CPT Bryce Wilkins (Army AI2C) & Dr. Robert Edman (CMU SEI)

Sensor Fusion for Automated Gathering of Labeled Data in Edge Settings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

BLUF

Need: High quality labeled data from operational settings, acquired rapidly and affordably.

Approach: Distributed sensing provides multiple views and multiple data modalities allowing that can be aggregated using sensor fusing methods, providing high enough aggregate confidence to allow automatic labeling.

Benefit: Faster adaptation and detection in operational environments. This is achieved through faster data labeling and model retraining in edge scenarios.

Competition:

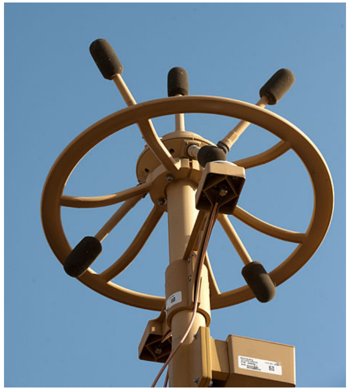
1. Store all data, collect it at a central location, and hand label it.
2. Few-shot/low shot methods

Operational Scenario/CONOPS

Forward Operating Base (FOB) Perimeter Defense

- **Sensors:** thermal imagery, motion sensors, UAS overhead imagery, acoustic sensors
- **Scenario:** at night, an adversary is attempting to emplace an IED near a FOB. Motion detectors near the perimeter are triggered, and thermal imagery is used to confirm that people are outside the FOB. Acoustic sensors pick up the sound of digging, which could be indicative of emplacing an IED. It is too dark to use the UAS overhead imagery, but the data from the other sensors are enough to achieve high system-level belief that a threat is present.
- **Reaction:** the high confidence of the AIF system triggers an immediate alert to the FOB security force, and they illuminate the suspected area for better visibility. The UAS overhead imagery is now able to quickly pinpoint the location of the adversaries.
- **Data Storage and Re-training:** as the data is being analyzed by the AIF instances, new labeled data can be generated when the system has high confidence in a threat detection. This newly labeled data is used to re-train the AIF instances to be able to react faster to this kind of threat in the future.

Sensor Variety



System Requirements

- Multiple Data Modalities
- Distributed Belief Aggregation Algorithm
- Detection/Label/Retrain Policies
- Independent Validation
- Model Exchange

- Time to Detect
- Ability to response
- False positive rates (validation)
- IFF (validation)

Key decisions

Belief Update Policy:

What automated method do I used to compute the belief of the system overall, rather than the (uncalibrated) probability of an isolated detection

- Baseline
- Max
- Average
- Voting
- Viterbi
- Bayesian

Labeling Policy:

How do I use the belief estimates to decide when to label a gathered example and how/where do I store that example.

- Simple Threshold
- Weighted Threshold and logic
- Threshold over time
- Spectral clustering
- Semi-supervised

Retrain Policy:

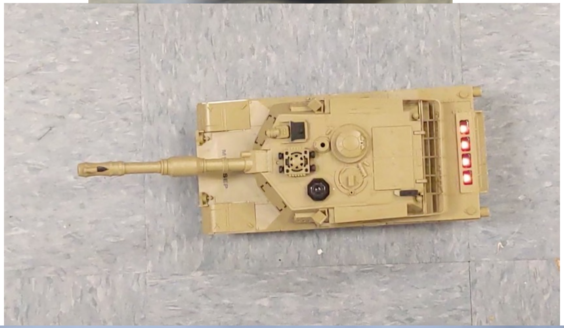
At what point to I have enough data to retrigger retraining and how do I retrain

- Online updates versus batch
- Quantity of new data
- Independence of new data

And, deploy policy/exchange

Toy Data

- Collected scavenged data for training from google maps, street view, image + gathered acoustic data
- Filmed in 4 locations: Office, Lab, Garage, Snowfall
  - Each location has 4 sensor streams: Audio, Overhead, Left, Right
  - And collected positive/negative examples from each location
- Both positive and negative scenarios include confuser objects like tables, chairs, cars, bicycles



Data Holding/Handling

- Pre-loaded training data
  - Backgrounds
  - Positive Examples
  - Negative Examples
- Harvested Data
  - Confidence Examples
  - Confidence non examples
- Validation Data

Automate as much possible the data processing

- Generate composites
- Balance dataset

	AIFI1 (Acoustic)	AIFI2 (Overhead)	AIFI3 (Stationary Ground)	AIFI4 (Panning Ground)
Positive Examples	540	17	32	32
Negative Examples	396	15	14	14
Backgrounds	N/A	16	24	24
Validation	205	204	204	204

Model Training

- Limited data available means we have to be smart about the training
- Test data is 70/30 split of the training data, validation data is fully held out.
  - Drawing randomly from test and validation data
  - Adding noise to sampling
- 5000 samples from a 44khz stream (around 1/9th of a second)

Vision classification models are resnet-50 fine tuned; input images are 224x224x3

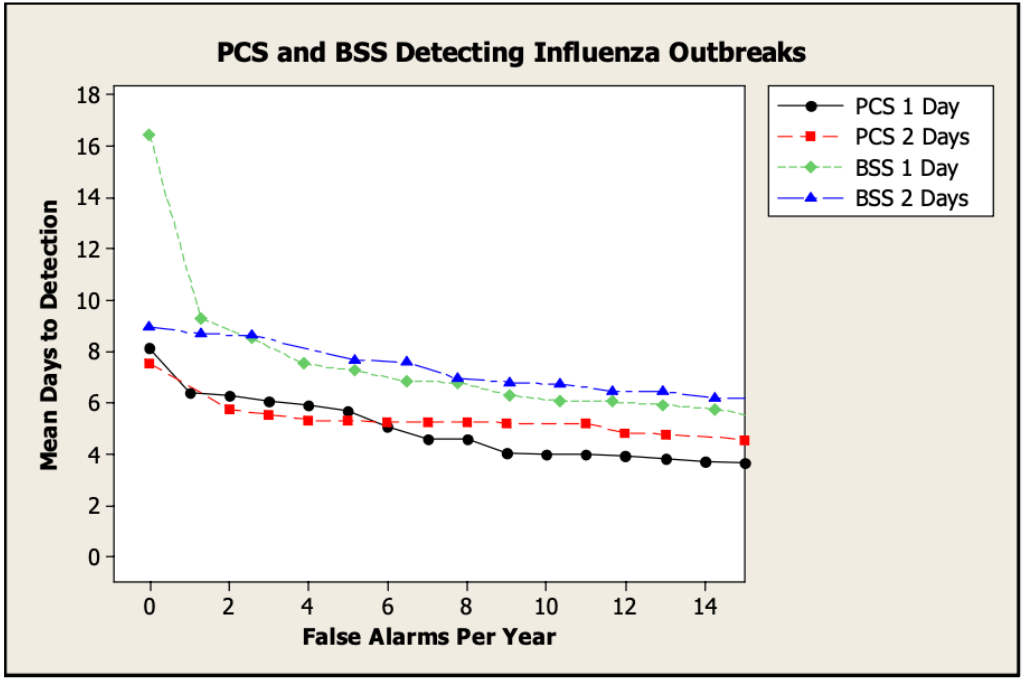
A fine-tuned mini-rocket time series classification model

Evaluation criteria

Using Activity Monitoring Operational Characteristic (AMOC) curve.

Method commonly used for disease outbreak detection

- Time to detection
- False positive rate
- Missed detections



AMOC

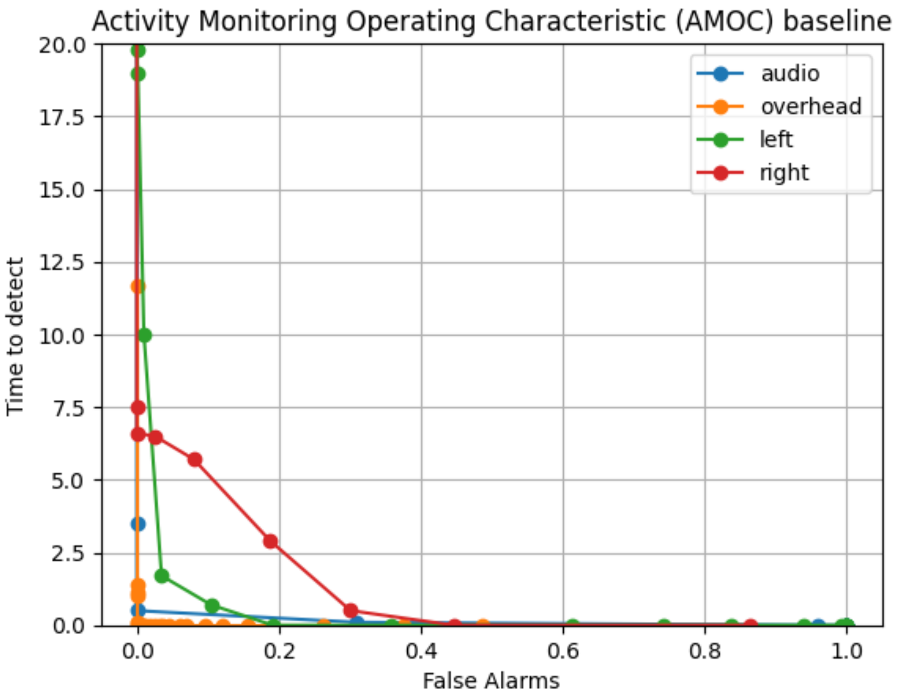
Using synthetic injection to produce enough examples to evaluate:

- Mean time to detect
- Detections outside of positive frame.

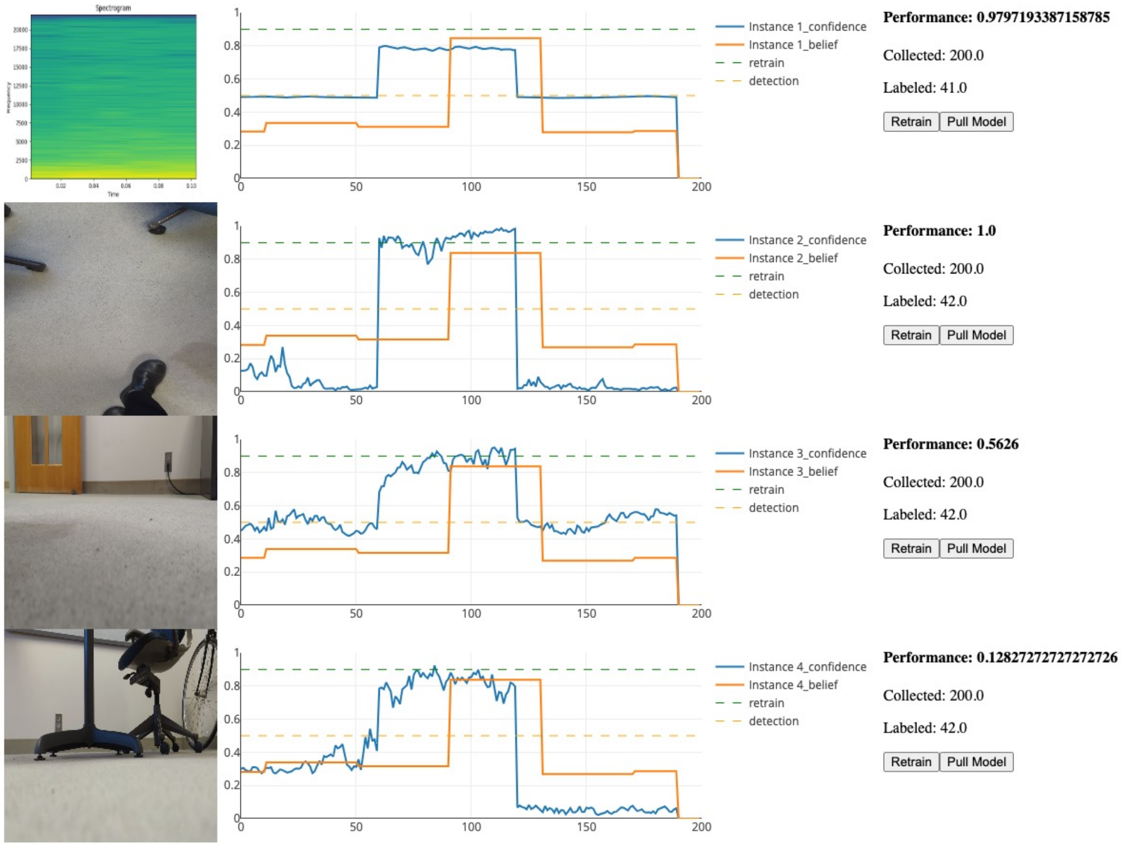
Synthetic Data generation:

- Randomize Scenario, use frames from positive and negative example
  - Randomize start frame
  - Randomize end frame

Limitation: Data in the middle is seen multiple times



Dashboard



Spectrogram
Time
Instance 1_confidence
Instance 1_belief
retrain
detection
Performance: 0.9797193387158785
Collected: 200.0
Labeled: 41.0
Retrain
Pull Model
Instance 2_confidence
Instance 2_belief
retrain
detection
Performance: 1.0
Collected: 200.0
Labeled: 42.0
Retrain
Pull Model
Instance 3_confidence
Instance 3_belief
retrain
detection
Performance: 0.5626
Collected: 200.0
Labeled: 42.0
Retrain
Pull Model
Instance 4_confidence
Instance 4_belief
retrain
detection
Performance: 0.12827272727272726
Collected: 200.0
Labeled: 42.0
Retrain
Pull Model

Partners and other collaborations

Carnegie Mellon University Software Engineering Institute

OUSD(R&E) Integrated Product Team 2 (IPT-2)

- Focused on cross-echelon, DDIL-resilient decision and control distributed AI architecture
- Collaboration between AI2C, ARL, NRL, and AFRL:
  - AI2C: Developing OS.AI
  - ARL: Model deployment and monitoring
  - NRL: Hierarchical C2
  - AFRL: Distributed AI Broker

CMU Basic Research

- Distributed model training
- Distributed logistics
- Distributed network design

AI2C's Distributed AI "Stack"

CMU: provides models that can be used by AIF instances
AI Fusion: provides the framework for belief sharing among AIF instances
IPT-2: provides the architecture for distributed AI

Lessons Learned

- Data Lessons
  - Data Augmentation allows limited data to go further
  - The gold-standard data for monitoring performance is critical
- Model Lessons
  - Transfer learning helps a lot
  - And careful chose of models
- System Lessons
  - Localization and object deduplication matters a lot
  - Separate model and system performance
    - Use model specific metrics (e.g. pixel counts) for model specific metrics
    - Use system metrics to evaluate design decisions

Thank You

Contact Information

CPT Bryce Wilkins
Data Scientist
Army Artificial Intelligence Integration Center

Email: bryce.d.wilkins.mil@army.mil

Dr. Rob Edman
Machine Learning Research Scientist
Software Engineering Institute/Carnegie Mellon University

Email: rbedman@sei.cmu.edu

Dr. Jose Morales
Senior Researcher
Software Engineering Institute/Carnegie Mellon University

Email: jamorales@sei.cmu.edu