School of
Information Systems

HDR-Nets,
October 13, 2020

Collaborative Edge-based Machine Intelligence: Promise and Challenges

Archan Misra

Acknowledge the creative contributions of:
PhD: Amit Sharma, Dulanga Weerakoon
Post-Docs: Tran Huy Vu, Kasthuri Jayarajah, Manoj Gulati, Meera Radhakrishnan
Post-doc & Engineers: Vengat Subramaniam, Dhanuja Wanniarachchige
Collaborators: Vigneshwaran Subbaraju, Tarek Abdelzaher, Rajesh Balan

My Research History

2010 - 2015

2010 — 2015

Mobile Sensing & Analytics

- Indoor Location
- Group Detection
- Queuing Detection

Key Research Thrusts

- Fusion of multi-modal sensing *(inertial)*
- Adaptive sampling & triggered sensing
- **Multiple live deployments** *(campus, malls, museums)* **+ licensing**

2014 - 2018

2014 — 2018

Wearable Sensing & Systems

- Eating *(Annapurna)*
- In-Store Shopping *(IRIS, I4S)*
- VR+ mobile *(Empath-D)*

Key Research Thrusts

- Optimize *(Energy, Accuracy, Latency)* tradeoffs
- Multi-modal sensor fusion *(inertial, image)*

2018-2022

2022

Wearable + IoT Systems

- Batteryless Wearables
- Wireless/RFID Sensing
- Fine-grained Gestural Tracking

Key Research Thrusts

- **Make Batteryless (or Utlra-Low Power) Sensing possible**
- **Method: Utilize new sensing modalities (*video, wireless*) & collaborative ML at edge**

W8-Scope: Exercise Monitoring using IoT Sensors

Percom 2020

Goals:

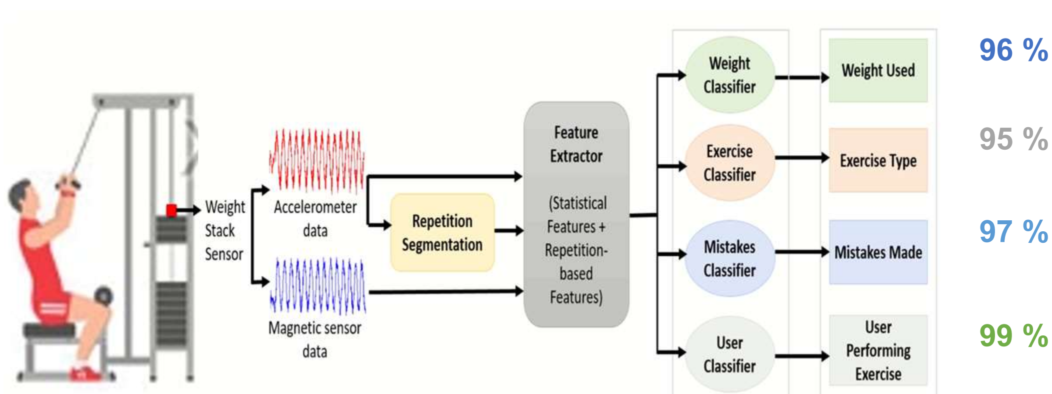
- Quantified insights on weight stack-based exercises→ provide personalized digital coaching

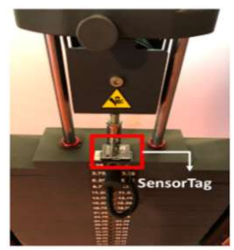
Techniques:

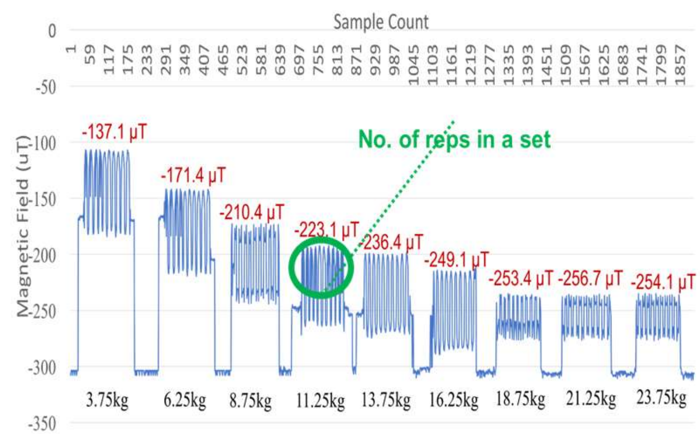
- Simple weight stack sensor *(accelerometer+ magnetometer)* to track & understand exercises

Results:

- Longitudinal Data Collection at 2 gyms ➨ 95+% accuracy & adaptation to medium-term evolutionary behavior







Magnetic Sensor on Wt. Stack→ {Weight, Type, User}

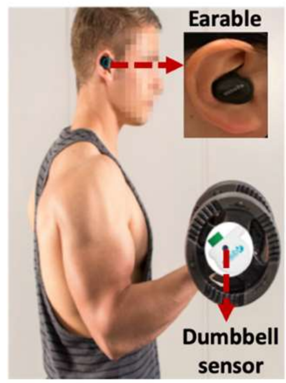
School of Information Systems

ERICA: Earable-based Real-Time Feedback for Free-weights Exercises

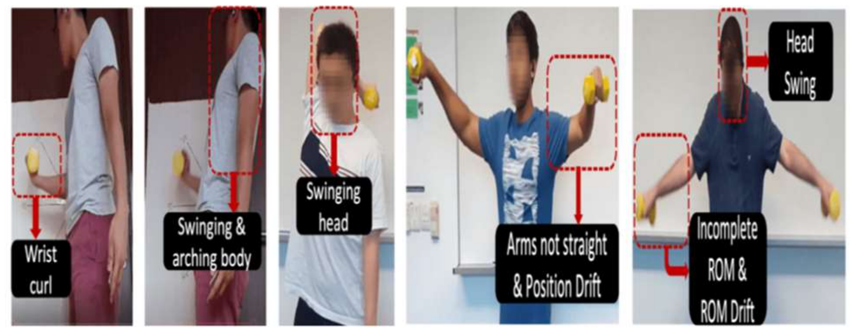
Sensys 2020

Goals:

- **Associate User's Earable with Dumbbell-mounted IoT sensors**
- **Perform exercise recognition & real-time mistake detection**
- **Provide "live" corrective feedback**



(b) Sensor Placement



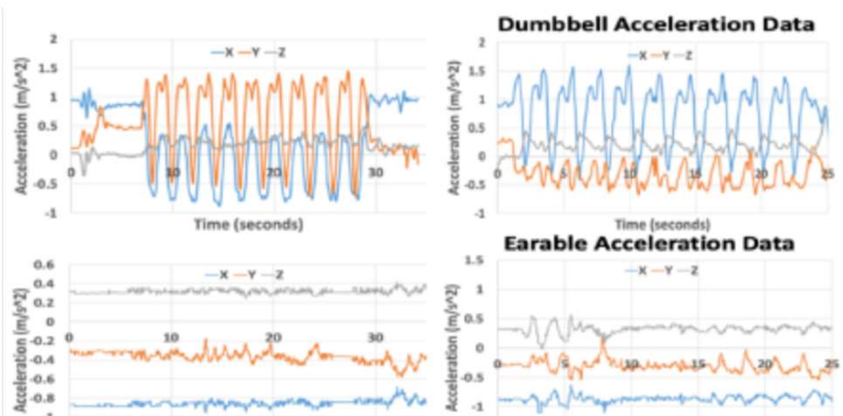
(a) Wrist Curl during Bicep Curls

(b) Swing body during Bicep Curls

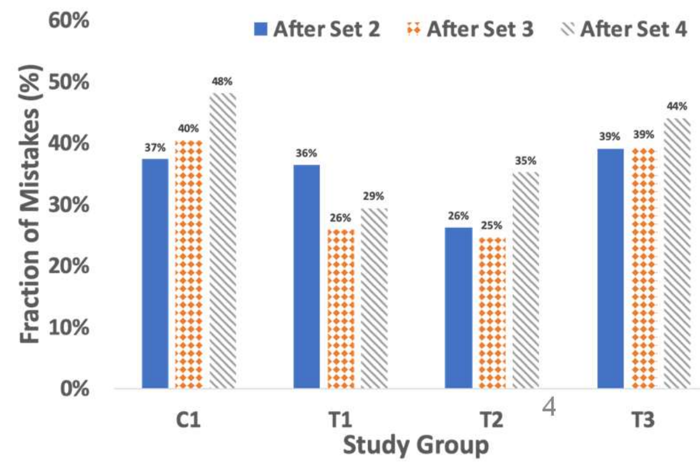
(c) Bend & swing head during Triceps Extension

(d) Arms not straight & Position Drift during Lateral Raises

(e) ROM Drift & Incomplete ROM during Lateral Raises



Feedback after every ~4 repetitions results in lower mistakes during set



Some Lessons Learnt

Pure Wearable/Mobile Sensing or Infrastructure Sensing isn't Enough

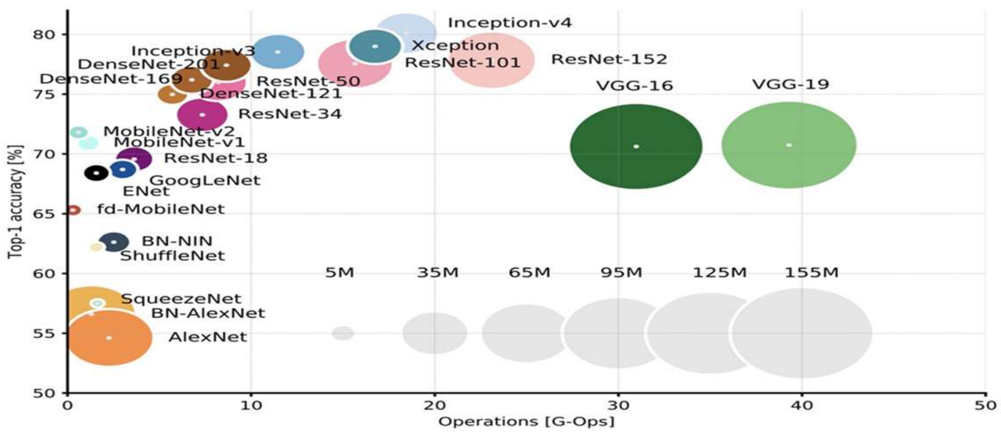
- Need to fuse inputs from personal and ambient sensors

Computation vs. Communication Tradeoffs are Changing

- Comms getting cheaper; computation more complex

	Proximity	NFC	ZigBee	BT	WiFi	LoRa
Distance	1 mm	10 cm	10–100 m	10–100 m	30–50 m	~km
Data rate	8–32 Gbps	0.021–0.48 Mbps	0.02–0.2 Mbps	0.8–2.1 Mbps	300 Mbps (11g) 7 Gbps (11ac, 11d)	200 Kbp
Energy-efficiency	4 pJ/b	1–50 nJ/b	5 nJ/b	15 nJ/b	5 nJ/b	1 μJ/b

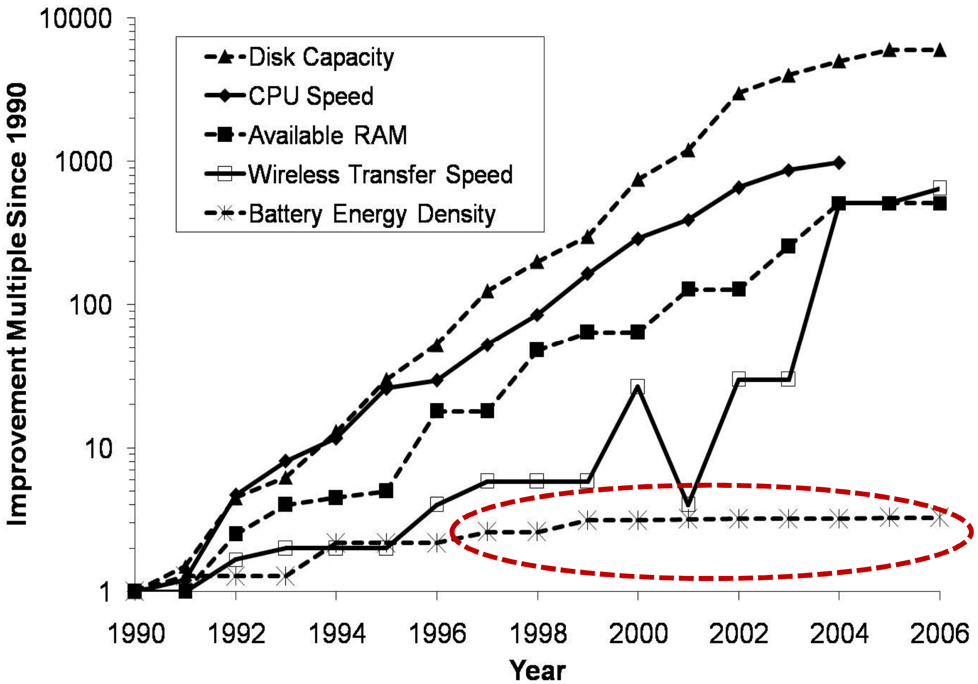
Source: doi: 10.1109/MIC.2018.011581520



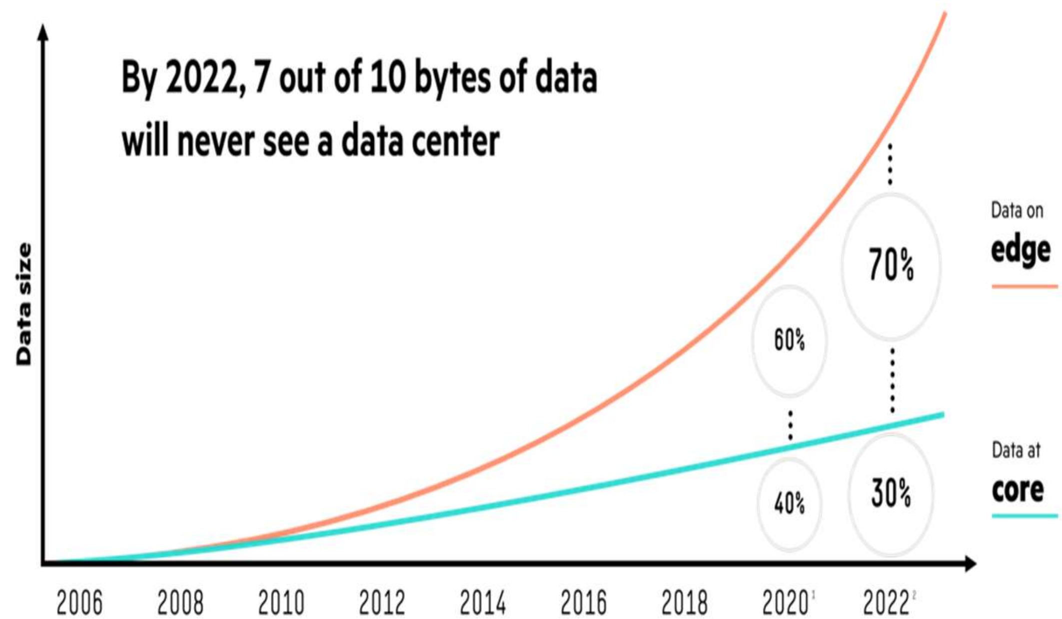
Source: A. Canziani, A. Paszke, E. Culurciello, An Analysis of Deep Neural Network Models for Practical Applications, *CoRR, May 2016*

Resource Bottlenecks & Trends

1. Where's the Resource Bottleneck?



2. The Rise of the "Edge"



1. International Data Corporation (IDC) https://www.idc.com/getfile.dyn?containerId=US41883016&attachmentId=47265871&id=null&bid=null&cid=null&patnerId=null
2. M2M Global Forecast & Analysis 2011-22

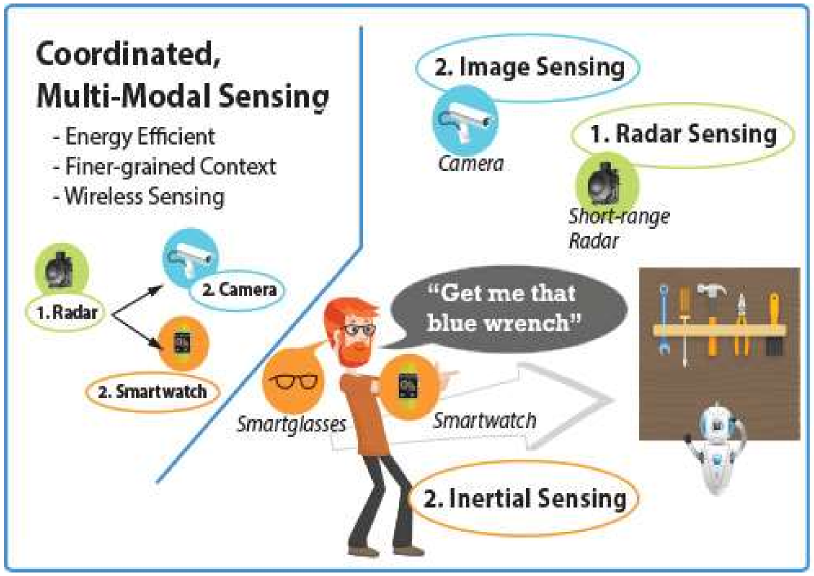
This Talk: Summary of Collaborative Machine Intelligence (CMI)

Collaboration is the Key to Realizing this Vision. Among:

- Wearable devices & Edge infrastructure
- Multiple IoT devices & Edge infrastructure

DS: Distributed & Triggered Sensing

Tightly coordinate Cheaper → Expensive Sensor Triggering



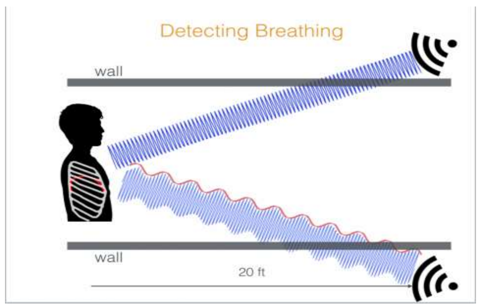
CMI: Collaborative ML-based Edge Intelligence

Distribute Inferencing Pipelines across multiple pervasive devices & across modalities → (Accuracy, Energy, Latency)

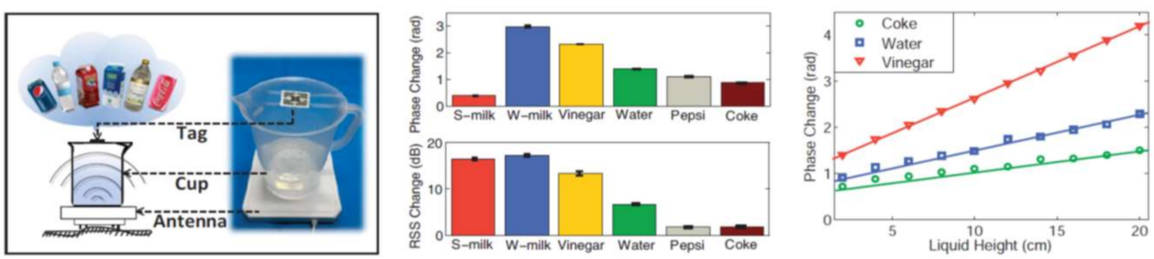
RF/Wireless: A Swiss-Army Knife

Sensing

- Use Radio signal reflections to capture gestures
  - WiSee: Doppler Shifts ➔ Movement Frequency
- Human Motion Artefacts
  - WiBreathe: Breathing Rate
  - Doppler Shift



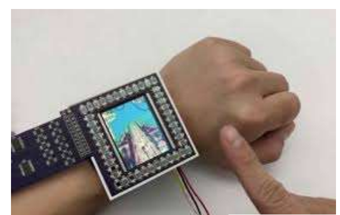
- Object Composition
  - RFID Phase Shift ➔ Shape & Liquid Detector



(a) Experimental setup. (b) Phase/RSS changes in different materials. (c) Phase changes vs liquid heights.

Energy Harvesting

- Multiple emerging modalities: light > vibration > temperature > RF
- Factors: size/form factor, on-body position, intrusiveness.



Ambient light

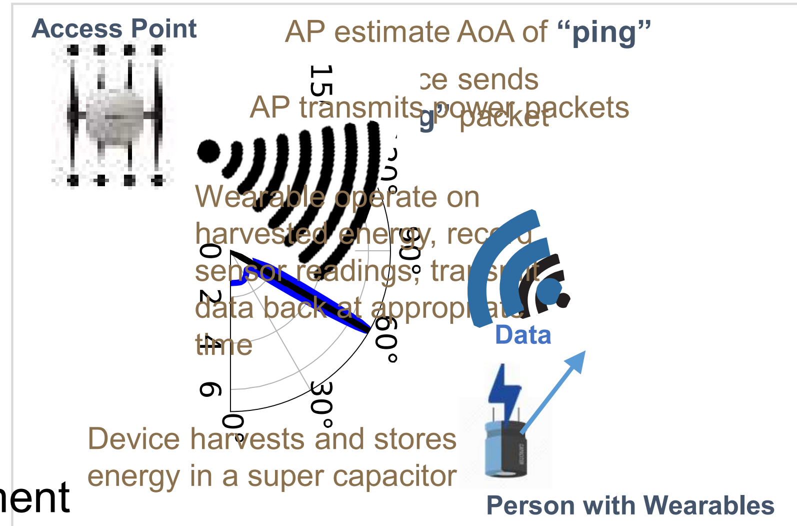
Vibration

Thermal gradient

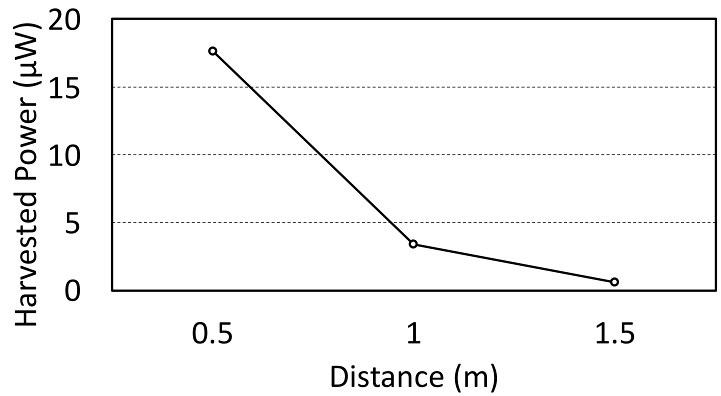
DS1. Battery Free Wearable/IoT Sensors

Percom 2019

Vision
- Utilize battery-free sensors on wearables & IoT devices to provide fine-grained tracking
- **Key breakthrough**: Charge devices wirelessly via WiFi "power packet" transmissions

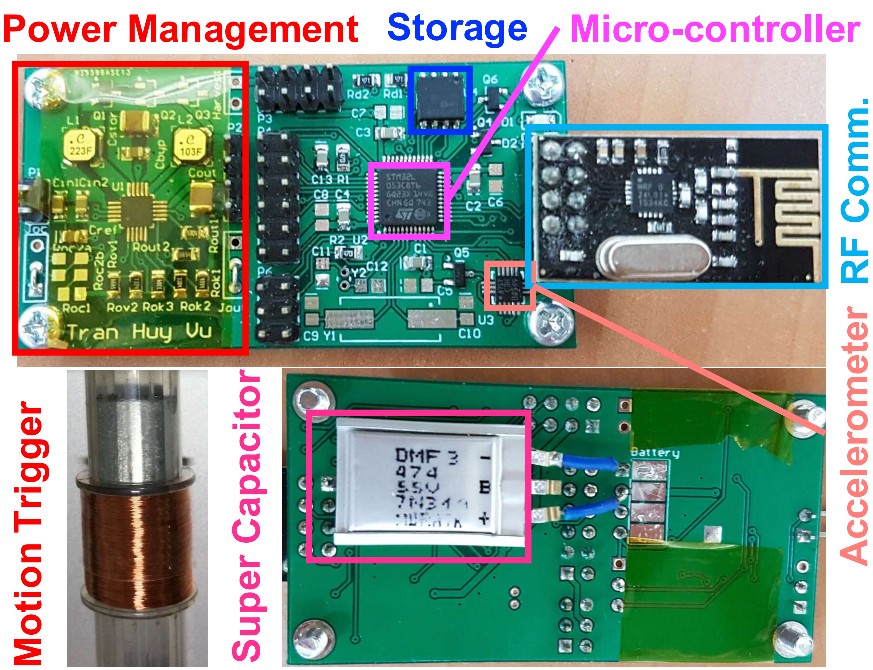


Applications
- Activity Tracking of Workers & Moving Equipment
- Product Monitoring in Warehouses
- Elderly Monitoring in smart homes



Challenges
- Low energy density using omnidirectional WiFi antenna *(< 1μW at 1.5m)*
- WiFi AP coordination to charge multiple devices

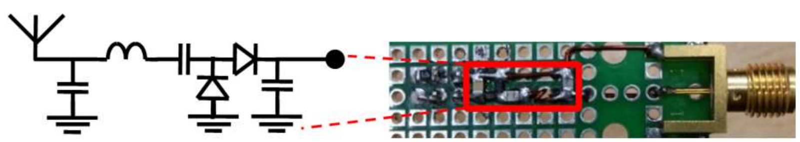
The Wearable + AP System



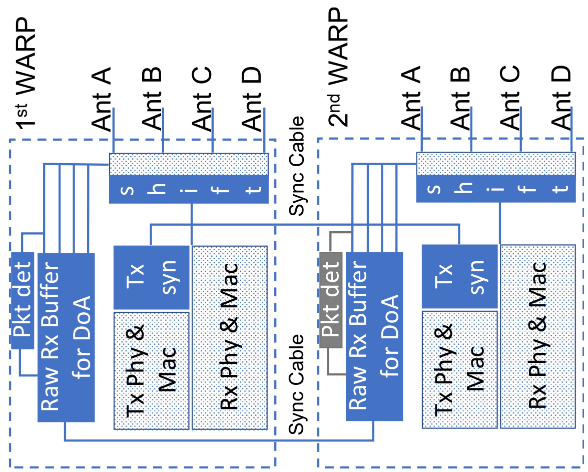
The Wearable

The Harvester:

- Matching Circuit
- Rectifier



The Beamforming AP

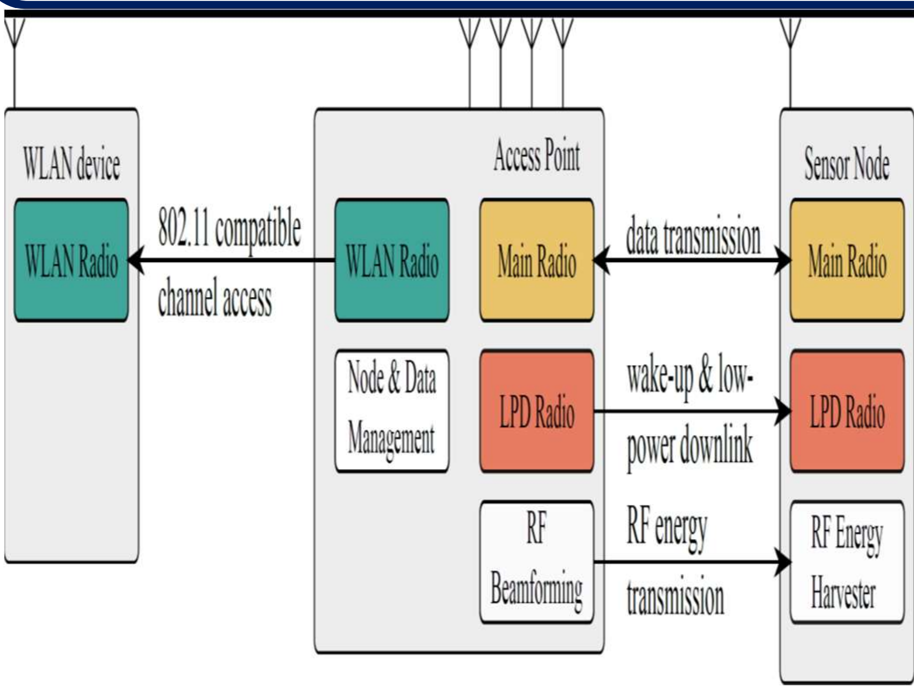


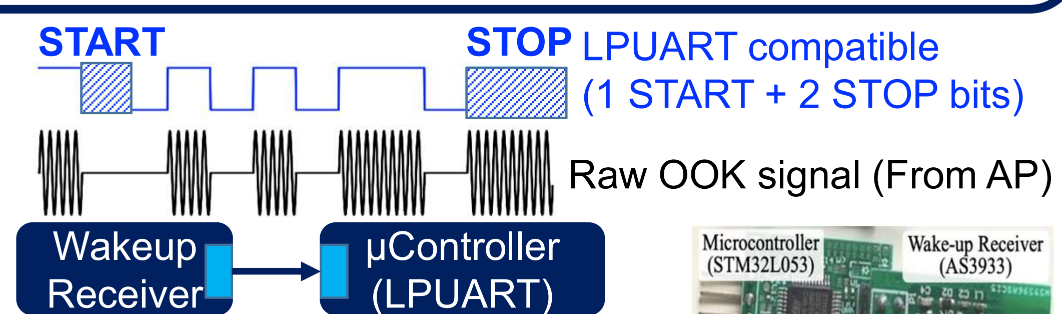
- Ping detection (nRF24L01+)
- Rx Buffer for AoA
- Tx Phase Sync for Beamforming

The WiWear++: Low-power downlink (LPD)

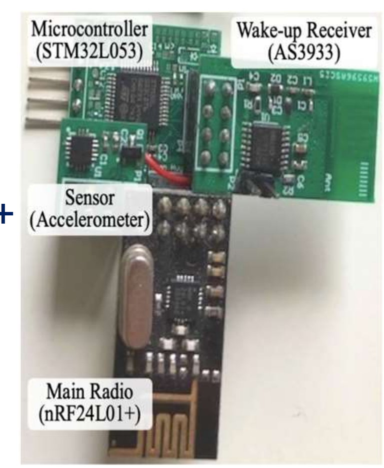
Under submission

- Base version: Ping triggered by significant motion;No MAC
- New: Use Wake-up Receivers to support low-power downlink (AP to device)
  - Proactive ping request (update orientation)
  - Content-free uplink trx





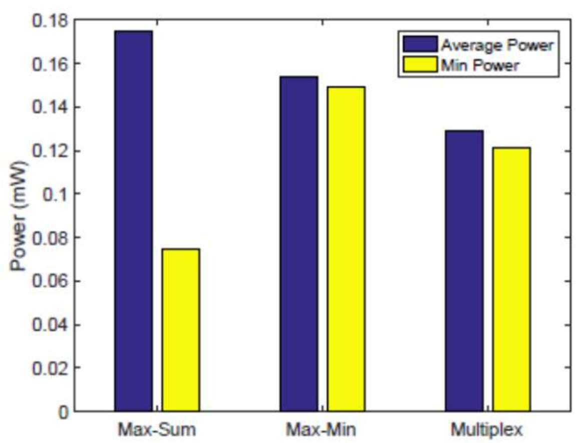
WiWear++ Prototype



µController only wakes up to read LPUART data register

The Cloud RAN & The Future of Multi-AP Operation

- Harvesting Power levels drop with multiple wearable devices

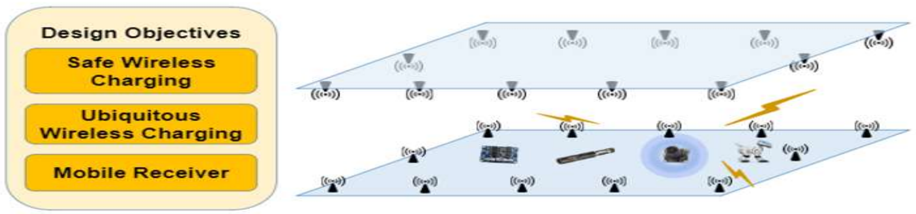


Power harvested (4 devices, 0.2m)

- **Future: What about multiple APs, that coordinate their transmissions?**
  - Complex balance between sensing, communication and energy transfer capacity

- Lots of distributed transmitters (915/964 MHz channels) surrounding the target.
  - Adjust phase→ distributed beamforming
  - 24 Trx (1.7W) in 20X20 m^2 → 0.6-0.7mW power harvested

EnergyBall, Ubicomp'19



Takeaways & Reflections

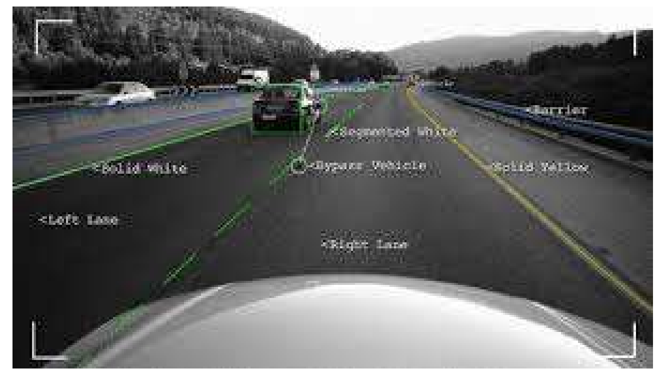
New opportunities:

- Edge-Coordinated Activation of sensing on wearable devices.
- Combination of passive RF sensing+ battery-less wearable/IoT devices
- **Edge ML needed to perform real-time *multi-modal* inferencing**

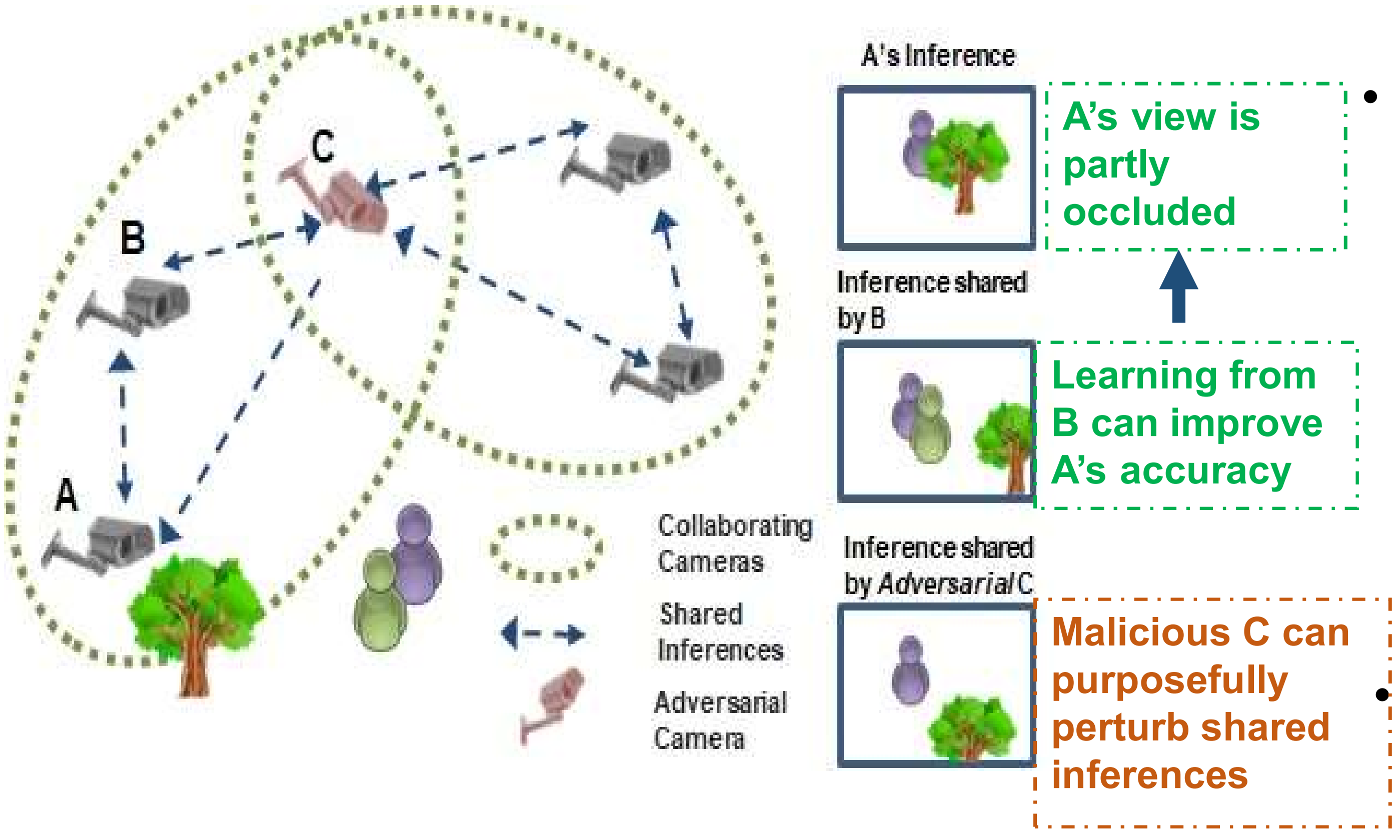
Collaborative MI: The Solution for Dependable Machine Intelligence

Key Idea: Overcome limitations in resource & fidelity by performing machine intelligence jointly

- **Real-time decision making**
  - Complex ML pipelines being executed on individual IoT devices or with edge-assistance
- **Key Resource & Performance Bottlenecks**
  - Latency of DNN execution
    - 550 msec+ for person recognition/frame on a Movidius co-processor (1W)
  - Low Accuracy
    - Individual sensors subject to environmental artefacts
  - Energy Overhead
    - Need to support battery-less operations



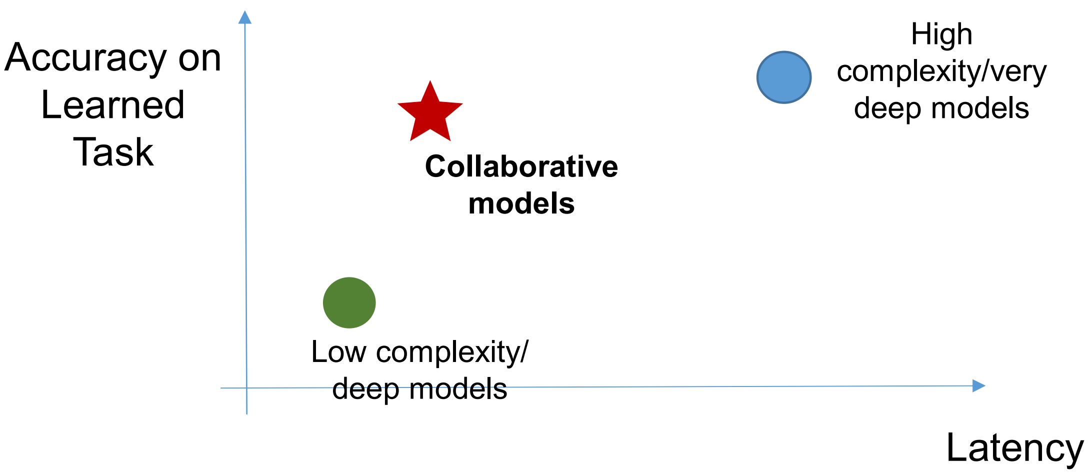
EA1: Collaborative IoT & The Edge: Ongoing Work



- **Collaborative Sensing**
  - Spatial and/or temporal overlap among sensors
    - ***Sensor Multiplicity***
  - Adjust Inferencing Pipeline ***on-the-fly***

- **Dependable Systems**
  - Resilience to Adversarial Attacks

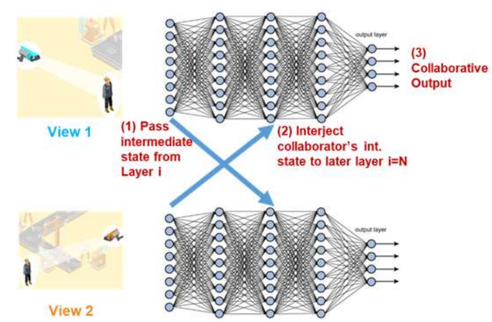
Collaborative IoT & The Edge: Ongoing Work



Closing the accuracy gap with collaboration

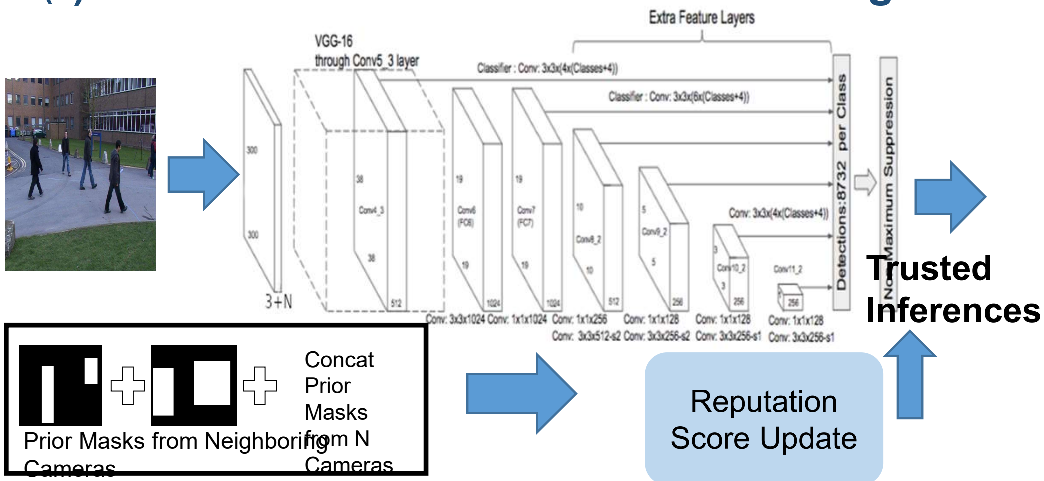
Design Goals

1. Requires NO re-training of the DNN models
2. Backward compatibility to non-collaborative mode when no collaborators are available
3. Minimal latency and bandwidth overhead for infusing collaborative input

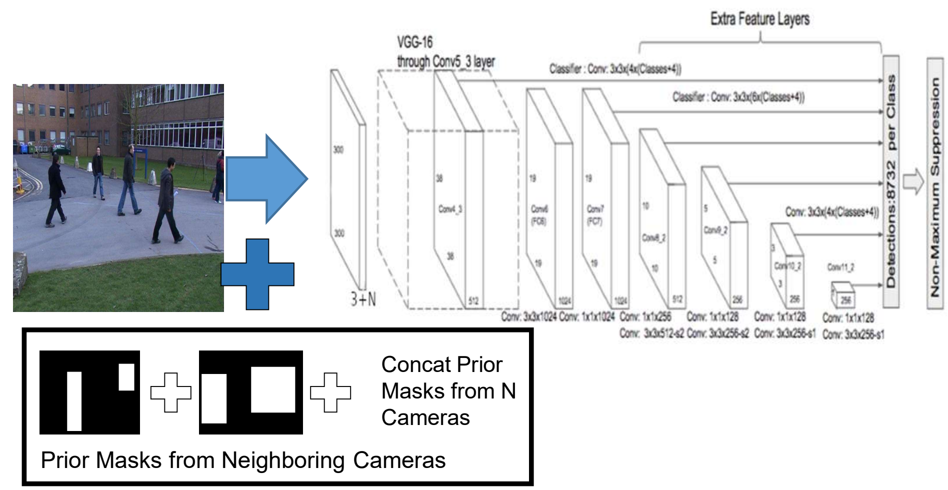


Approach #1: Run-Time Collaborative Inferencing

(a) CNMS: Collaboration at Decision Stage



(b) CSSD: Collaboration at Input Stage



PETS Dataset *(8 cameras)*

- *Person detection using SSD300; Homographic View mapping*

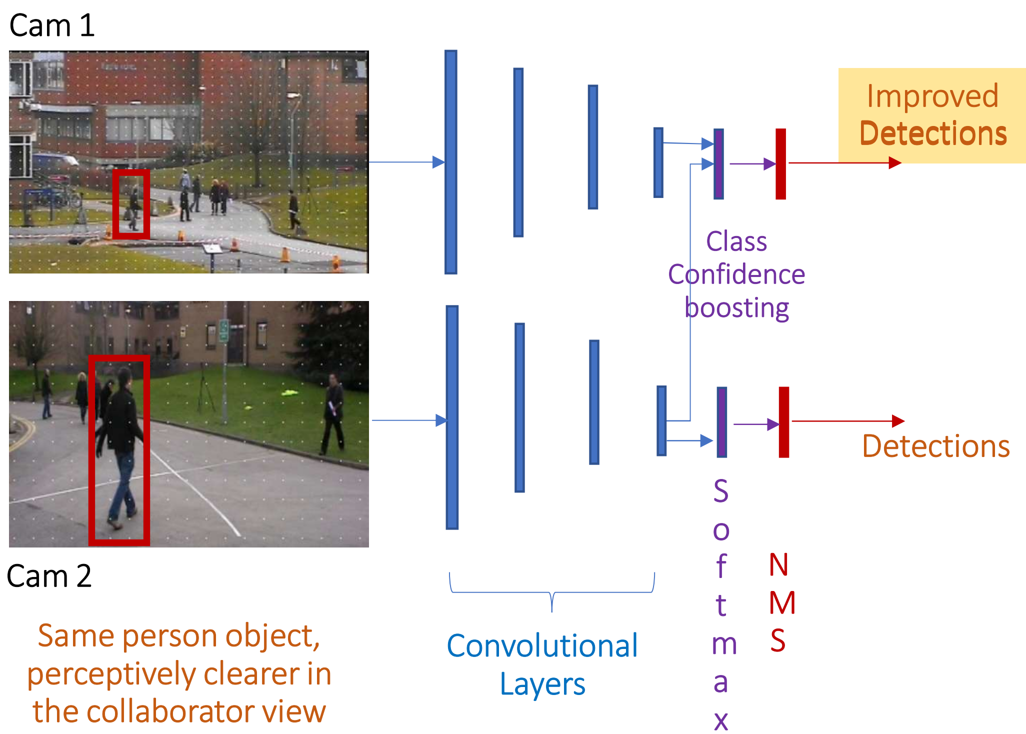
(a) Reference camera view (b) Collaborative camera 1 (c) Collaborative camera 2 (d) Collaborative camera 3

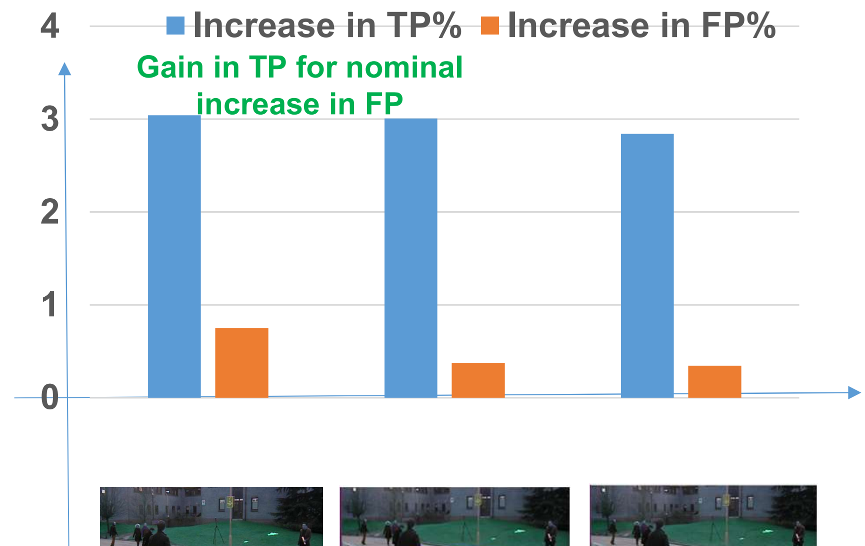
High accuracy improvement with minimal latency

	Inference Time	Accuracy
SSD Baseline	80ms	71%
Collaborative SSD	**85ms**	**82.2%**
CNMS	100ms	75.5%

Approach #2.1: Adapting the ML Pipelines *“On the Fly”* for Improved Accuracy

Exploration #1
Improving Accuracy through Sensor Multiplicity

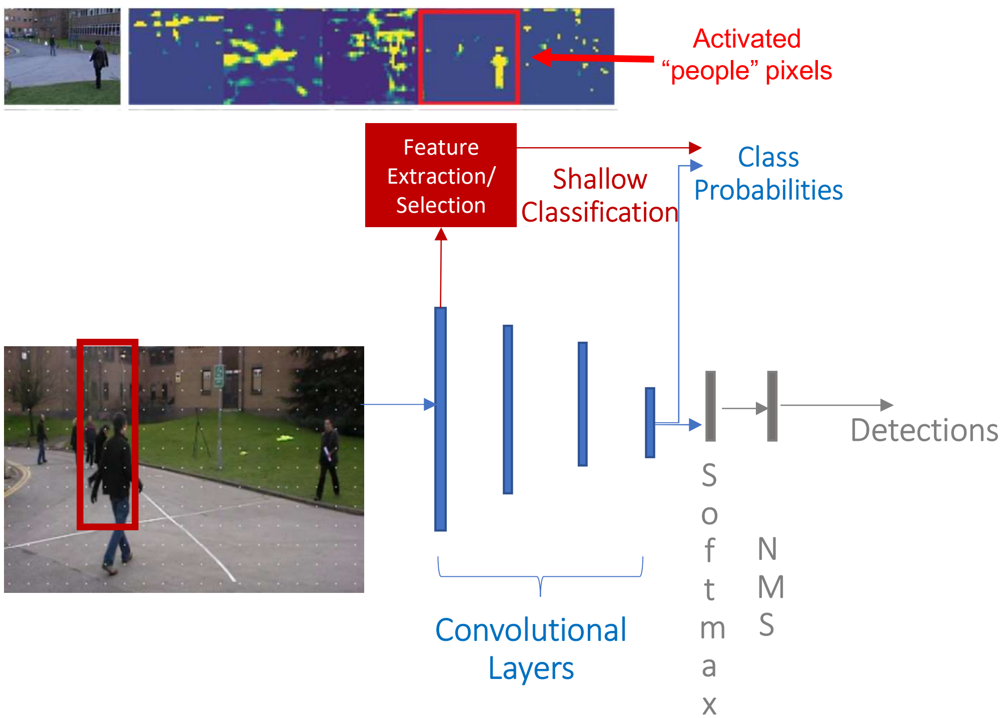




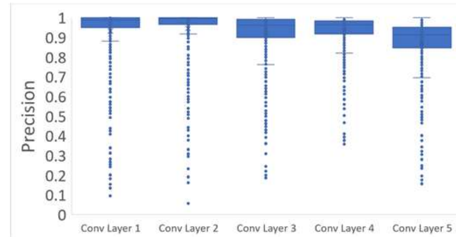
View 8+5 (33% overlap) View 8+6 (62%) View 8+7 (38%)

Approach #2.2: Using Hidden Layer State for Collaborative Classification

Exploration #2
Early Estimation and Hybrid Classification



Accuracy of Discriminant FMaps in Detecting People Pixels



- PETS Dataset, Camera 7 as Reference View (over 795 frames).
- **Precision =**

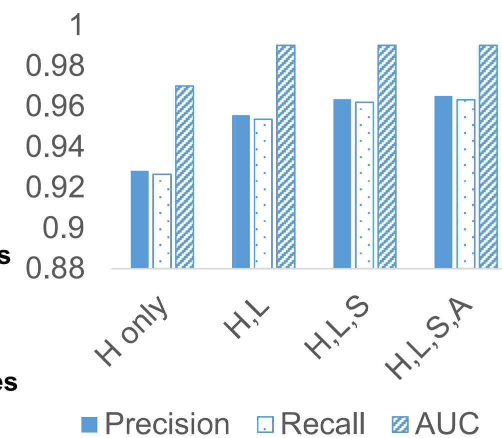
$$\frac{\textit{Pixels activated by discriminant fmaps that overlap with Ground truth Bboxes}}{\textit{Pixels activated}}$$

Average precision of over 93% in detecting target-specific pixels as early as Layer 1

Accuracy of Shallow Person Classification

AUC > 95%
as early as Layer 1

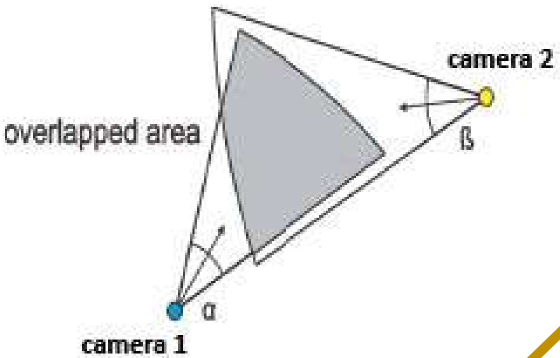
H – Histogram of Feature values only
L – Location of Anchor Boxes
S – Scale of anchor boxes
A – Aspect ratio of anchor boxes



Coupled Collaboration+ Sensing: CollabCam

Reduce Sensor Sampling Energy by Reducing Camera Image Resolution

Overlapping FoVs



Object Matching

Estimated Overlap

Resolution Reduction

CollabCam: Mixed Resolution & Accuracy

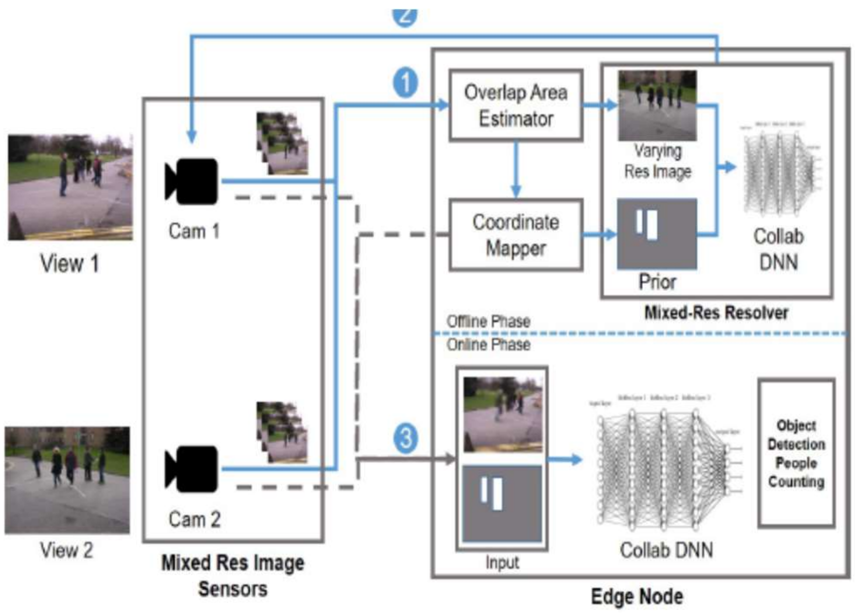
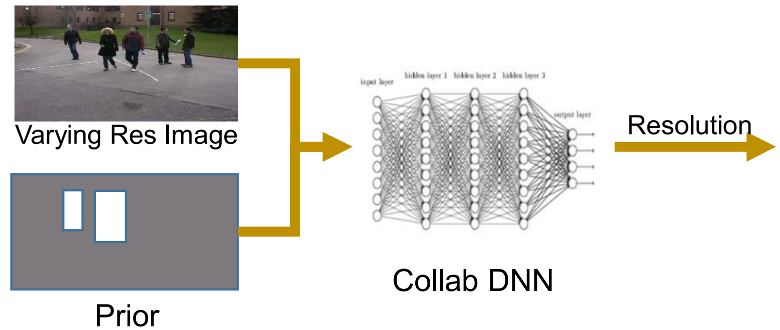


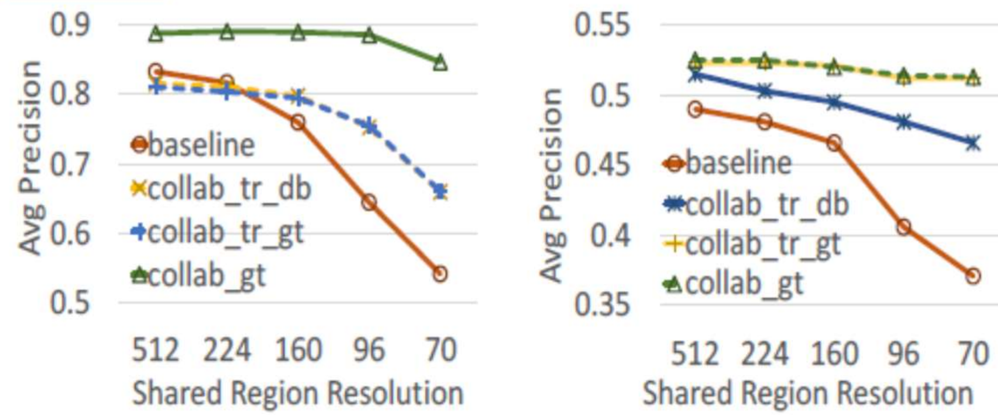
Figure 4: *CollabCam*: Functional Components on Vision Sensor & Edge

Overall Networked Vision Sensing Architecture

Shared-Area Resolution Estimation



Mixed-Resolution vs DNN Accuracy



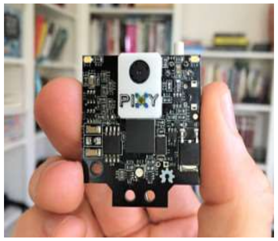
(a) PETS (Cam 8,5) (b) WILDTRACK (Cam 1,4)

SMU SINGAPORE MANAGEMENT UNIVERSITY | School of Information Systems

Opportunity #3: Sensing Energy Savings

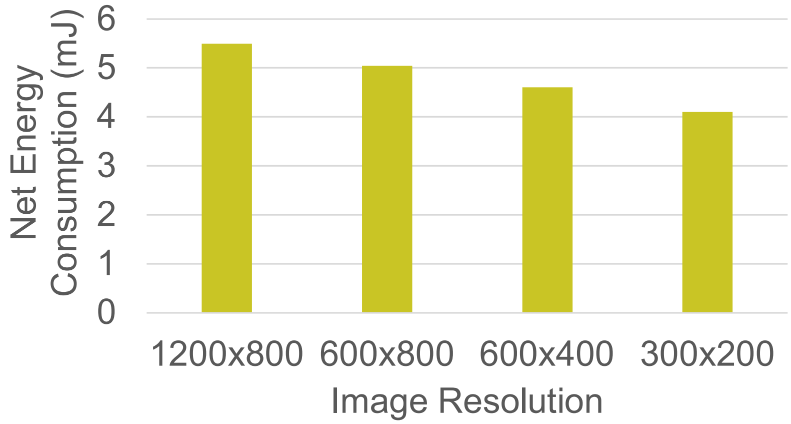
Camera Prototype

- CMUCam-5 (Pixy 2 Platform)
- Camera: Aptina MT9M114 CMOS
- NXP LPC4330 dual-core ARM processor @ 204MHz
- 264kb SRAM | 2 Mb Flash Memory
- Firmware modification → mixed resolution capture



Sensing Energy Savings

Net Energy = Total Energy – Baseline Energy



1200x800 → 300x200 |~25% Energy Reduction

Observation from Experiments:

- ***Reduced Resolution lowers sensing energy***
- ***Energy proportionality requires additional adaptive clocking of sensor***

ML & Network Status Coupling:

- DNN can adapt to differing resolution and data rates from individual sources
- Data rate selection can depend on network congestion+ device state

Takeaways & Reflections

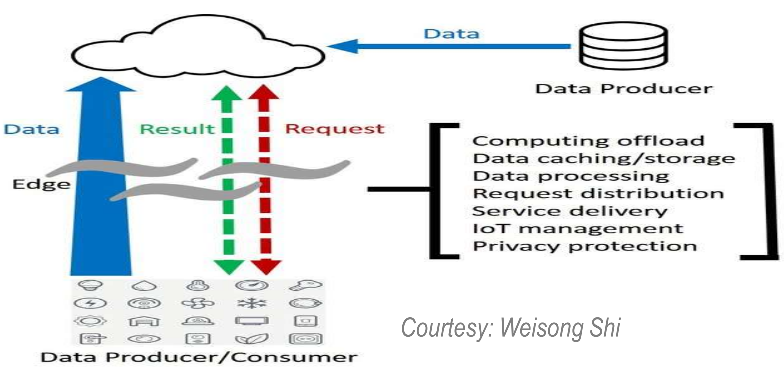
New opportunities:
- Edge-Coordinated Activation of sensing on wearable devices.
- Combination of passive RF sensing+ battery-less wearable/IoT devices
- **Edge ML needed to perform real-time multi-modal inferencing**

- ML Coordination between a set of distributed edge (IoT) & wearable devices
- Run-time Collaboration: Improve Accuracy, Energy & Latency
- **Collaborative ML (Training) requires**
  - **new DNN architectures**
  - **network-aware DNN adaptation**

Edge Computing at Present

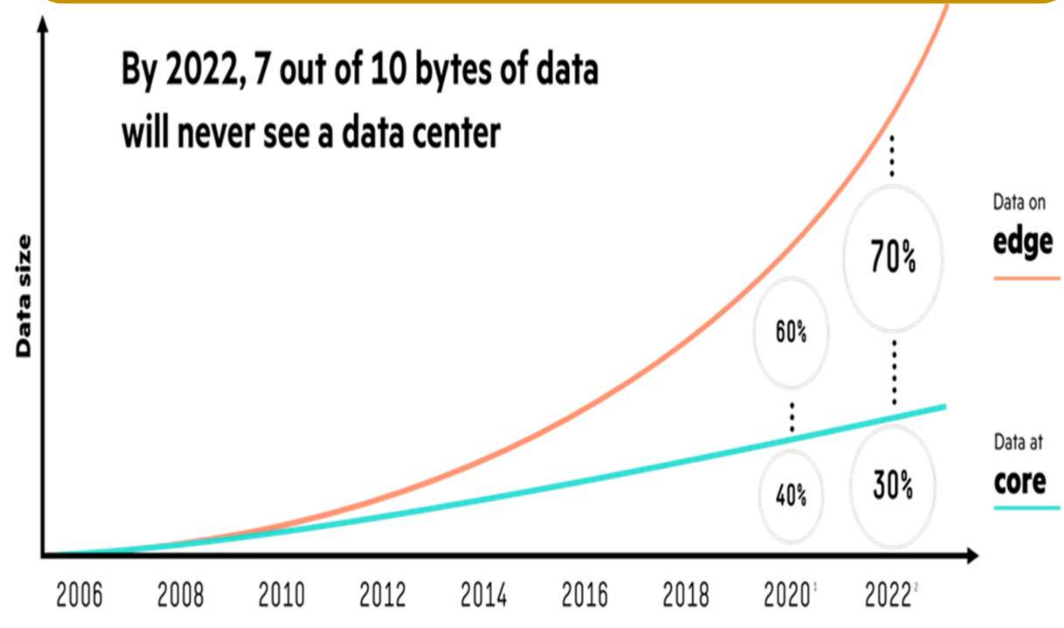
- Offload computation to a nearby, powerful-computational entity
  - **Edge provides isolation and resource augmentation**

- **Advantages**
  - Low-latency, real time ML pipelines
  - Data privacy
  - Energy-efficiency



Courtesy: Weisong Shi

- **Isolated Interaction between individual device & "cloudlet"**



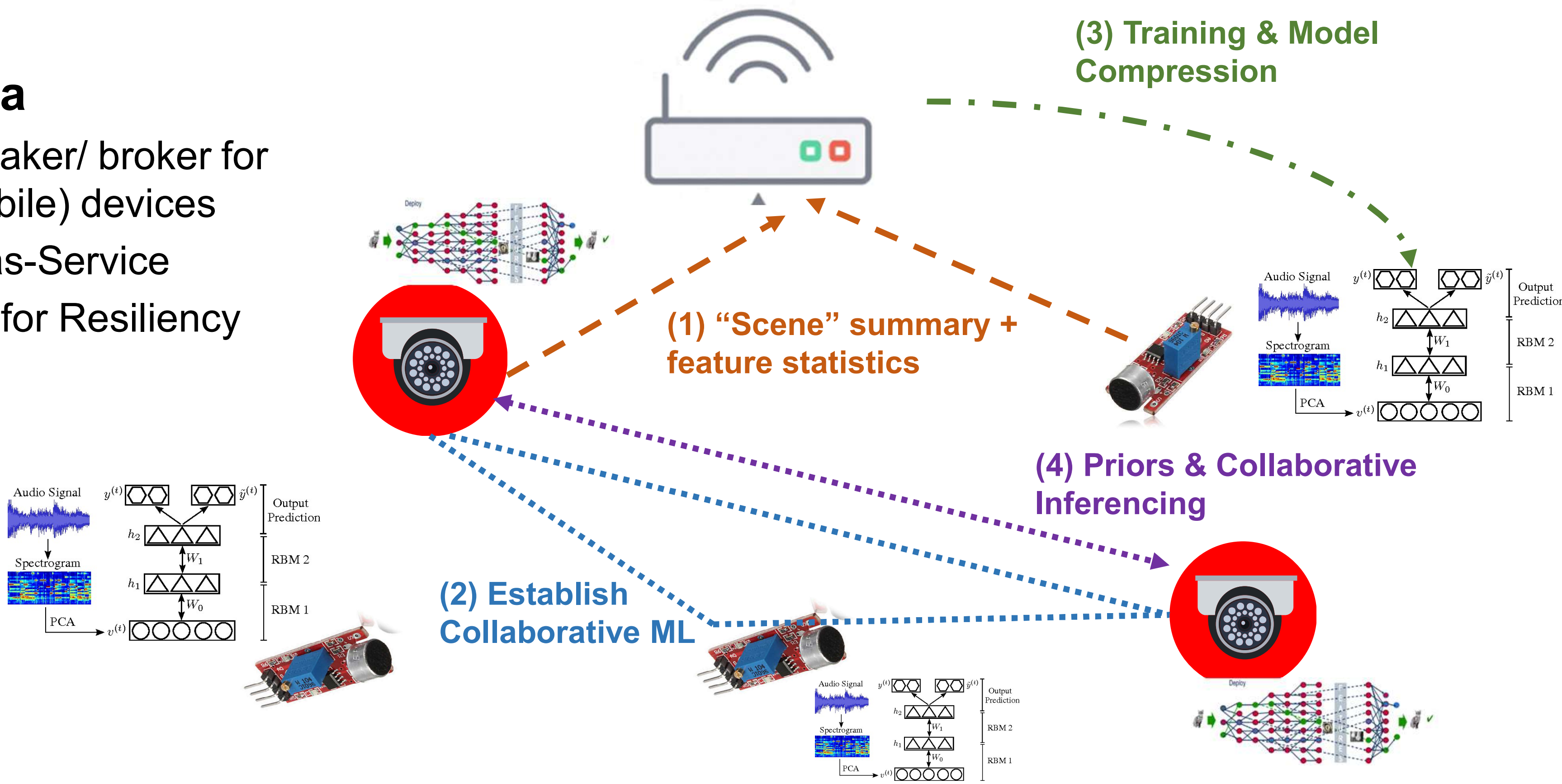
1. International Data Corporation (IDC) https://www.idc.com/getfile.dyn?containerId=US41883016&attachmentId=47265871&id=null&bid=null&cid=null&patnerId=null
2. M2M Global Forecast & Analysis 2011-22

My Vision: Cognitive Edge for IoT

Edge enables CMI

Edge as a

- Matchmaker/ broker for IoT (mobile) devices
- ML-as-as-Service
- Monitor for Resiliency



Challenges for The Cognitive Edge

- **Find Useful Spatiotemporal Correlations among Devices**
  - Minimizing Communication Overhead
  - Handling Disparate Sensing Modalities
  - Handle Redundancy in Dense IoT Deployments

- **Enable trusted interactions among Devices**
  - Find Correlations from non-sensitive Metadata/Features
  - Identify and isolate malicious/non-conformant devices

- **Handle Dynamic Workloads**
  - Mobile devices that temporarily reside in specific areas
  - Changes in spatiotemporal human/event patterns

Takeaways & Reflections

New opportunities:

- Edge-Coordinated Activation of sensing on wearable devices.
- Combination of passive RF sensing+ battery-less wearable/IoT devices
- **Edge ML needed to perform real-time multi-modal inferencing**

- **Edge as a Dynamic Matchmaker & Orchestrator between "dumb" IoT Devices**

- ML Coordination between a set of distributed edge (IoT) & wearable devices
- Run-time Collaboration: Improve Accuracy, Energy & Latency
- **Collaborative ML (Training) requires**
  - **new DNN architectures**
  - **network-aware DNN adaptation**

SMU Singapore Management University | School of Information Systems

Conclusion

- **Need for greater interaction between wearable devices & edge computing/network entities**
  - **Key to 100-fold decrease in power consumption on pervasive platforms**

- **Need for inferencing orchestration among edge devices**
  - Significant opportunities for scaling up ML-based applications
  - Need for standardized models for distributing computational state
  - Need for stackable ML models for accommodating sensing diversity

- **Need for Edge Platforms to be enablers of such multi-device orchestration**
  - Need to rethink the role of edge computing
  - Adaptive computational resources to support DNN vs. network tradeoffs

(E) archanm@smu.edu.sg
(U) https://sites.google.com/view/archan-misra