Next Class (Oct 22): Checkpoint-2

no class on Oct 15 (Fall Break)

online score sheet

What we are trying to figure out:

- **Have you made substantial progress?**
- **Do you have any working code yet?**
- **Have you run into any difficulties?**
- **Do you know how you will resolve them?**
- **Are you clear on steps to completion?**
- **Have you articulated the final demo?**
- ***Is it time to panic yet?*** ☺

~30 minutes per project

Encourage mentors to attend

Sensing and Actuation

15-821 / 18-843
Fall 2024

Mahadev Satyanarayanan
School of Computer Science
Carnegie Mellon University

Importance of Topic

Link between physical and cyber worlds sensing = physical → cyber actuation = cyber → physical

Sensing is critical to inferring *user context*

- essential for successful user interaction
- especially important when user attention is scarce (e.g. when mobile)

***New applications & efficiencies* enabled by sensing**

supply chain management, inventory control, traffic control, preventive maintenance, …

Actuation enables impact on physical world

taking a picture, dimming lights in room, changing temperature, moving a robotic arm, leg, wheel, …

Autonomous robots have been driver for 30+ years

- energy issues dominated by physical movement of robot (actuation)
  energy cost of sensing minor by comparison
- energy cost of sensing is a dominant concern
  actuation is much less frequent, and often indirect (e.g. dimming room lights)

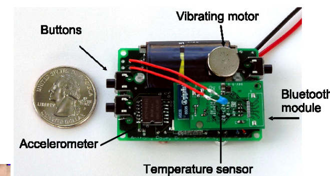
Approaches to Sensing

Sensing

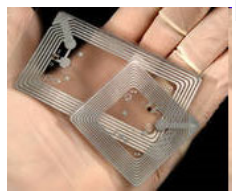
Embedded

- **Mote-based ad hoc wireless networks ("Smart Dust")**

- **Body-worn Sensors**



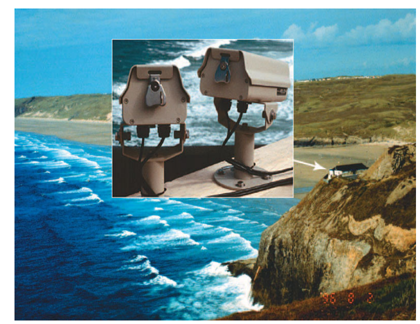
- **RFID Tags**



Remote

Typically based on computer vision

- **satellite/aerial imaging**
- **widely used in robotics**



Planned or opportunistic

- **Deployed camera networks**
- **Crowd-sourced imaging**

Tradeoffs

Embedded	Remote
+ Broader range of attributes can be sensed e.g. temperature, pressure, toxicity, …	+ Surprising versatility e.g. spectroscopy detected He in stars first
– Requires physical attachment much more intrusive	+ Non-invasive does not require cooperation of subject
	– Computationally intensive
– Battery charging is major consideration both for computation/communication (except for RFID)	Energy issues less onerous
+ Less ambiguity, easier to interpret	– Greater ambiguity, harder to interpret multiple viewpoints helpful

Readings for Today

[Efstratiou2007]	Efstratiou, C., Davies, N., Kortuem, G., Finney, J., Hooper, R., Lowton, M. ***Experiences of Designing and Deploying Intelligent Sensor Nodes to Monitor Hand-Arm Vibrations in the Field***
[Chebrolu2008]	Chebrolu, K., Raman, B. Mishra, N., Valiveti, P., Kumar, R. ***BriMon: A Sensor Network System for Railway Bridge Monitoring***
[Simoens2013]	Simoens, P., Xiao, Y., Pillai, P., Chen, Z., Ha, K., Satyanarayanan, M. ***Scalable Crowd-Sourcing of Video from Mobile Devices***
[Ha2014]	Ha,K., Chen, Z., Hu, W., Richter, W., Pillai, P., Satyanarayanan, M. ***Towards Wearable Cognitive Assistance***
[Wang2017]	Wang, J., Amos, B., Das, A., Pillai, P., Sadeh, N., Satyanarayanan, M. ***A Scalable and Privacy-Aware IoT Service for Live Video Analytics***
[George2023]	George, S., Turki, H., Feng, Z., Ramanan, D., Pillai, P., Satyanarayanan, M. ***Low-Bandwidth Self-Improving Transmission of Rare Training Data***

Mapping Papers to Techniques

Embedded Sensing

- **distributed sensor-based**
  - Chebrolu2008 (BriMon)
- **body-worn sensor-based**
  - Efstratiou2007 (Nemo)
  - Ha2014 (Gabriel)

Remote Sensing

- **deployed sensors**
  - Wang2017 (RTFace)
  - George2023 (Hawk)
- **crowd-sourced sensing**
  - Simoens2013 (Cloudlet-based)

The Role of Machine Learning

Sensors are imperfect

- **noisy** (random fluctuations)
- **drift** (consistent error that varies slowly over time)
- **affected by external factors** (imperfect control)
- **sensing may itself perturb value being sensed**
- …

Complex model maps sensor readings to true system state

- **making correct inferences is hard**
- **body of machine learning techniques have evolved to cope**
- **most sensor-related papers include some use of machine learning**

"Machine Learning: A Crucial Tool for Sensor Design"
Zhao, W., Bhushan, A., Santaaria, A.D., Simon, M. G., Davis, C. E.
Algorithms, 2008 1(2):130-152

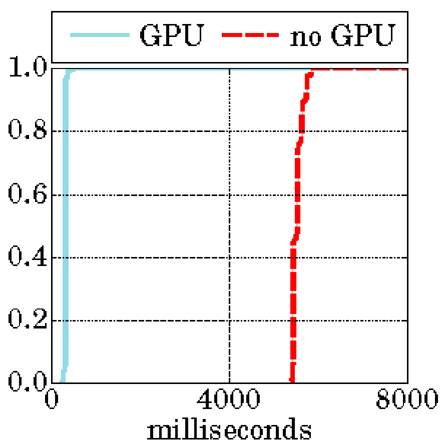
The Role of Offloading

Machine learning has two components:

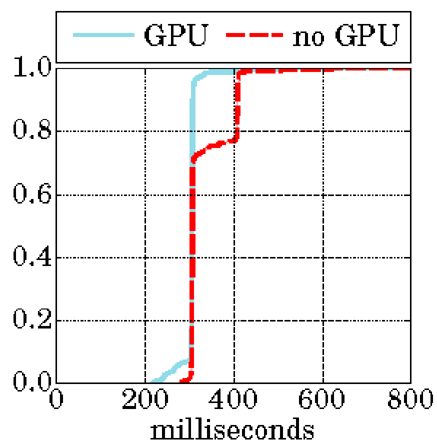
- *training* is typically done offline in the cloud
- *inferencing* has to be done as part of sensing

In early types of ML (e.g. SVM) inferencing was relatively cheap

With *deep neural networks (DNNs)* this is no longer true



(a) Sandwich



(b) Face

from Chen et al 2017, "An Empirical Study of Latency …"

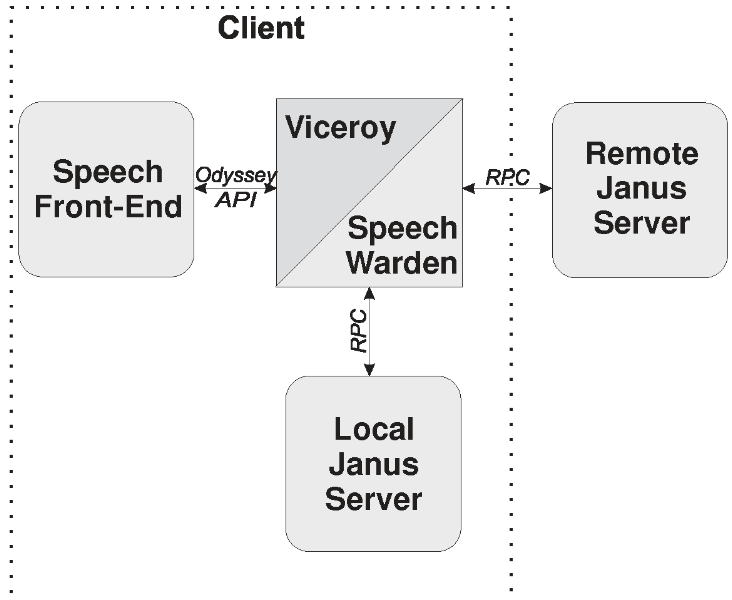
Tier-3 may not perform inferencing fast enough

Solution: *wireless offload to powerful remote server*

not encumbered by weight, size, thermals, battery life, etc.
aka "cyber foraging" (~2001-2014)

First demonstrated in 1997 at CMU!

- **See "How we created edge computing"**
  (Nature Electronics, 2(1), January 2019)
- **See "A Brief History of Cloud Offload"**
  (GetMobile, Vol. 18, #4, October 2014)



October 1997 (10 years before iPhone and Siri)
"Agile-Application Aware Adaptation for Mobility"
Noble, B., Satyanarayanan, M., Narayanan, D., Tilton, E., Flinn, J., Walker, K.
Proceedings of the 16th ACM Symposium on Operating Systems Principles, St. Malo, France, October 1997

Where should remote server be located?
- **"nearby" makes a huge difference!**

Does Latency Really Matter?

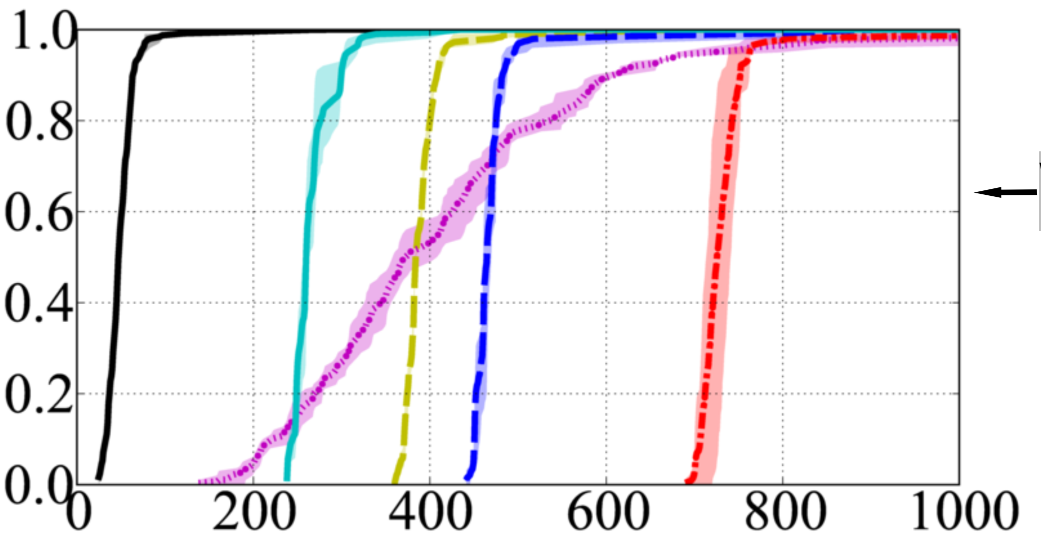
"The Impact of Mobile Multimedia Applications on Data Center Consolidation"
Ha, K., Pillai, P., Lewis, G., Simanta, S., Clinch, S., Davies, N., Satyanarayanan, M.
Proceedings of IEEE International Conference on Cloud Engineering (IC2E), San Francisco, CA, March 2013

"Quantifying the Impact of Edge Computing on Mobile Applications"
Hu, W., Gao, Y., Ha, K., Wang, J., Amos, B., Pillai, P., Satyanarayanan, M.
Proceedings of ACM APSys 2016, Hong Kong, China, August 2016

Augmented Reality

E2E Response Time CDF

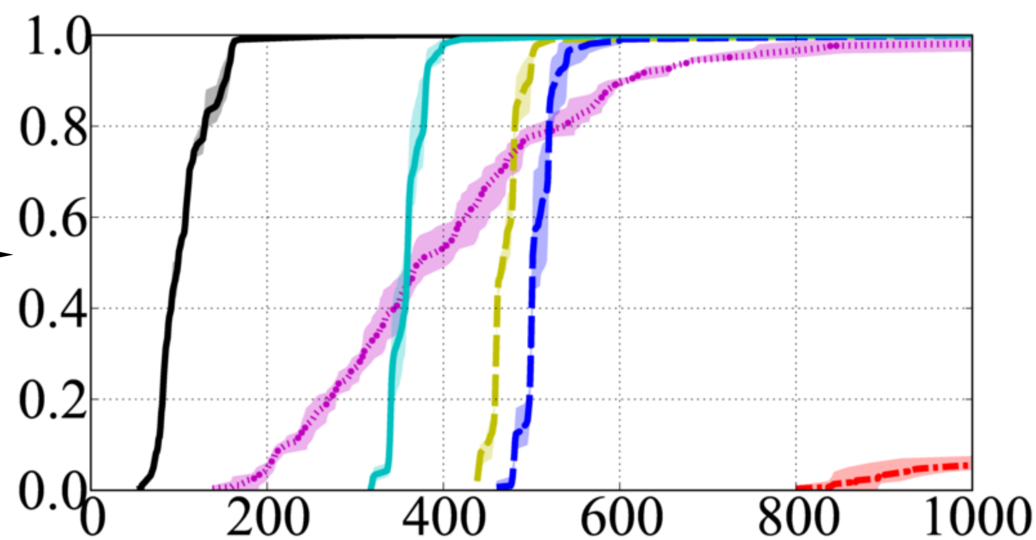
1. Send JPG image from device to cloud/cloudlet
2. Recognize landmark buildings using computer vision
3. Send labels & coordinates back to device



Wi-Fi
802.11n

X axis values are milliseconds

4G LTE
T-Mobile for Cloud
In-lab Nokia eNodeB for Cloudlet



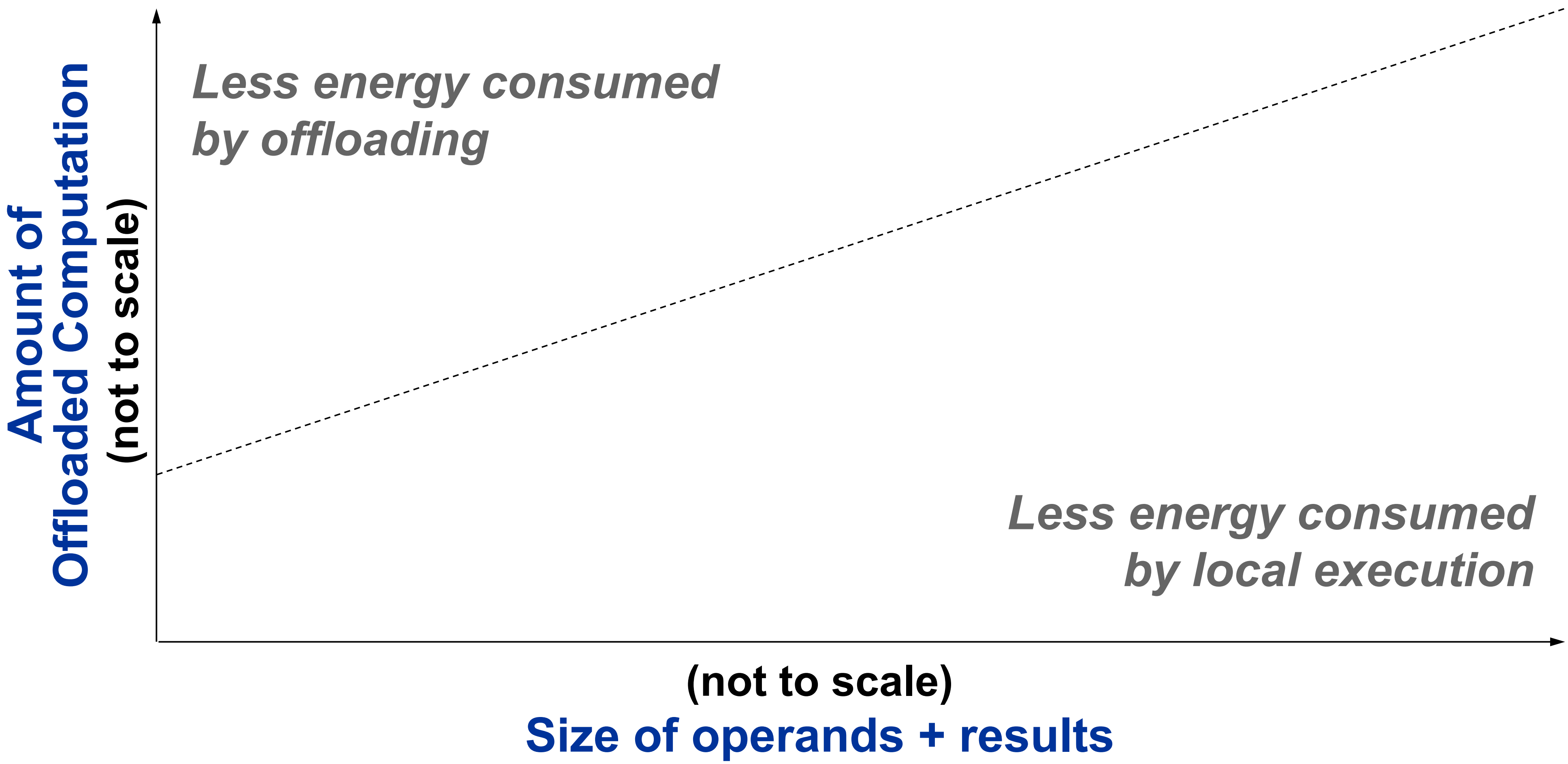
We have confirmed same trend for many other applications

Per-Operation Energy Use by Device

Face Recognition		Augmented Reality
12.4 J	Mobile-only	5.4 J
2.6 J	Cloudlet	0.6 J
4.4 J	Amazon East	3.0 J
6.1 J	Amazon West	4.3 J
9.2 J	Amazon EU	5.1 J
9.2 J	Amazon Asia	7.9 J

We have confirmed same trend for many other applications

Extending Battery Life



Observations on Offloading

1. *Network Proximity is important*
   - **weakly correlated with physical proximity**
     200 km per millisecond in fiber at speed of light
   - **physical proximity is neither sufficient nor necessary for network proximity**
2. *"Motion-to-Photon Latency" (MTPL) is key*
   - **complex end-to-end metric spanning compute and transmission**
   - **correlated with network latency and bandwidth**
   - **but also correlated with CPU speed/load, accelerators (e.g. GPUs), cache state, …**
3. *Non-technical considerations matter*
   - **business considerations**
   - **societal priorities** (e.g. moral equivalents of handicapped parking)
   - **legal constraints** (e.g. GDPR privacy constraints on data placement)

Is There a Killer Use Case?

"Towards Wearable Cognitive Assistance"
Ha, K., Chen, Z., Hu, W., Richter, W., Pillai, P., Satyanarayanan, M.
Proceedings of the Twelfth International Conference on Mobile Systems, Applications, and Services (MobiSys 2014), Bretton Woods, NH, June 2014

"Early Implementation Experience with Wearable Cognitive Assistance Applications"
Chen, Z., Jiang, L., Hu, W., Ha, K., Amos, B., Pillai, P., Hauptmann, A., Satyanarayanan, M.
Proceedings of WearSys 2015, Florence, Italy, May 2015

"An Empirical Study of Latency in an Emerging Class of Edge Computing Applications for Wearable Cognitive Assistance"
Chen, Z., Hu, W., Wang, J., Zhao, S., Amos, B., Wu, G., Ha, K., Elgazzar, K., Pillai, P., Klatzky, R., Siewiorek, D., Satyanarayanan, M.
Proceedings of SEC 2017, San Jose, CA, October 2017

Wearable Cognitive Assistance

"Look and feel of AR, with functionality of AI"

"An Angel on Your Shoulder"
Project Gabriel
http://gabriel.cs.cmu.edu

Wearable UI with wireless offload to cloudlet

Real-time cognitive engines **on cloudlet (microservices)**

- scene analysis
- object/person recognition
- speech recognition
- language translation
- planning, navigation
- question-answering technology
- voice synthesis
- real-time machine learning
- …

Low latency response is crucial

Human Cognition is Amazing

Fast, accurate and robust

• face detection under hostile conditions (low lighting, distorted optics)	< 700 ms
• face recognition	370 ms – 620 ms
• is this sound from a human?	4 ms
• VR head tracking	< 16 ms

To be "superhuman" we need to beat these speeds

Leave time for additional software processing (e.g. database lookup) to add value to user

Task-specific Assistance

Example: cooking

passive recipe display

versus active guidance

“Wait, the oil is not hot enough”

Inspiration: GPS Navigation Systems

Turn by turn guidance

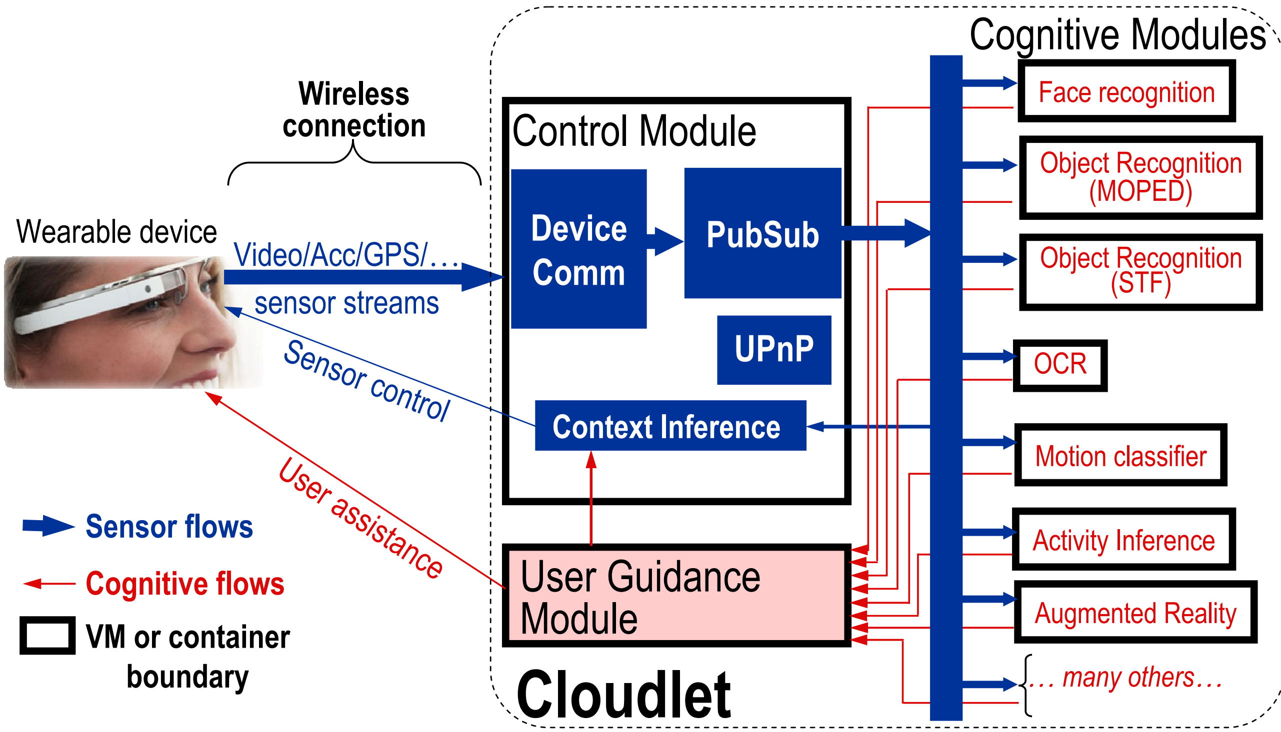
- **Ability to detect and recover**
- **Minimally distracting to user**

Uses only one type of sensor: location from GPS

Can we generalize this metaphor?

Gabriel Architecture

(PaaS for Wearable Cognitive Assistance)



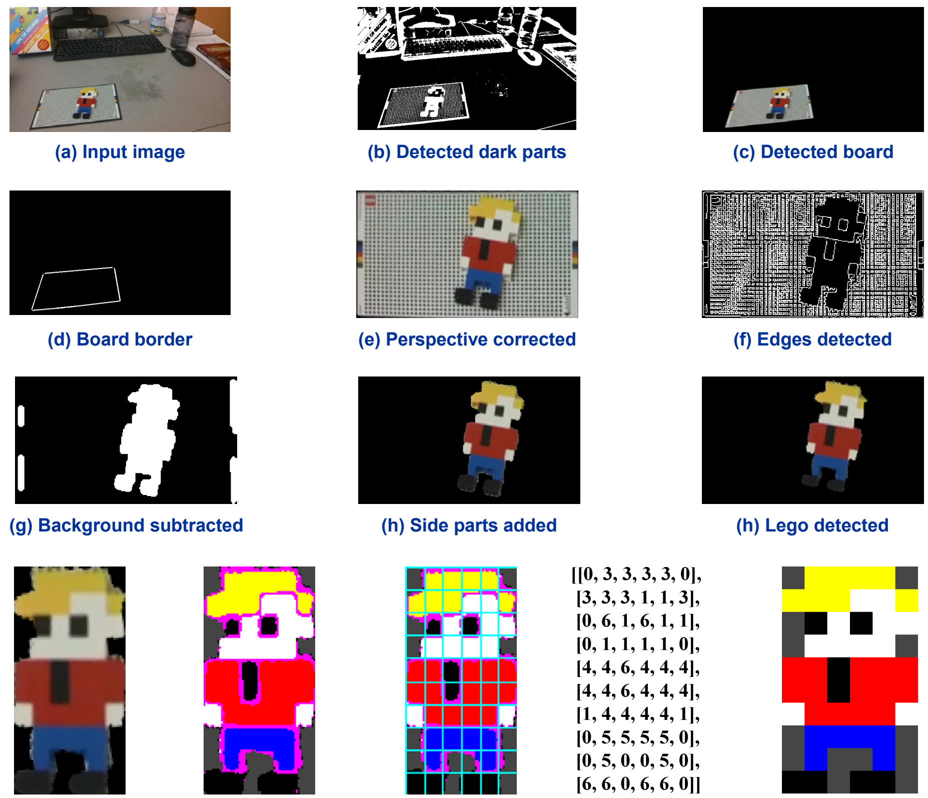
Baby Steps: 2D Lego Assembly

Very first proof-of-concept (September 2014)

Deliberately simplified task to keep computer vision tractable

2D Lego Assembly (YouTube video at http://youtu.be/uy17Hz5xvmY)

On Each Video Frame



(a) Input image
(b) Detected dark parts
(c) Detected board
(d) Board border
(e) Perspective corrected
(f) Edges detected
(g) Background subtracted
(h) Side parts added
(h) Lego detected
(i) Unrotated
(i) Color quantized
(j) Partitioned
(j) Matrix
(k) Synthesized

When Milliseconds Matter

Ping-pong assistant
(https://www.youtube.com/watch?v=_Ip32sowyUA)

Assembling an IKEA Kit

IKEA kit assistant

(https://www.youtube.com/watch?v=qDPuvBWNIUs&index=5&list=PLmrZVvFtthdP3fwHPy_4d61oDvQY_RBgS)

Many Use Cases …

Assembly instructions

Industrial troubleshooting

Medical training

Self-Instrumentation

Strong willpower

Escalation to Human Expert

Perfecting software is a long and arduous process

How hard do we have to work on WCA applications?

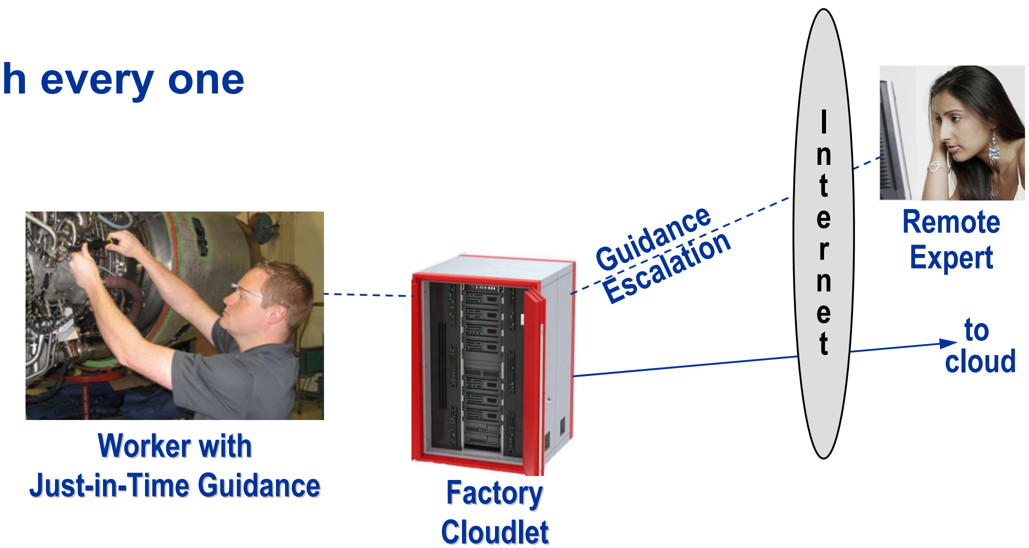
- **Many common errors by humans in a specific task**
- **But also rare errors, not often repeated**
- **Expensive to implement software to catch every one**

Solution: ***Escalate to a human via Zoom***
(exception handling)

Example: factory setting

WCA enables many workers per expert

Contrast with Microsoft 365 Dynamic Remote Assist **(one expert per worker)**



Low-Bandwidth Self-Improving Transmission of Rare Training Data

Shilpa George, Haithem Turki, Ziqiang Feng, Deva Ramanan,

Padmanabhan Pillai[†], Mahadev Satyanarayanan

Carnegie Mellon University and [†]Intel Labs

Many things in ML simplified if you already have a good training set

But what if you are trying to assemble that training set?

from data only found in remote and inaccessible places?

of a new, rare event?

Extreme Sensor to Backhaul Mismatch

4K Video Camera → 30 Mbps demand

future higher resolution, multispectral cameras will demand even higher bandwidths

Unmanned probes often have very poor wireless backhaul connectivity

- **deep space and inter-planetary networks** (10 – 100 kbps, 10^2 – 10^6 s one-way latency)
- **underwater acoustic networks** (10 – 100 kbps)
- **LoRa networks** (1 – 100 kbps)

Many exciting discoveries await us in these remote locations

A Perfect Storm

Convergence of three factors

- **Extreme mismatch of sensing vs transmission data rates**
  can't blindly ship all data
- **Rare unlabeled events**
  < 0.1% of frames, possibly much rarer, can't do random sampling
- **New phenomenon**
  no pre-built accurate detector/classifier, data is unlabeled

How to Retrieve Almost All Events Seen?
"Event" = True Positive (TP)

Our Solution: Live Learning

Interative human-in-the-loop workflow that combines

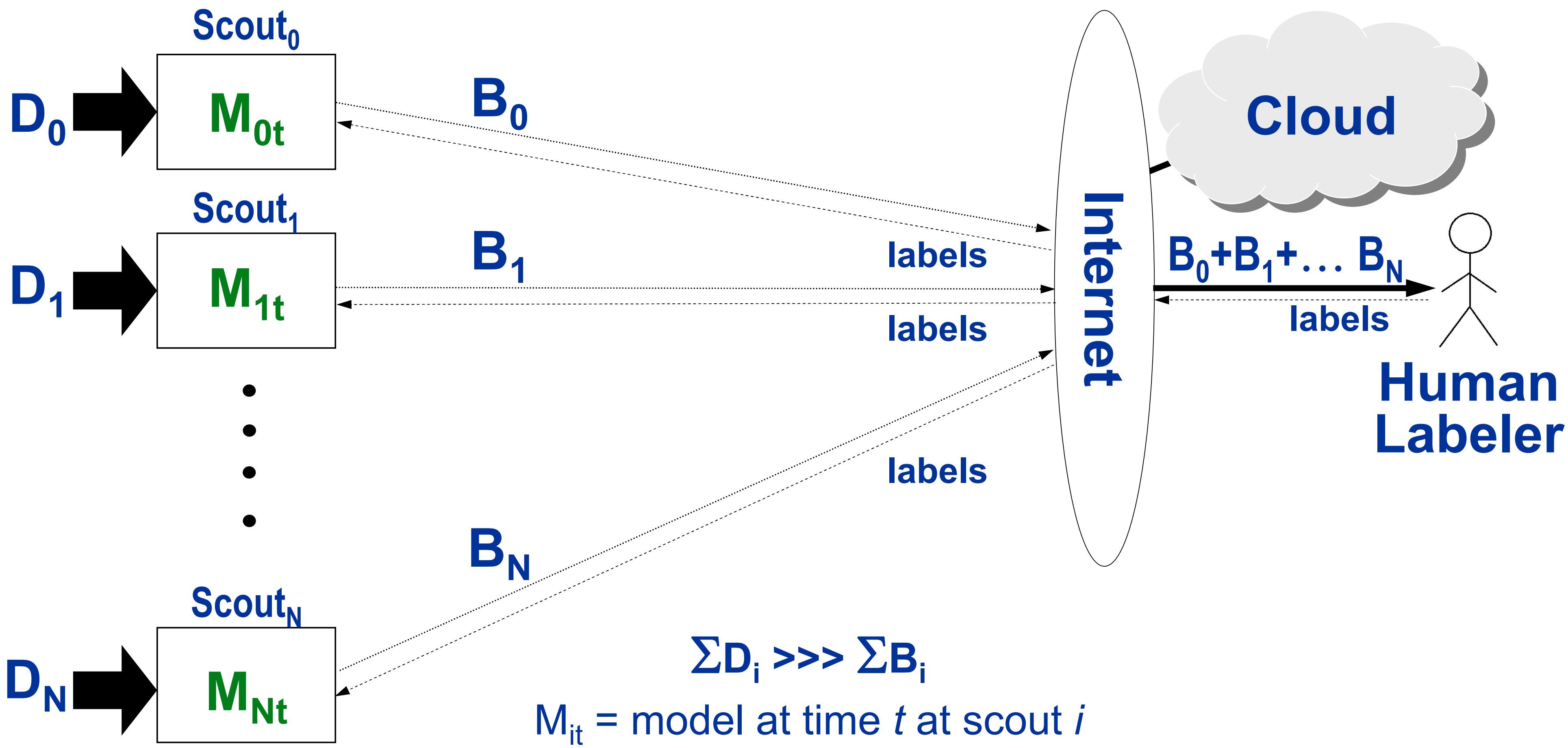
- **Semi-Supervised Learning (SSL)**
- **Active Learning**
- **Transfer Learning**

***Pipeline* sensing, inferencing, transmission, labeling, and training**

Key steps in pipeline

1. **Bootstrap with weak initial model** (few-shot learning)
2. **Grow training set with newly-discovered TPs** (human confirmation for every TP)
3. **Train new model and replace current model asap** (cloud or edge training site, bandwidth-adaptive)
4. **Iterate steps 1–3 during mission**

Live Learning Overview



$\Sigma D_i >>> \Sigma B_i$

M_{it} = model at time *t* at scout *i*

Hawk: Open Source Implementation

https://github.com/cmusatyalab/hawk

Based on *ZeroMQ* delay-tolerant messaging

Completely model-agnostic (easy plugin of new DNN models)

Paper reports extensive investigations re detailed design choices

- Top-K vs MaxEnt selective transmission
- Hybrid SVM-DNN model evolution vs pure DNN evolution
- Revisit policy to collect missed positives
- Importance of tiling high-res data
- …

Experimental Evaluation

1. ***In spite of extreme low bandwidth, can scouts discover most TPs encountered?***
2. ***How close is Hawk to an ideal system?***
   - **Oracle (perfect precision and recall)**
   - **BruteForce (imperfect precision and recall)**
     - model with the same architecture but trained in advance on fully labeled incoming data.
     - grossly overfitted to the data that will be seen during the mission.
     - requires all incoming data to be seen in advance, and transmitted to the cloud for labeling and training.
     - may not have perfect precision and recall.
3. ***Can Hawk use additional bandwidth effectively?***
4. ***Is Hawk DNN-agnostic?***

… many more questions …

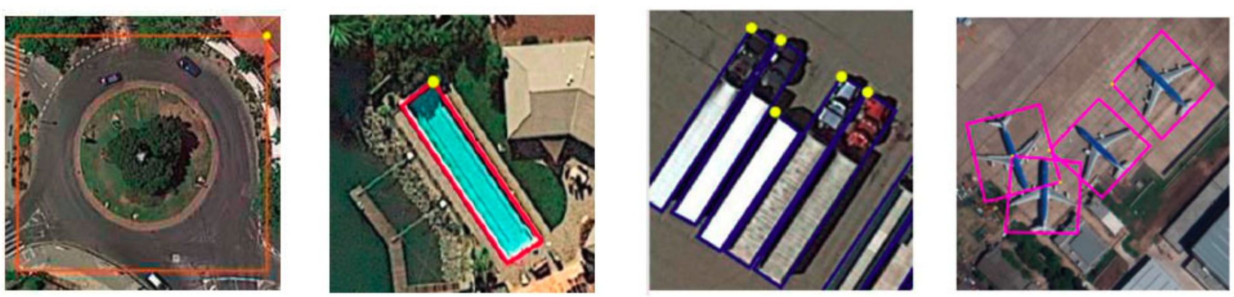
Dataset: Aerial Drone Surveillance

DOTA: Dataset for Object deTection in Aerial Images (published in 2018)

Consists of 2806 fully labeled images across 15 classes

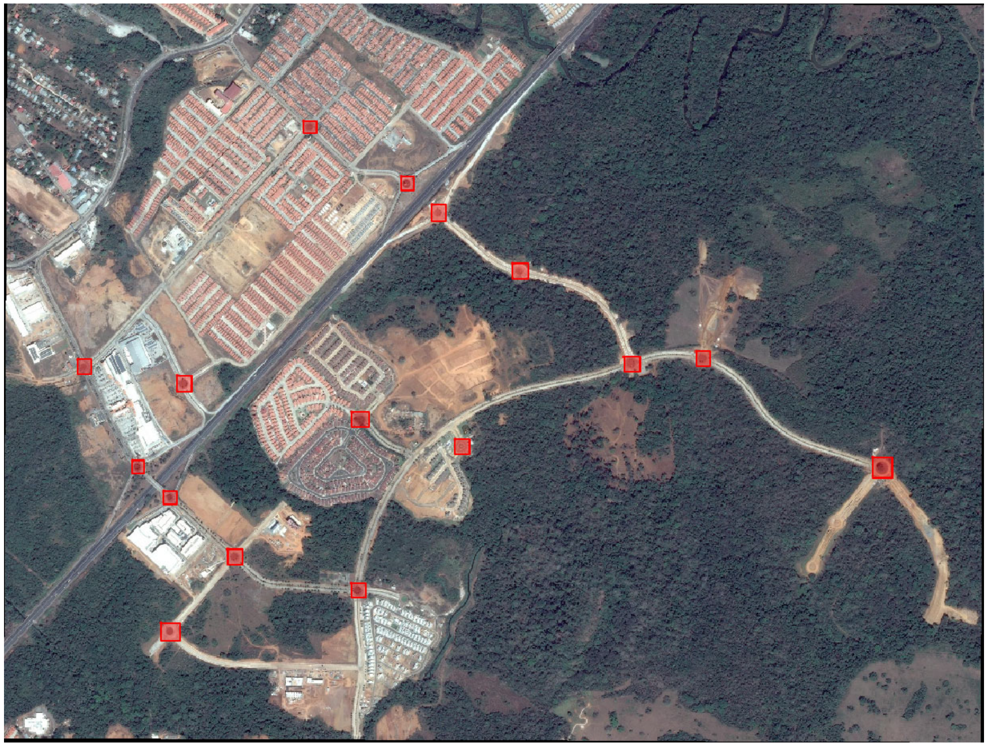
Image Resolution: Ranges from 800x800 to 4000x4000

Derived dataset has 252231 labeled tiles having base rate of 0.1%



(a) Roundabout (TPs=336) (b) Swimming Pool (TPs=335) (c) Large Vehicle (TPs=357) (d) Airplane (TPs=350)

256x256 tiles from large 4K images



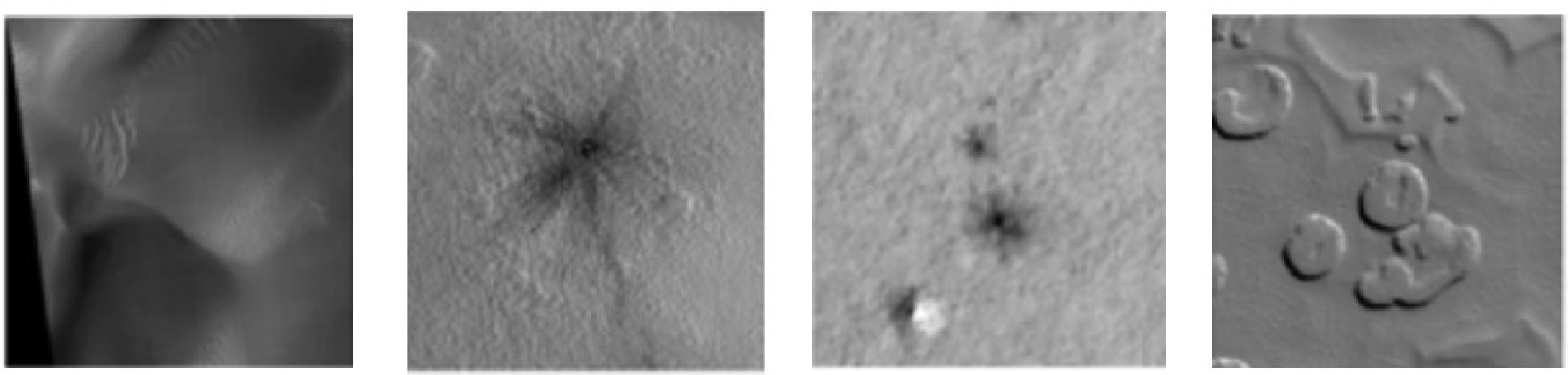
Dataset: Planetary Exploration

HiRISE: High Resolution Imaging Experiment from Mars (published in 2019)

Images collected by Mars Reconnaissance Orbiter

Dataset has 7 classes of landmarks on Martian terrain

Consists of 73,031 labeled images of size 227x227



(a) Dark Dune (TPs=64) (b) Impact Ejecta (TPs=64) (c) Spider (TPs=64) (d) Swiss Cheese (TPs=64)

Dataset: Underwater Sensing

Brackish: Marine dataset (published in 2019)

Images of marine animals in a brackish strait with varying visibility

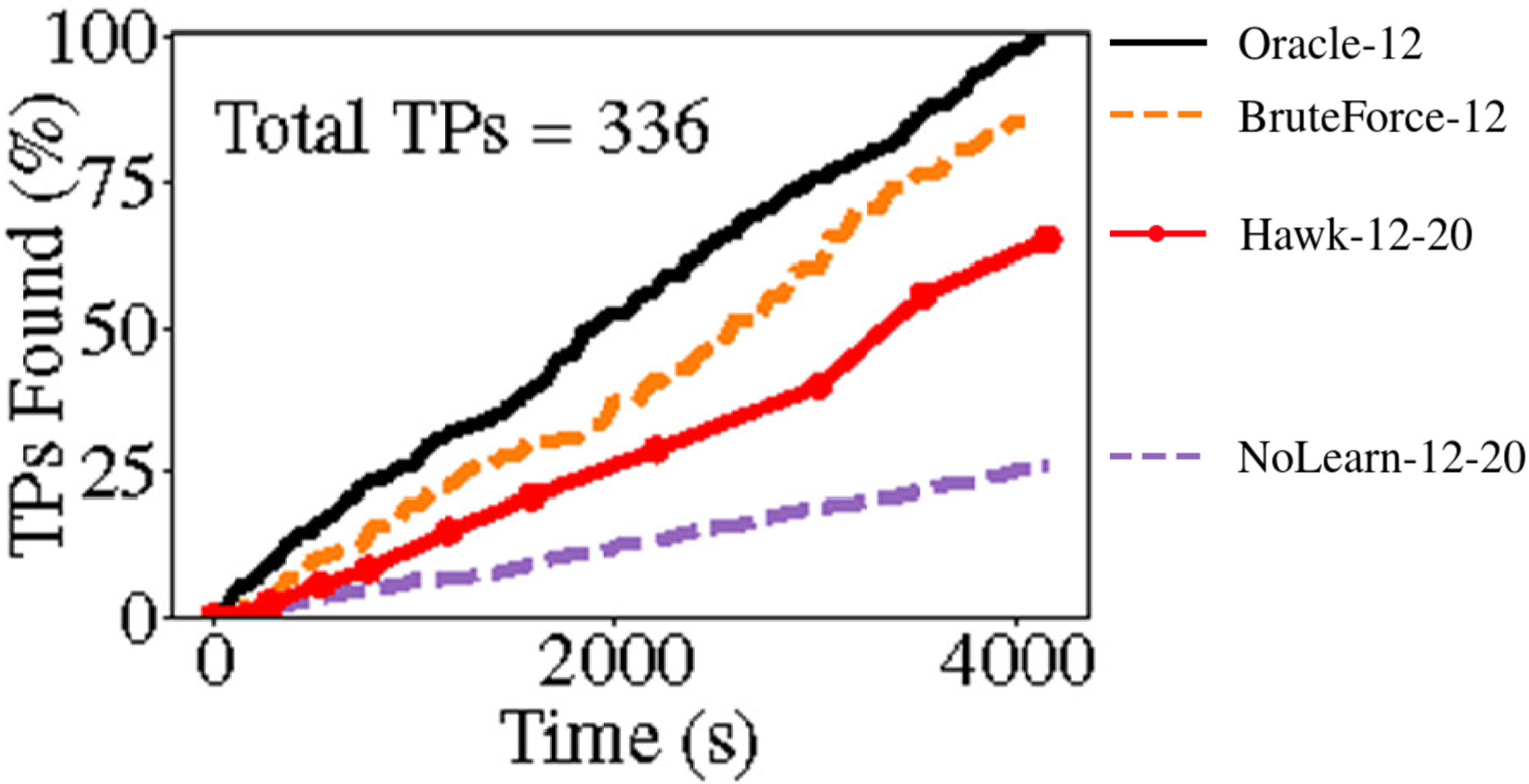
Consists of 14,518 labeled images of 1080p resolution

Derived dataset has 563,829 tiles across 6 classes with target baserate of 0.1%

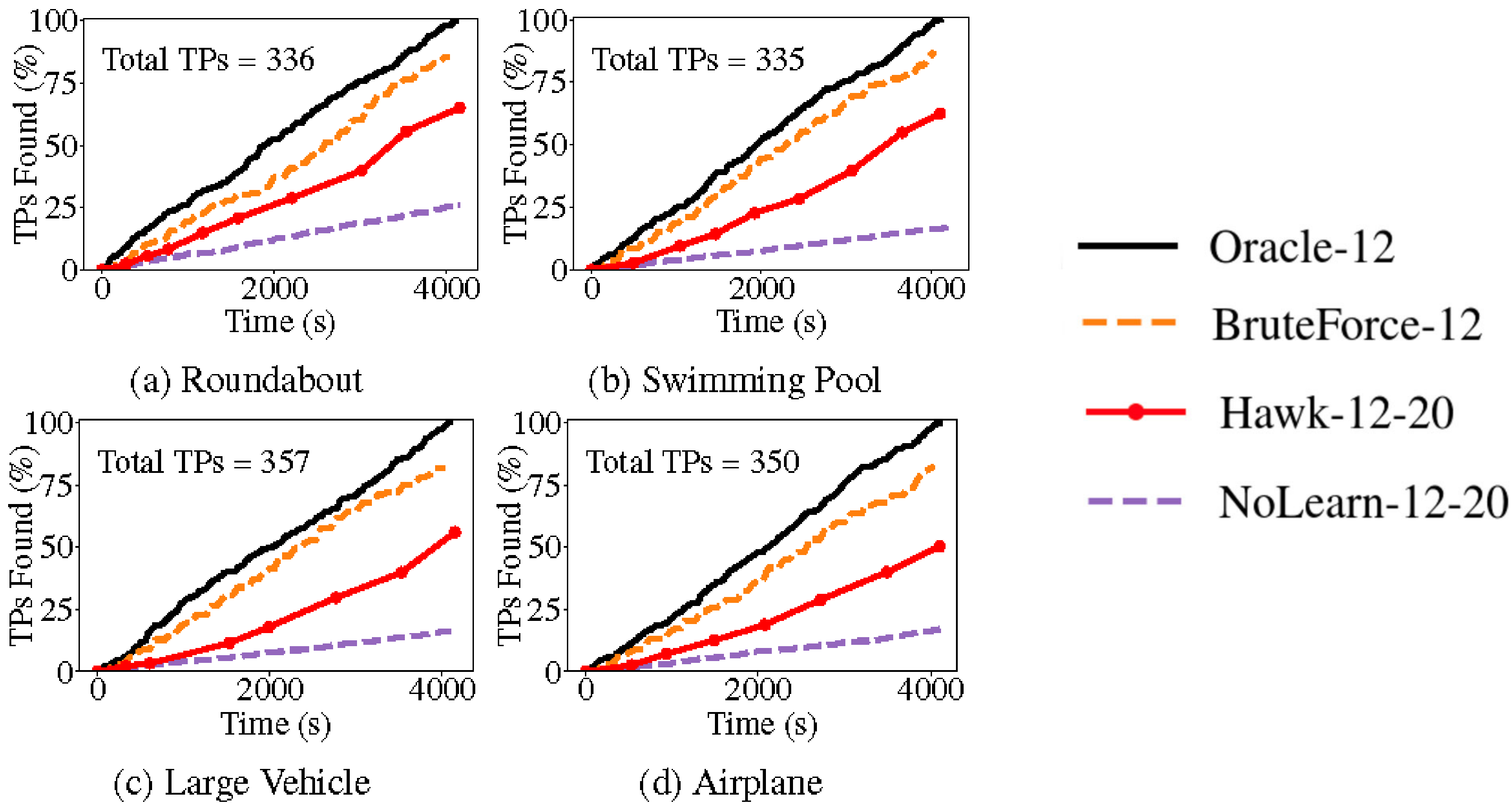
(a) Starfish (TPs=370) (b) Shrimp (TPs=564) (c) Small Fish (TPs=564) (d) JellyFish (TPs=584)

Scout-based Training - 12kbps (DOTA)

Class: Roundabout



Scout-based Training - 12kbps (DOTA)



See Paper for Many More Results

1. **Robust results across many datasets and classes** (aerial drone, Mars, underwater)
2. **Value of revisiting discard pile** (result caching of old scores)
3. **Live Learning is DNN agnostic** (ResNet-50, YOLOv4, ExtremeNet results)
4. **Ability to use higher bandwidth effectively** (12 kbps, 30 kbps, 100 kbps)
5. **Dynamic choice of cloud training versus scout training**
6. **Diversity Sampling to improve recall**
7. **Integration with Few-Shot Learning**

Take-Away Message

Gross bandwidth mismatch in remote sensing will grow worse

Live Learning is a viable solution to this problem

Key idea: *Integrate Learning with Selective Transmission & Human Labeling*

Hawk discovers up to 87% of the TPs discovered by BruteForce

Bonus: *Hawk also helps with limited human bandwidth*