Efficient Edge Learning for Thermal Vision

Closing the Robustness Gap in Edge-Based Perception

Zubin Bhuyan

Civil and Environmental Engineering
University of Massachusetts Lowell

NEW ENGLAND ITS 2025 ANNUAL INTERCHANGE
Boston, MA
October 9, 2025

Edge AI vs. Foundational Models

Foundational Models (Cloud / General-Purpose)

- Massive scale & resources.
- **Require centralized compute**, connectivity, and energy.
- Offer broad generalization but **lack contextual grounding** in real-world conditions.

Edge AI (Task-Specific & Adaptive)

- Runs on-device, **real-time perception** and decision-making.
- **Optimized for low power**, low latency.
- **Environment-aware and self-tuning**: adapts to sensor noise, lighting, or hardware variation, ideal for deployed systems.

Manufacturing
Predictive maintenance
Safety

Agriculture
Navigation
Harvesting
Spraying

Medical Imaging
Disease Monitoring
Disease diagnosis

ITS
Traffic motoring
Analysis

Sports/ Fitness
Telehealth
Vital tracking
Training

Security
Surveillance

Edge use cases or examples

Active Learning for Edge AI

Hybrid Workflow

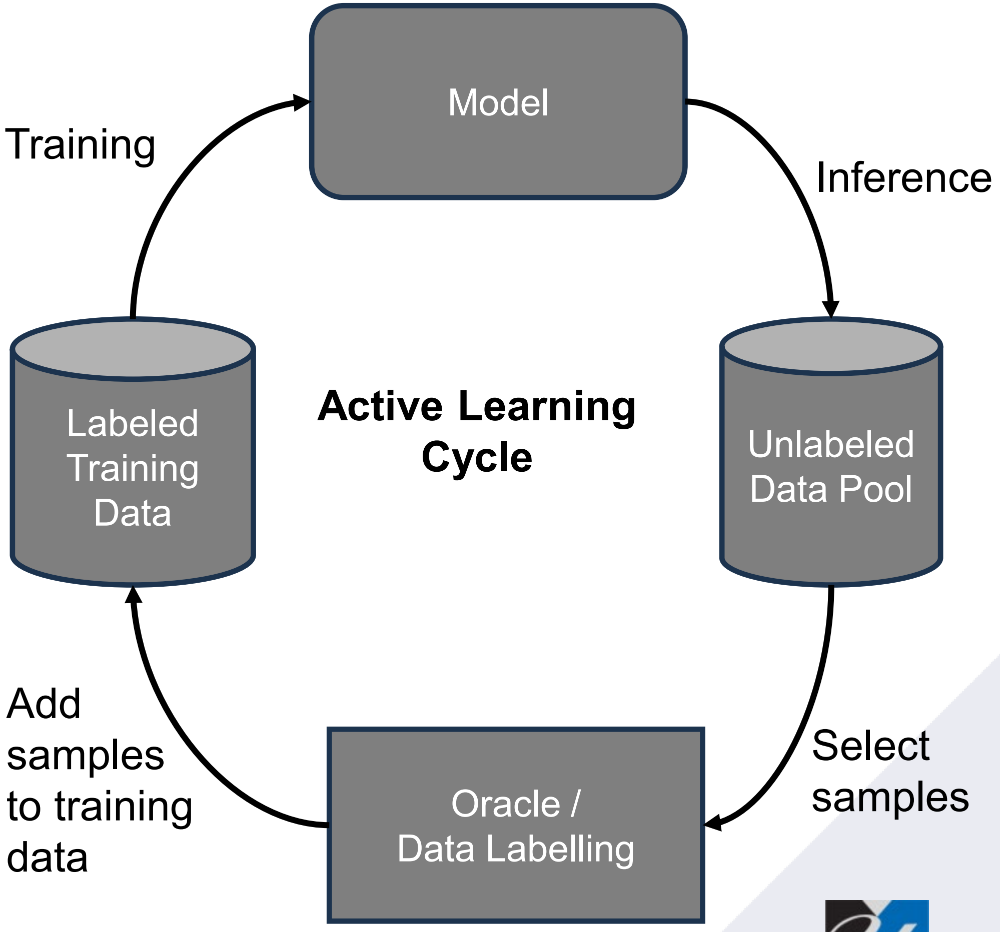
- The model is trained on a central server and deployed to the edge device for continuous inference.

Continuous Inference & Data Selection

- While running on-device, the model **analyzes incoming data streams in real time** and selects informative samples for future training, **without interrupting inference**.

Edge Device Setting

- Platforms: Raspberry Pi CM + HAILO Accelerator 26T.
- Connected via 4G, LoRa to servers and nearby devices such as flashing signs or alarms for local coordination



Task: Pedestrian Detection

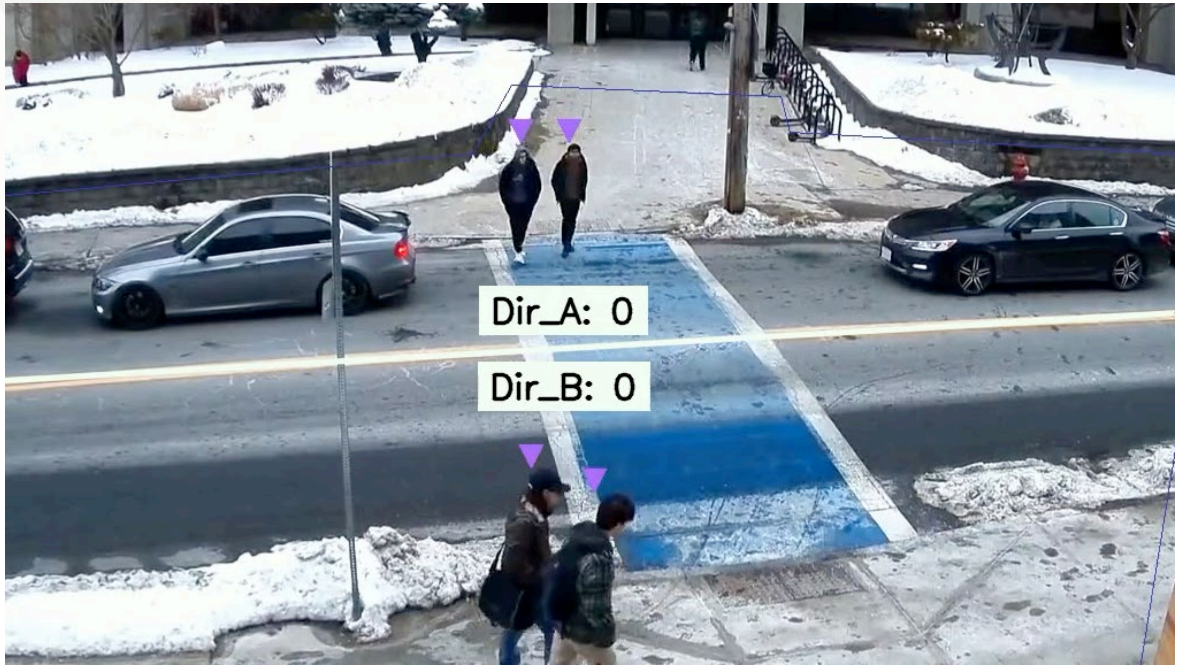
- **Environmental Diversity & Generalization**: Our base dataset spans urban parks, dense cityscapes, and tree-covered suburban roads.

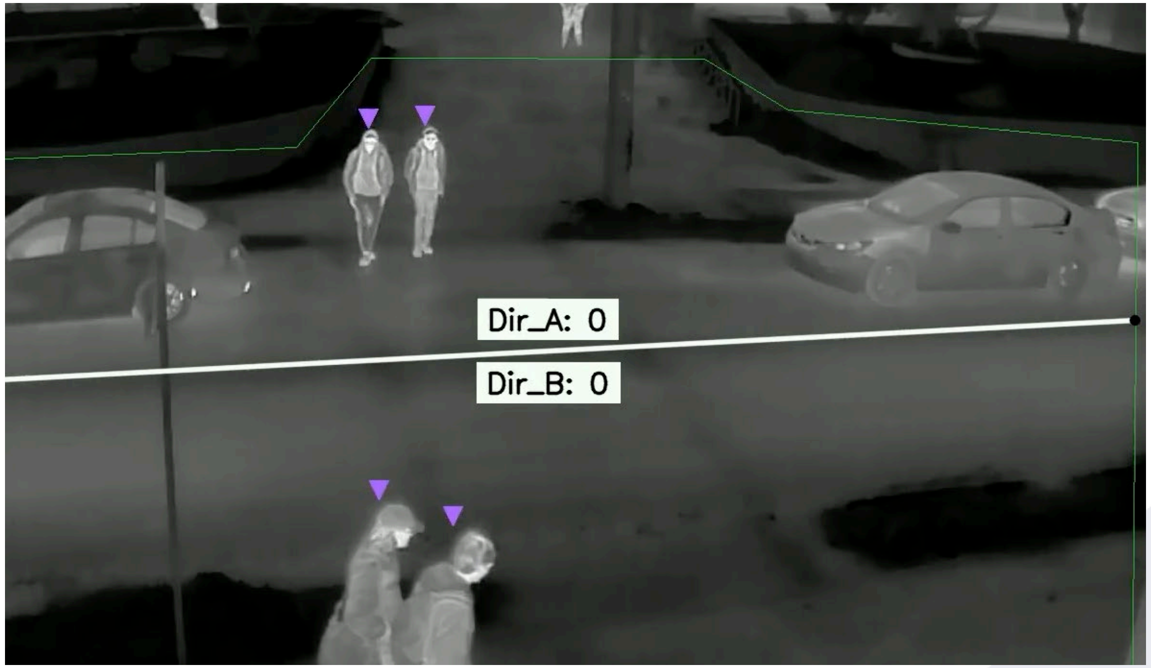
- **Improved Generalization**: Diverse locations and conditions help the model more accurately detect pedestrians across various real-world scenarios.

Sample images from our thermal dataset used to train the deep learning detection model.

Why Thermal?

- **Privacy-Preserving**: Converts people and vehicles into non-identifiable heat maps, ideal for privacy-sensitive monitoring (e.g., smart mobility, infrastructure).





Why Thermal?

- **See Beyond Light**: Captures scenes in complete darkness or glare, maintaining reliable object shape and size detection where RGB fails.

- **All-Weather Robustness**: Thermal imaging can retain >90% detection reliability in fog, rain, and low light, while RGB accuracy can drop >30% under the same conditions. [1]

- **Lightweight Processing**: Uses a single intensity channel, reducing bandwidth and compute demands- ideal for efficient edge inference.

Comparison of nighttime images: RGB vs. thermal.

[1] J.M.R. Velázquez, et al., Analysis of Thermal Imaging Performance under Extreme Foggy Conditions: Applications to Autonomous Driving, J. Imaging, Nov 2022.

Dataset for Pedestrian Detection

- **Environmental Diversity & Generalization**: Our base dataset spans urban parks, dense cityscapes, and tree-covered suburban roads.

- **Improved Generalization**: Diverse locations and conditions help the model more accurately detect pedestrians across various real-world scenarios.

Challenges:

- **Limited Training Data**: Annotating data for every new location or condition is costly.

- **Domain Shift**: When models trained in one location are deployed in a new environment, detection accuracy can drop sharply.

Sample images from our thermal dataset used to train the deep learning detection model.

Domain Adaptation

"Performance drop due to domain-shift is an endemic problem."

- **Domain Adaptation:** The scenario where model is initially trained with a dataset, which is usually not small, but its **distribution is different from the target environment** and is later updated with the collected data.
- **The "Edge Case" Problem**: Handling long-tailed and edge-case instances is a major challenge for deploying real-world computer vision models.

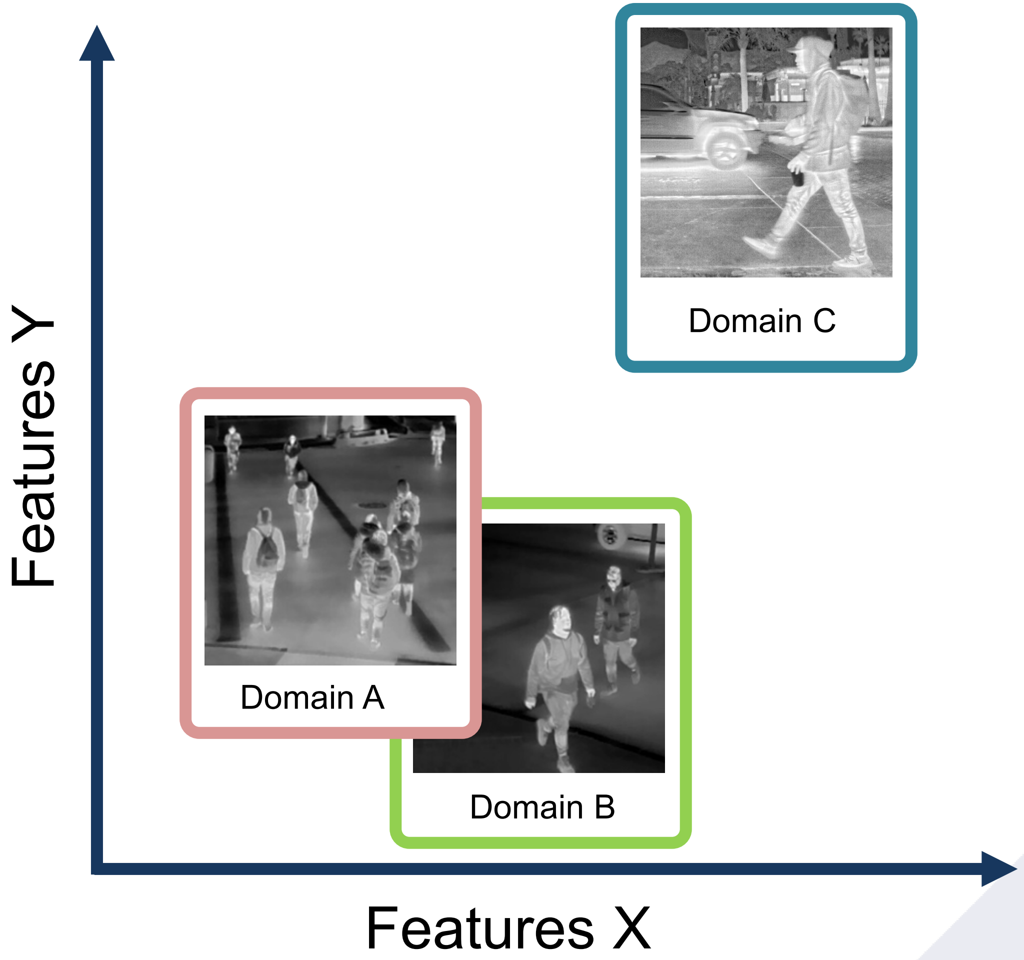
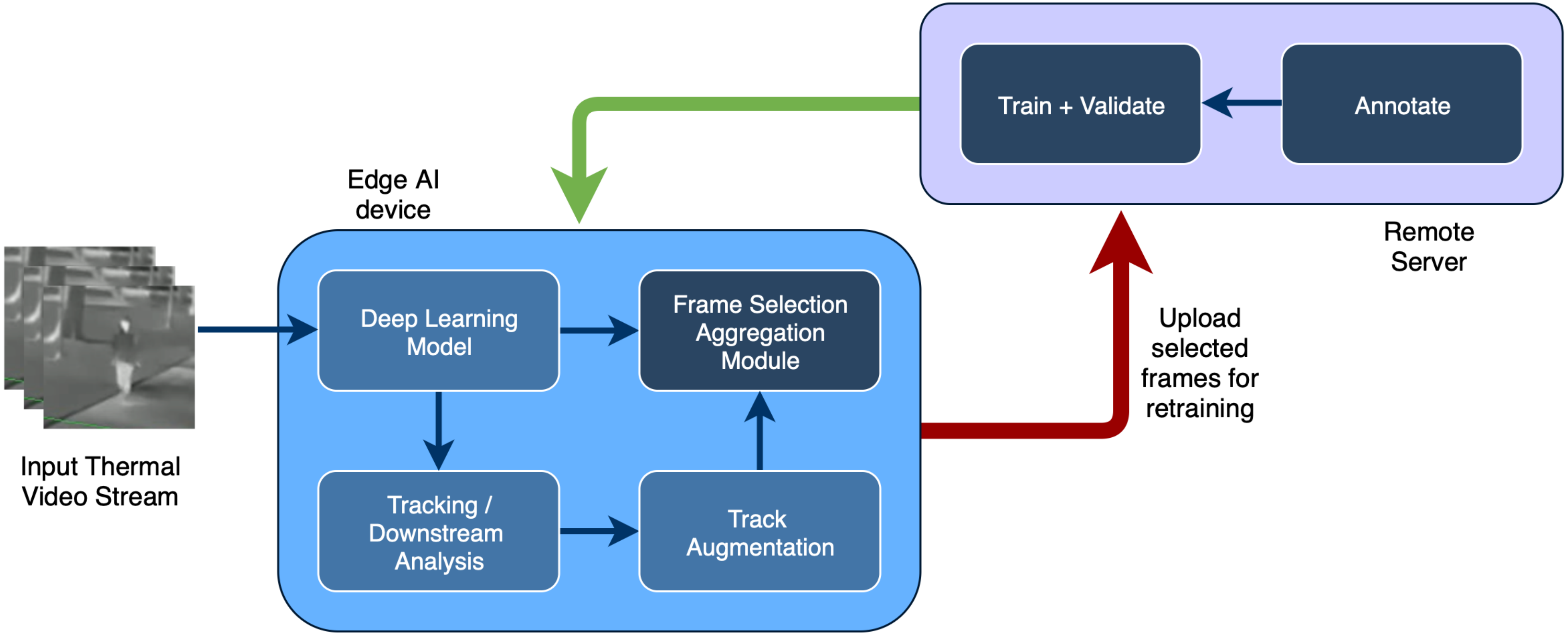
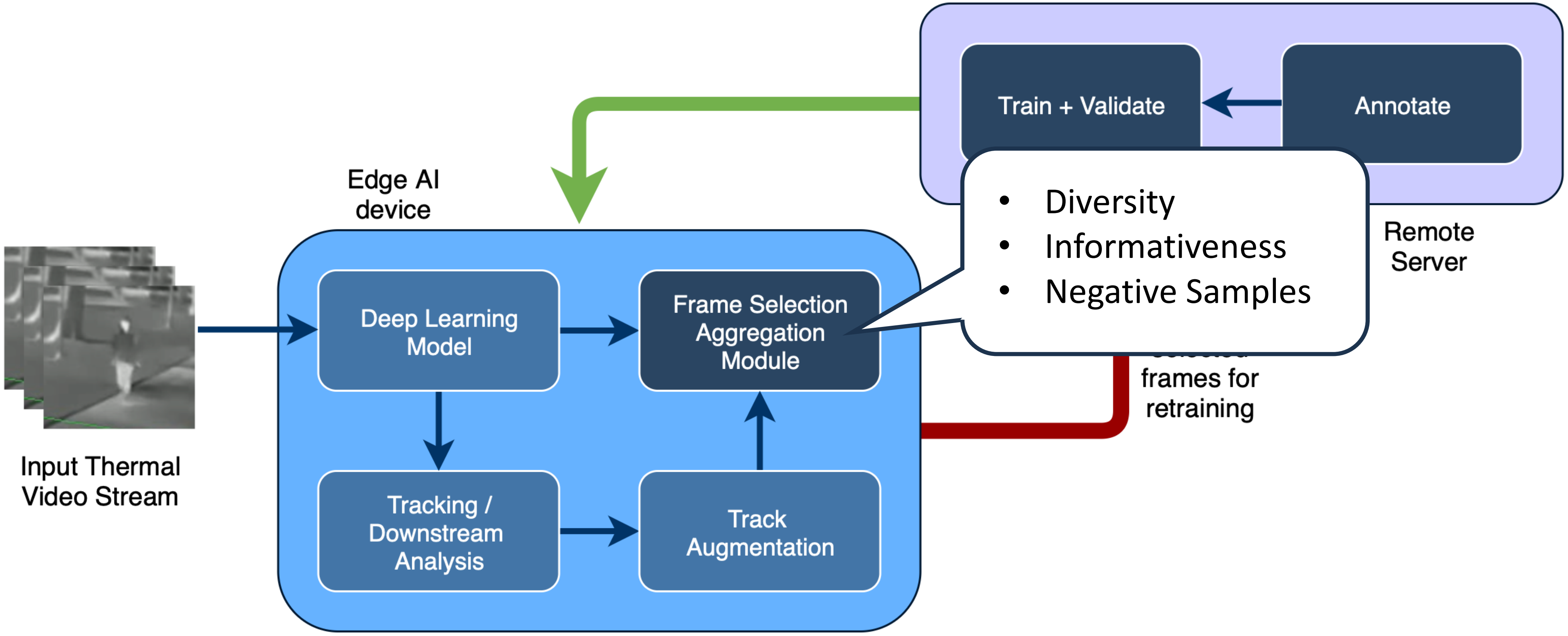


Illustration of a domain shift in feature space

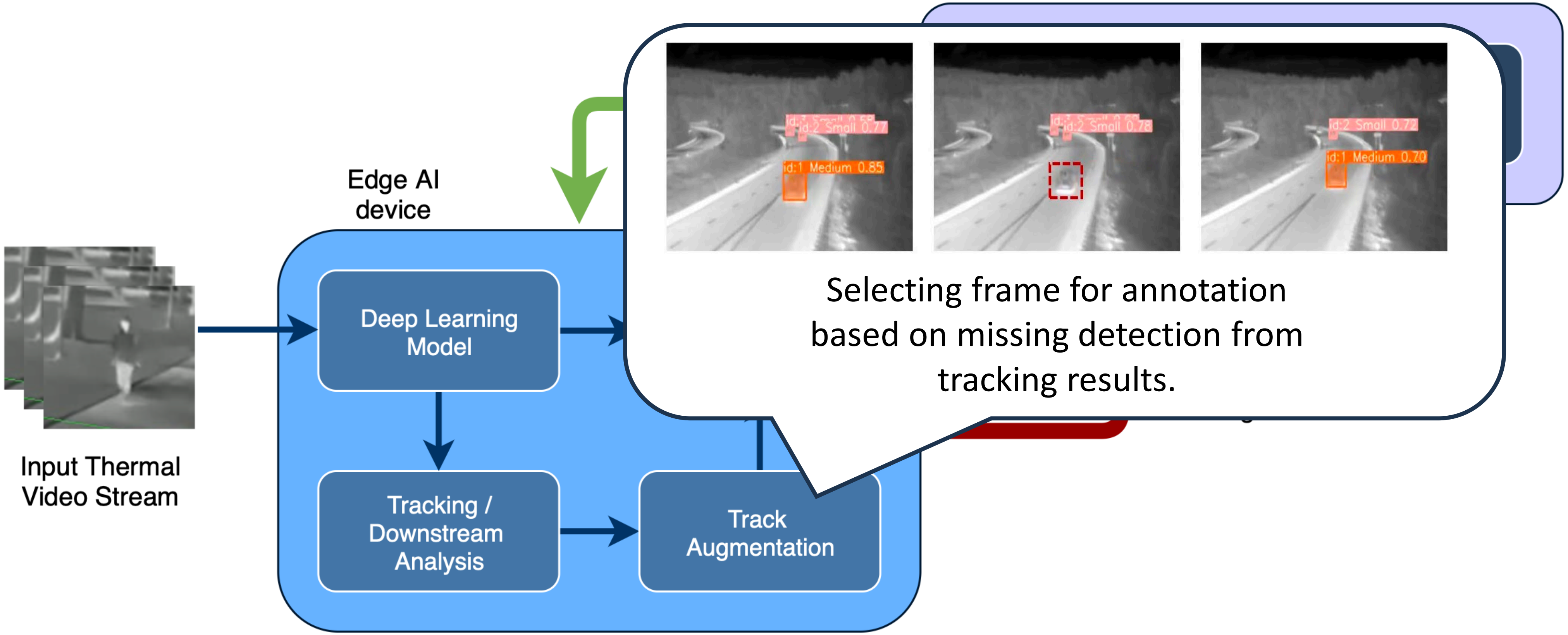
Submodular Active Learning Loop



Submodular Active Learning Loop



Submodular Active Learning Loop



Results

Every iteration of the Active Learning step selected 30-45 images from 4 sites.

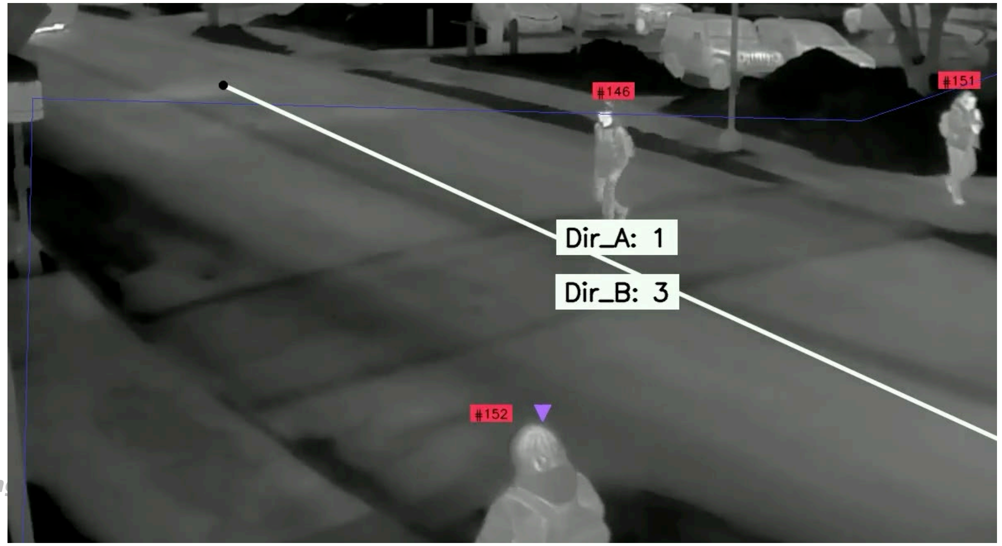
Ablation Results (Yv11-S Base model only)

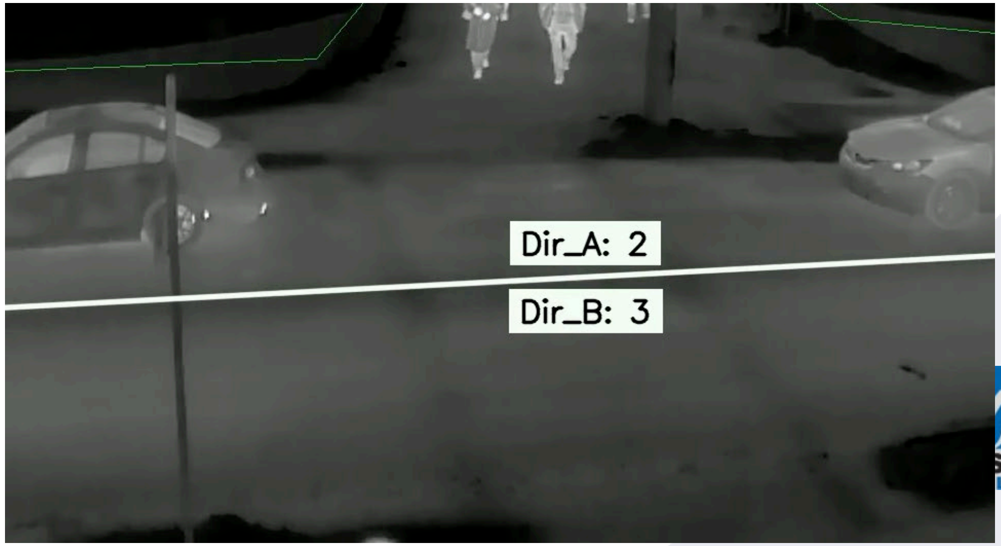
Model Variation	mAP50	mAP50-95
Baseline	77.6	61.4
+ cosine_lr_scheduler	79.9	62.7
+ label_smoothing	80.4	63.9
+ warmup_epochs	**82.4**	**64.4**

Model	Class	Precision	Recall	mAP50	mAP50-95
YOLOv11 (Base)	*Large*	*88.6*	*76.1*	*87.2*	*71.9*
	Small*	81.2	72.8	82.4	64.4
Active Learning (Yv11-small)	λ_1	84.7	74.5	85.3	66.3
	λ_2	86.2	75.7	86.6	68.1
	λ_3	**88.6**	76.0	**87.4**	71.4

Base model | After λ_{+3} (+ 135 img)

Visualization of Detection and Tracking





Conclusions

- **Improved model efficiency and robustness:** Implemented an on-device active learning framework that actively selects informative thermal frames for retraining.
- **~3x fewer parameters:** YOLOv11-small (~9.4M params) matches the pedestrian mAP@50 of YOLOv11-large (~25.3M params) after active sub-modular selection and server retraining.
- **Edge Deployment:** Deployed on Raspberry Pi and HAILO platforms with real-time inference, connected via 4G / LoRa mesh for distributed coordination.
- **Future Direction**: Explore federated or continual learning strategies for collaborative model improvement across devices.

Thank you!

This study was undertaken as part of a project funded by MassDOT. The authors are solely responsible for the facts, the accuracy of the data and analysis, and the views presented herein.

UMASS LOWELL