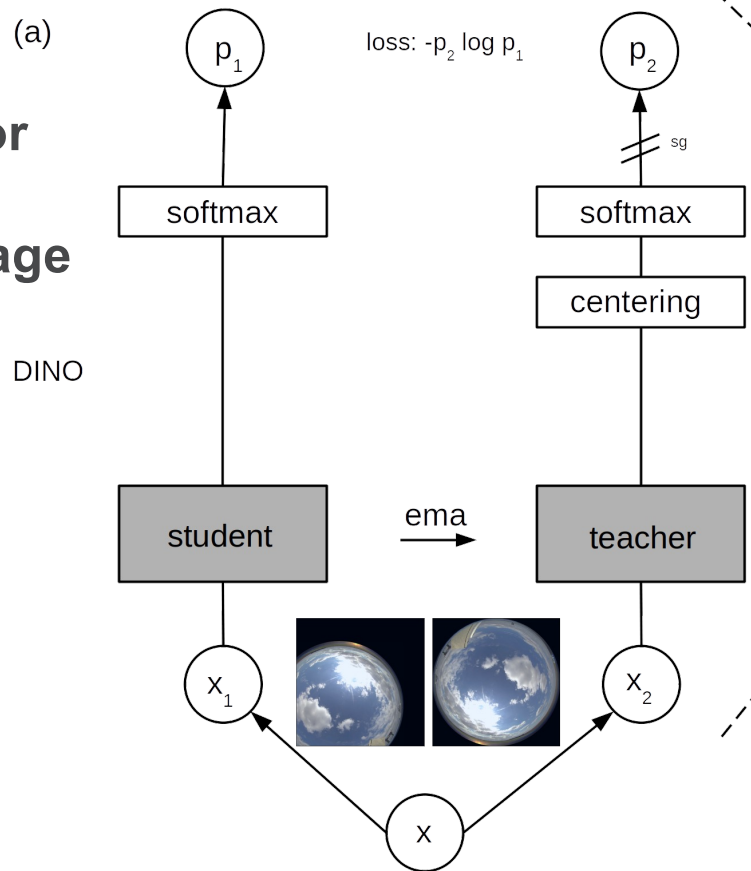
Sage Cloud

HPC/Cloud

Previous examples used *Supervised Learning*, requiring large annotated data-sets.

Self-supervised learning methods are well suited for Edge computing.

SSL for cloud coverage

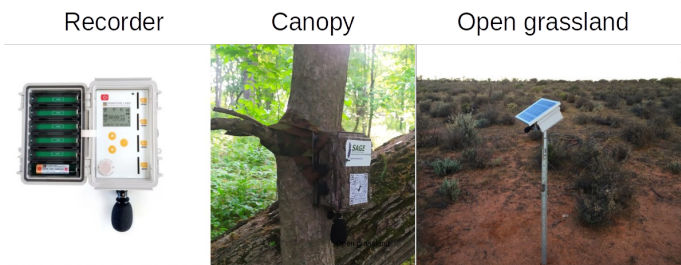
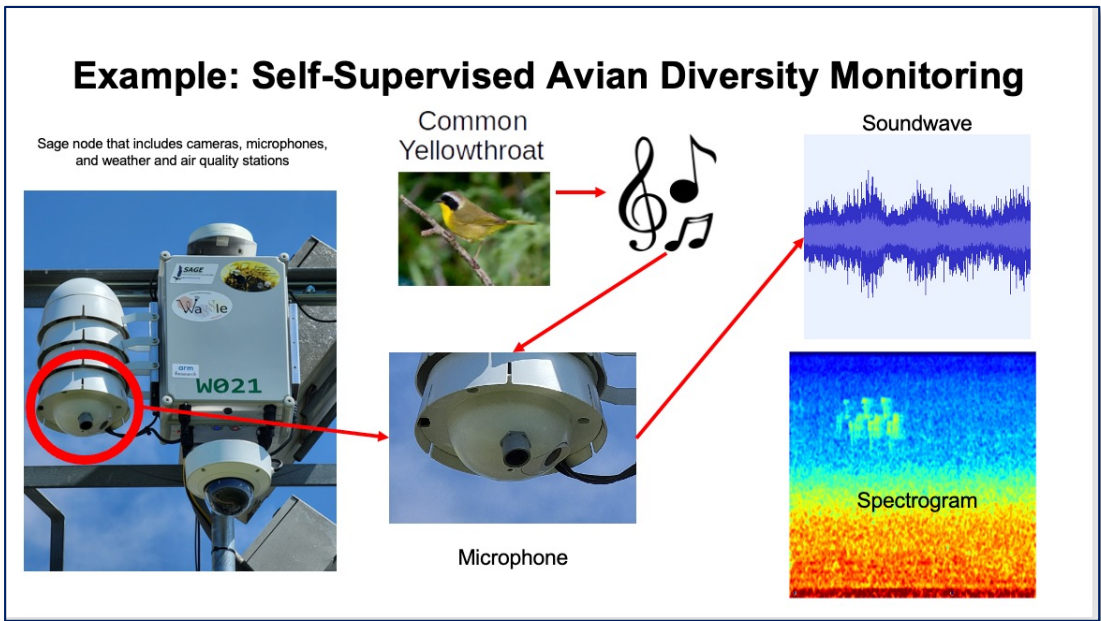
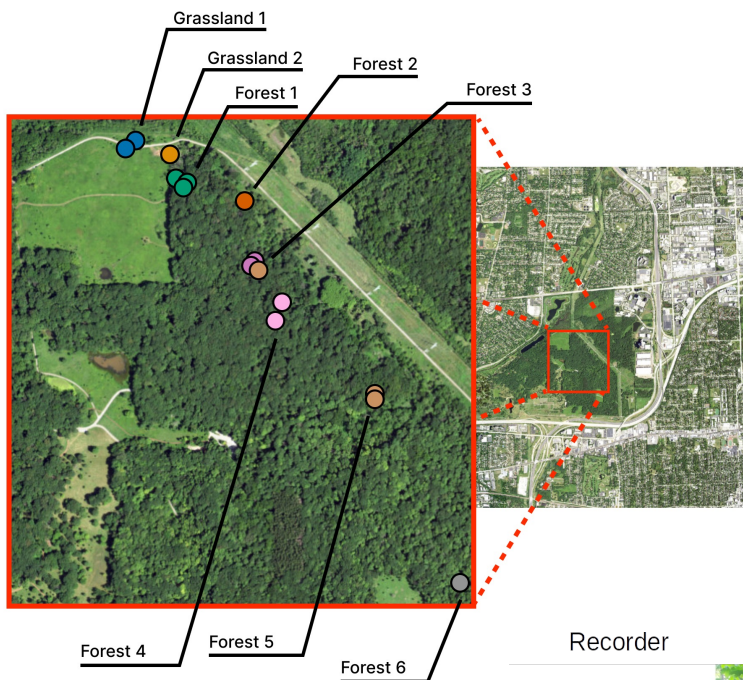


Dematties, D., Raut, B.A., Park, S., Jackson, R.C., Shahkarami, S., Kim, Y., Sankaran, R., Beckman, P., Collis, S.M. and Ferrier, N., 2023. Let's Unleash the Network Judgment: A Self-Supervised Approach for Cloud Image Analysis. *Artificial Intelligence for the Earth Systems*, 2(2), p.220063.

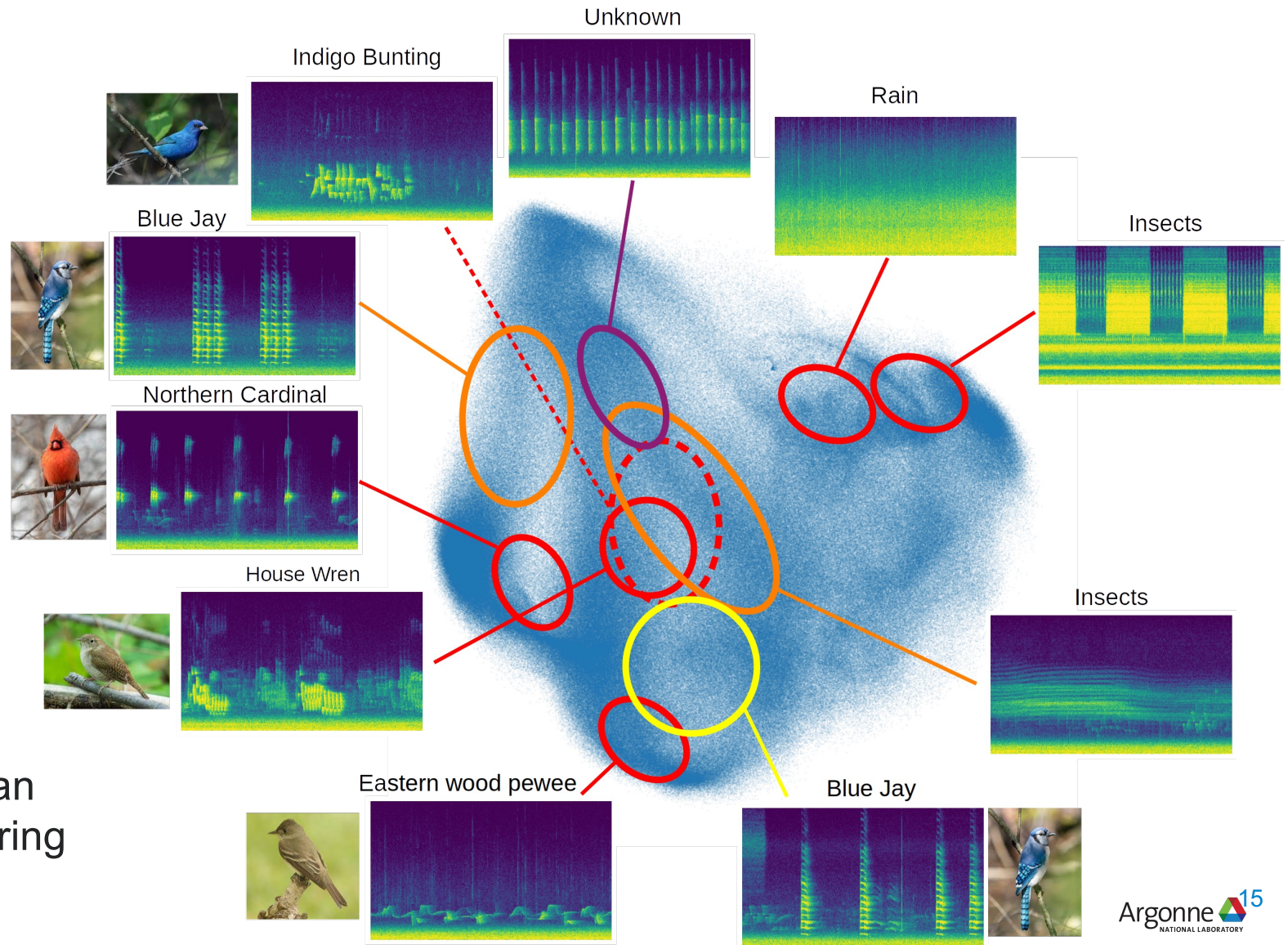
The Morton Arboretum, Lisle, IL

Device ID

- 4862
- 4879
- 23734
- 23764
- 23795
- 23757
- 23771
- 23700



Self-Supervised Learning for Avian Diversity Monitoring



The Future? A (multi-modal) LLM at the Edge to solve hard problems using only text prompts

Model RAM Usage 8.00 GB

ge 8.00 GB
0.00 %

PsiPi · llava v1 5 13B Q2_K gguf Vision

Eject Model

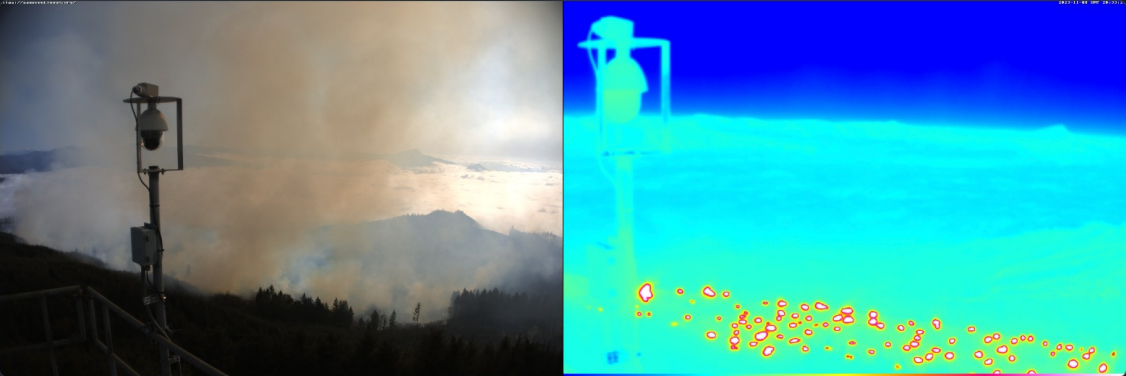
Export

Model Info Architecture Llama 13B Q2_K

Plaintext Markdd

Please describe in detail, what is happening in this image

USER

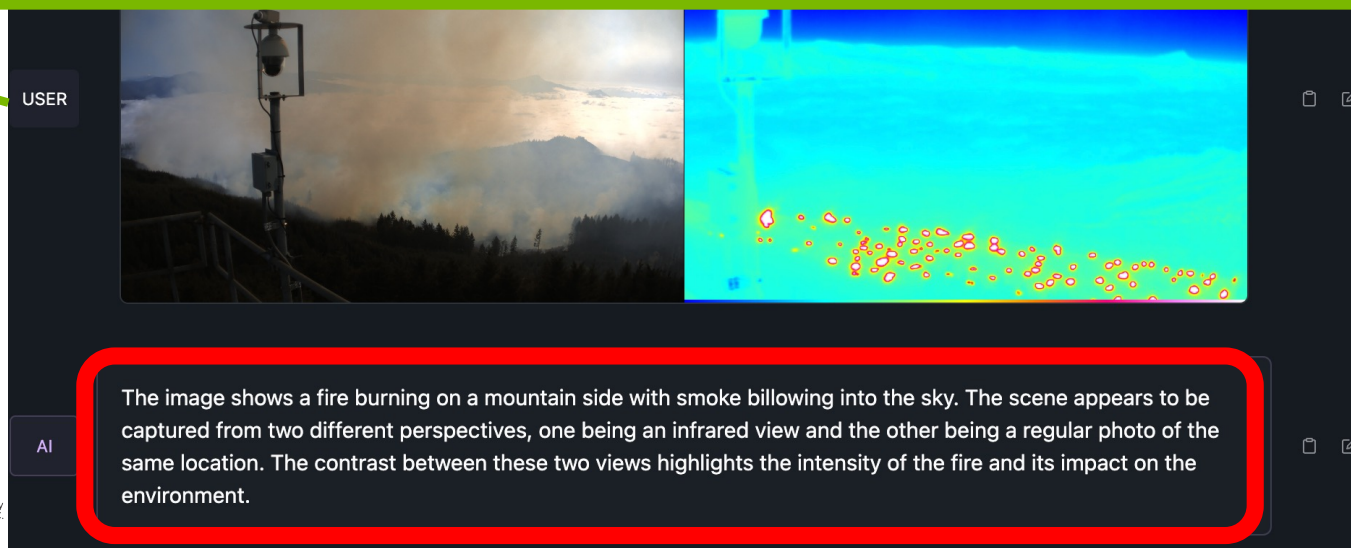


AI

The image shows a fire burning on a mountain side with smoke billowing into the sky. The scene appears to be captured from two different perspectives, one being an infrared view and the other being a regular photo of the same location. The contrast between these two views highlights the intensity of the fire and its impact on the environment.

The Future? A (multi-modal) LLM at the Edge to solve hard problems using only text prompts

The image shows a fire burning on a mountain side with smoke billowing into the sky. The scene appears to be captured from two different perspectives, one being an infrared view and the other being a regular photo of the same location. The contrast between these two views highlights the intensity of the fire and its impact on the environment.



AI provides a powerful tool to analyze data streams. Moving the compute to the edge can reduce bandwidth and provide a measurement of the “important information”. We need approaches that make it easier for a scientist to “answer their question” without extensive CS background (and/or tedious labeling of large datasets).

QUESTIONS/DISCUSSION?