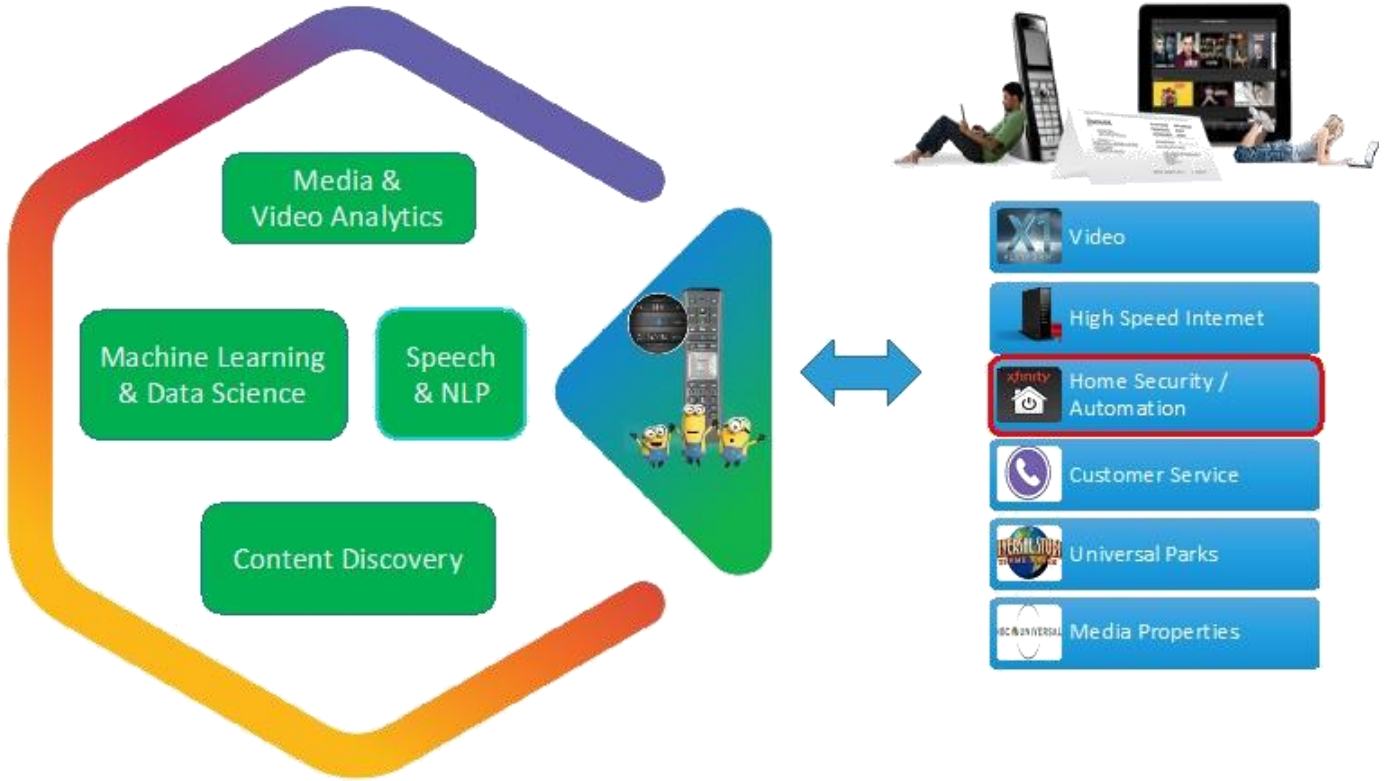
Architecting a Smart Home Monitoring System with Millions of Cameras

COMCAST

Hongcheng Wang

May 23, 2018

Comcast Applied AI Research

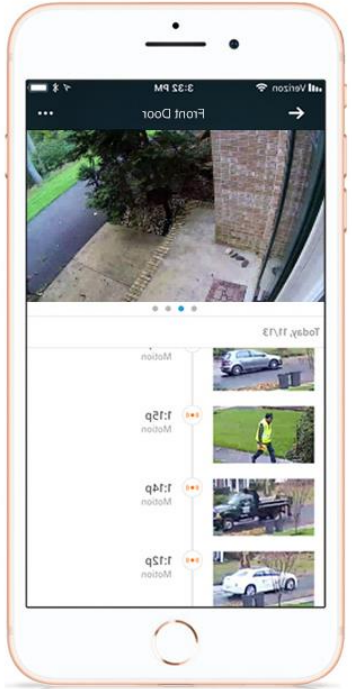


Xfinity Smart Home

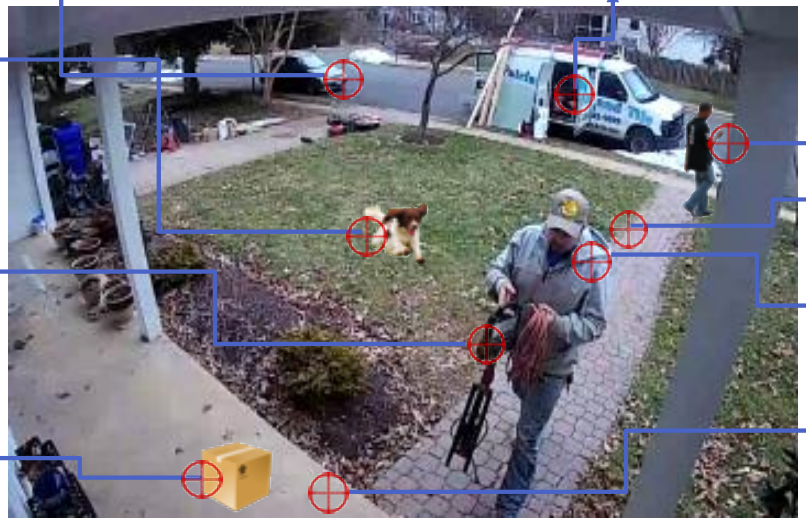
Comcast is becoming a leading platform to integrate smart home technologies with XCams, door/window sensor, WiFi, voice, remote control, etc.

1M+

Xfinity Home Subscribers



What Are Homeowners Interested In?



Whose car is coming to my driveway?

A truck stopped in front of my house. (I don't care about the cars passing by the street)

Is it my dog?

How long has this person been loitering around my home?

How long is my home service taking (mowing lawn, cleaning, repair)?

Is any family member at home?

Who is this person?

When was my package delivered?

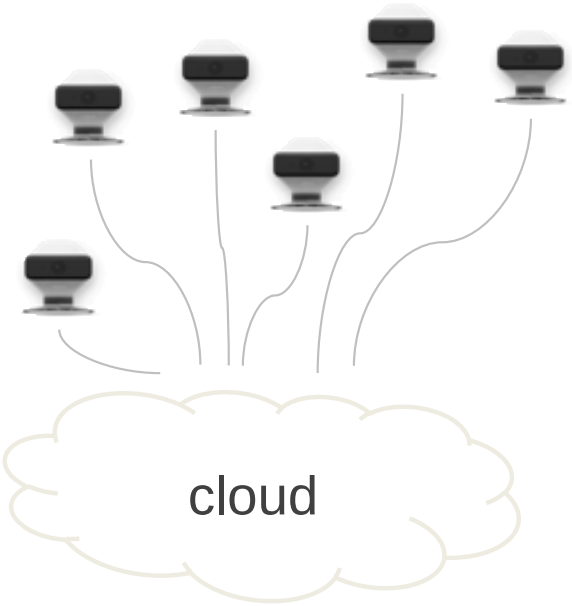
When were my kids back from school?

What We Need to Consider to Design a Hybrid System

- Edge vs. Cloud vs. Cost (Computing) vs. Accuracy (Algorithm)
  - Video resolution: HD(720p)/FHD(1080p)/UHD(4k)
  - Camera computing power: CPU/GPU
  - Cloud: CPU/GPU, storage, latency
  - Bandwidth: image/video/metadata
  - Algorithm: Deep learning or traditional; image quality vs. algorithm performance
- Optimize all these parameters to build a cost-effective system
  - Some parameters may be fixed in practice (e.g., camera)
  - General guideline is to push as much computation as possible to the edge

Edge vs. Cloud-based Analytics

Comcast Camera 720p, No GPU on camera



24/7 CVR (Continuous Video Recording)

Video analytics on the edge

- Limited computing but millions of cameras
- High quality metadata
- Reduced bandwidth for metadata

Deep learning algorithms on the cloud

- Cost effective CPU/GPU utilization
- Object detection
- Efficient motion event detection algorithm

Comcast Hybrid, Cost-Effective and Robust Analytics

On the edge: RDK-Camera

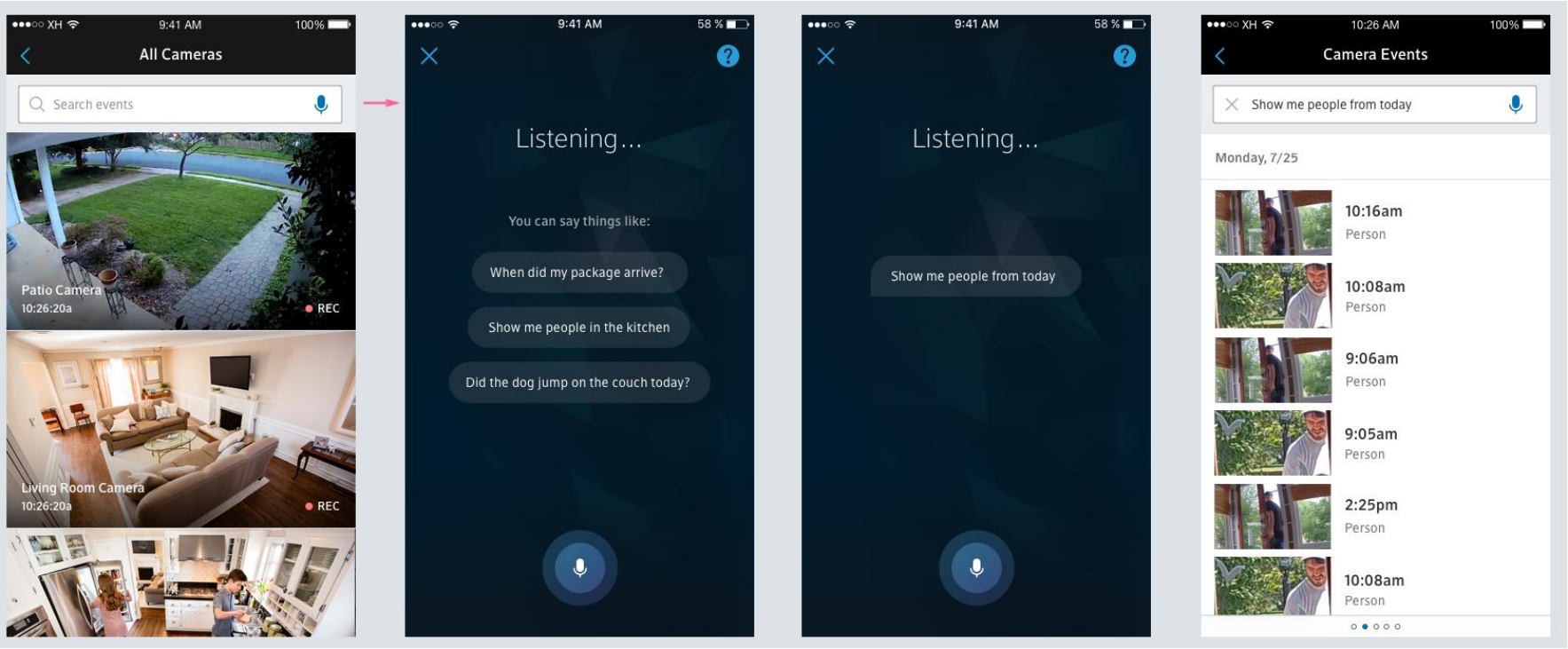
- Motion event detection
- Motion object tracking
- Smart thumbnail generation
- Face detection
- Audio analytics

In the cloud

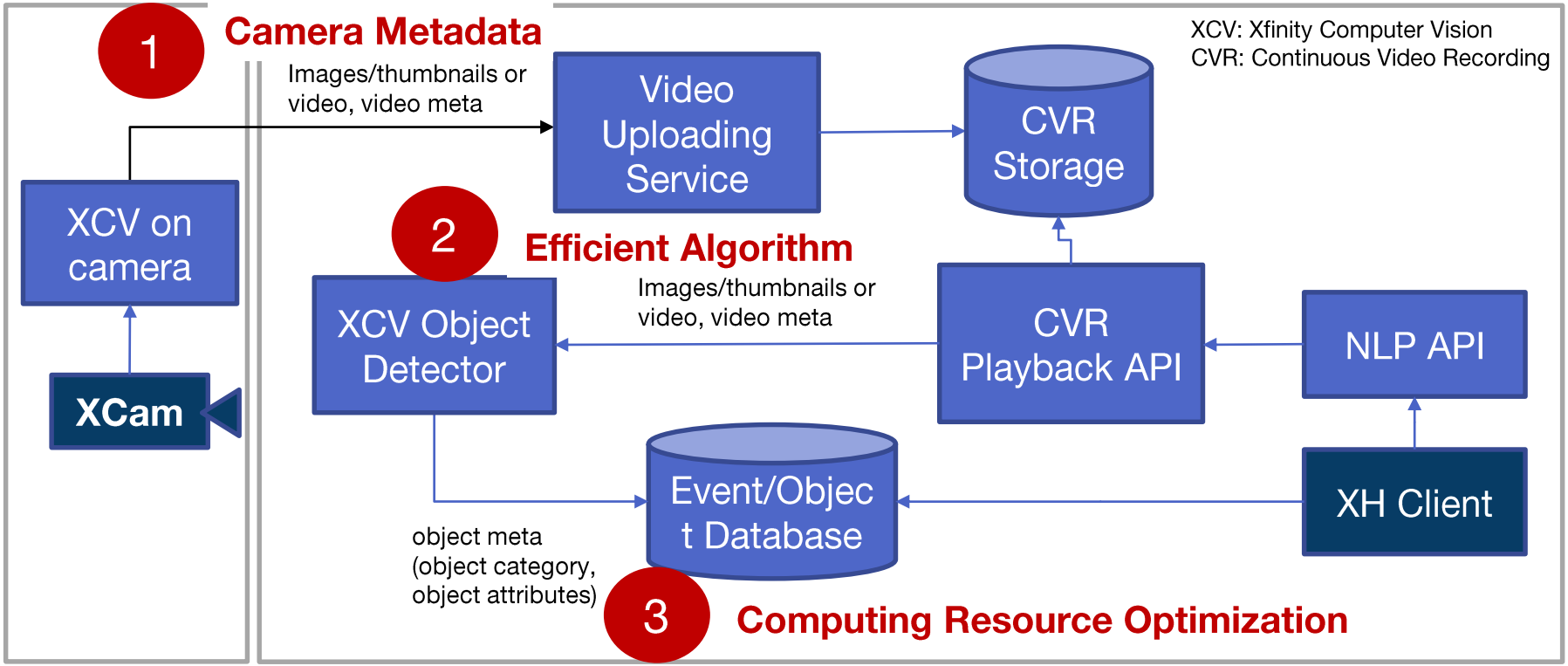
- Anomaly Detection
- Object Detection (Person/Vehicle/Pets)
- Summary of My Day
- Face Recognition
- Event/Activity Detection
- Semantic Video Search

Video → Motion detection and preliminary object detection → Advanced Video Analytics → Relevant Events

Use Case: Semantic Video Search

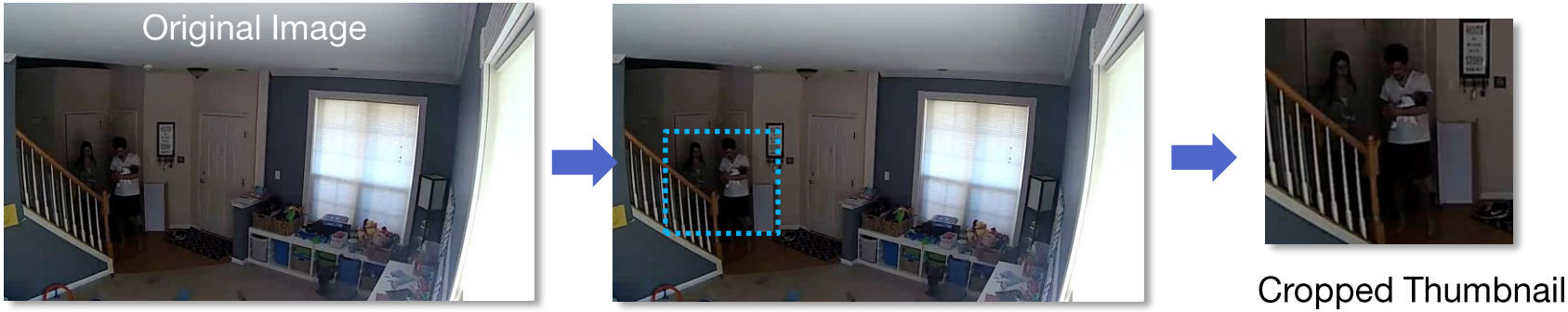


Hybrid Architecture



What Metadata are Extracted from Camera?

- Video resolution is limited in surveillance camera, and video compression is usually high to save bandwidth
- Cropped high resolution, large-size thumbnails from camera with multi-buffer parallel processing
- Other metadata include: track id, motion score, time stamps for selected frames



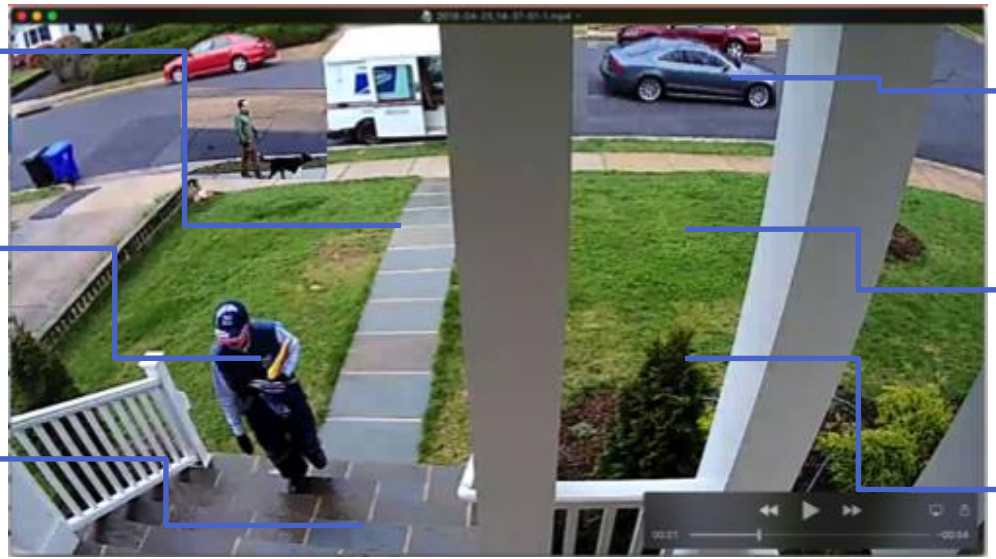
Relevancy Based Evaluation Metrics

- Event relevancy from 0 to 5

Relevancy 3:
Significant objects in a relevant area

Relevancy 4:
Significant objects entering/existing a relevant area

Relevancy 5:
Events worthy of an unprovoked notification (fire, burglary, etc.)

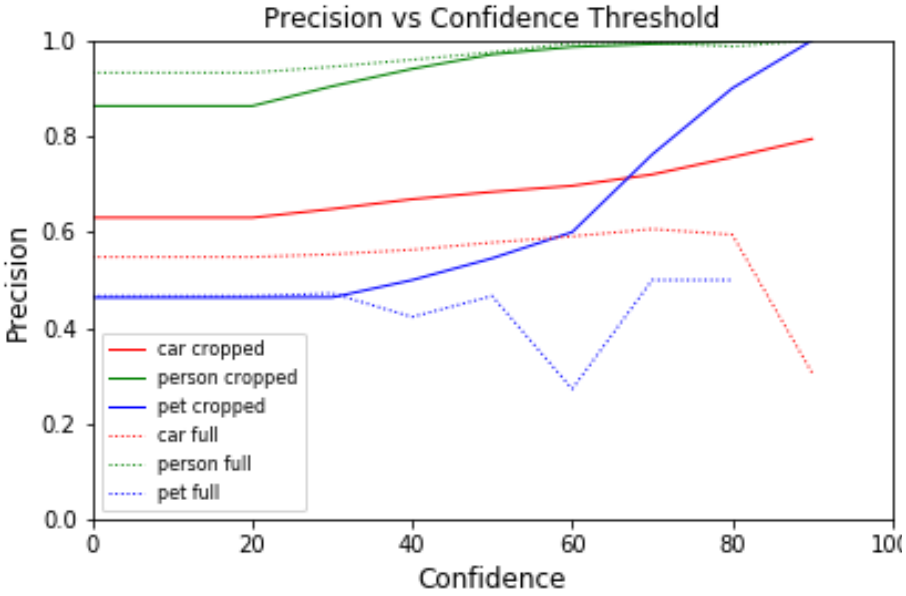


Relevancy 2:
Significant objects outside of a relevant area

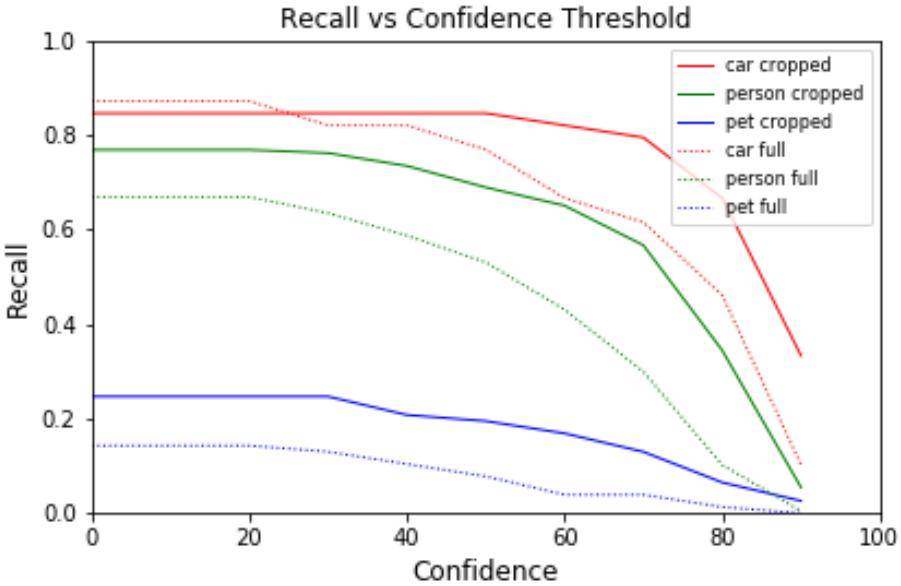
Relevancy 1:
Insignificant object (e.g., bugs, squirrels)

Relevancy 0:
Rain, wind, shadows

Thumbnail Cropping Improves Object Detector



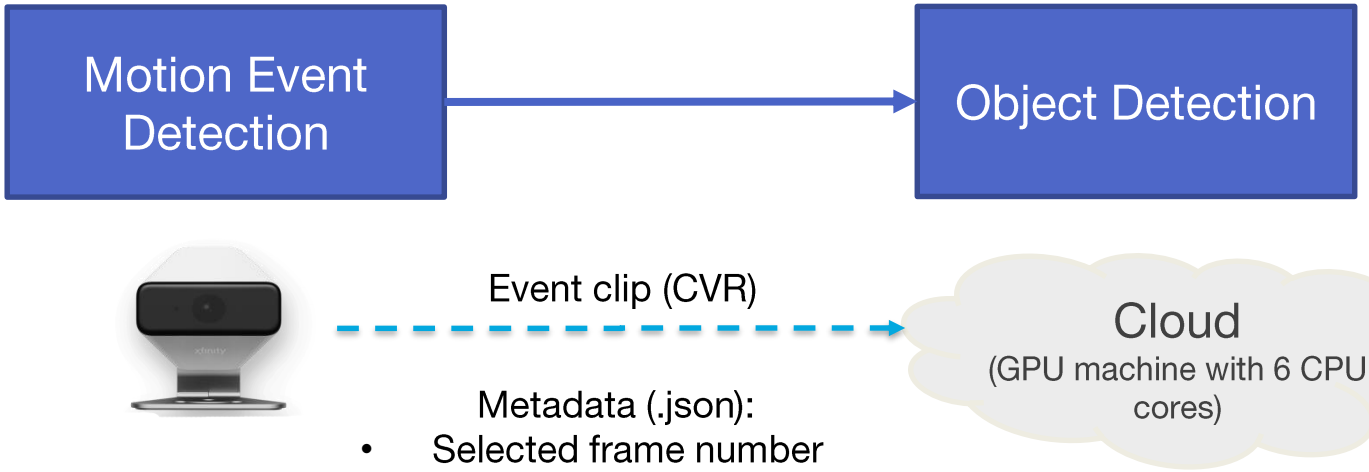
Relevancy >= 1



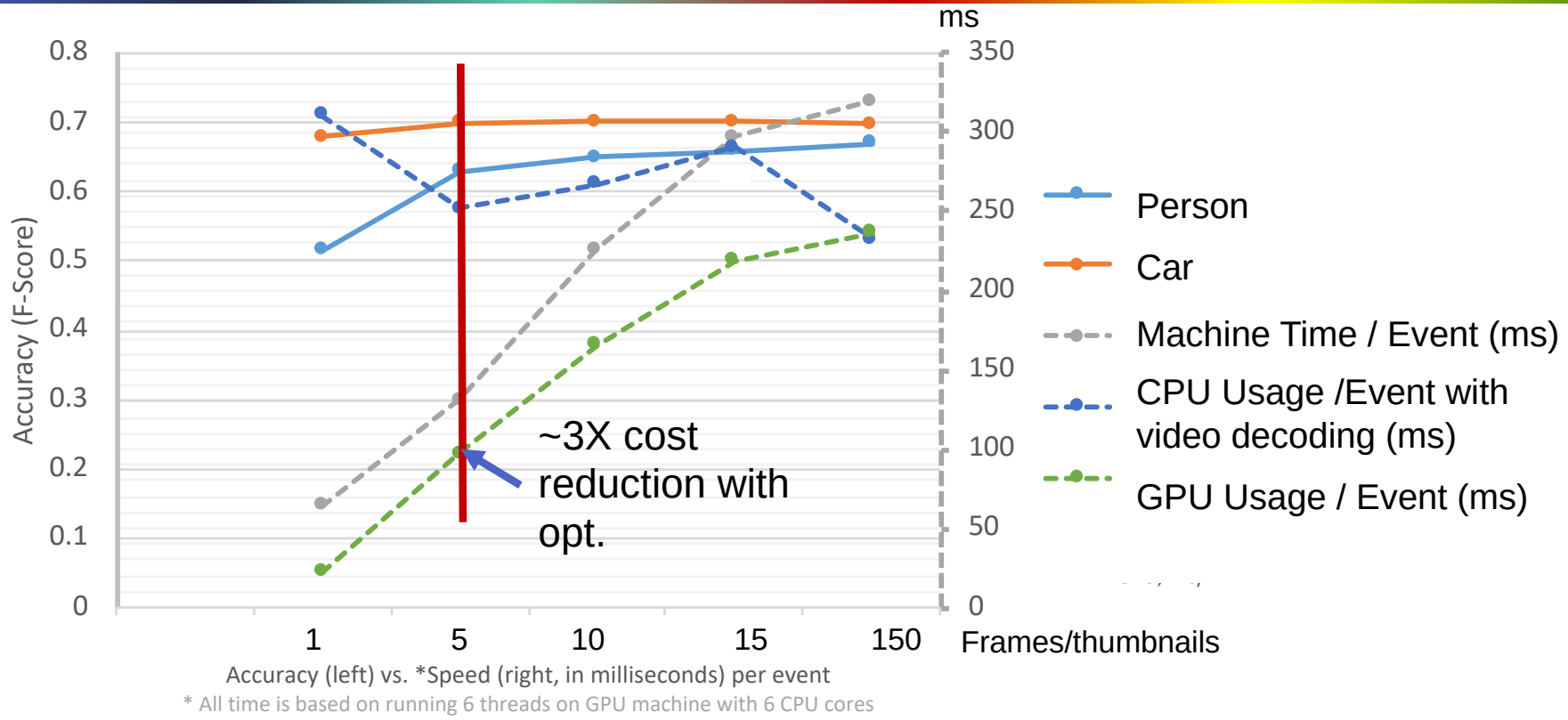
Relevancy >= 4

Comcast dataset consists of 8111 videos

CPU/GPU Optimization



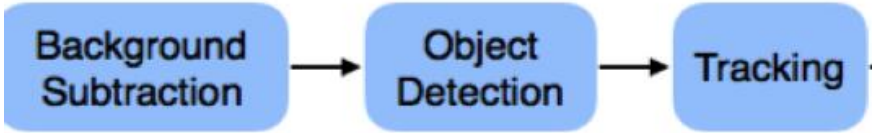
CPU/GPU Optimization Reduces Cost



Accuracy (left) vs. *Speed (right, in milliseconds) per event

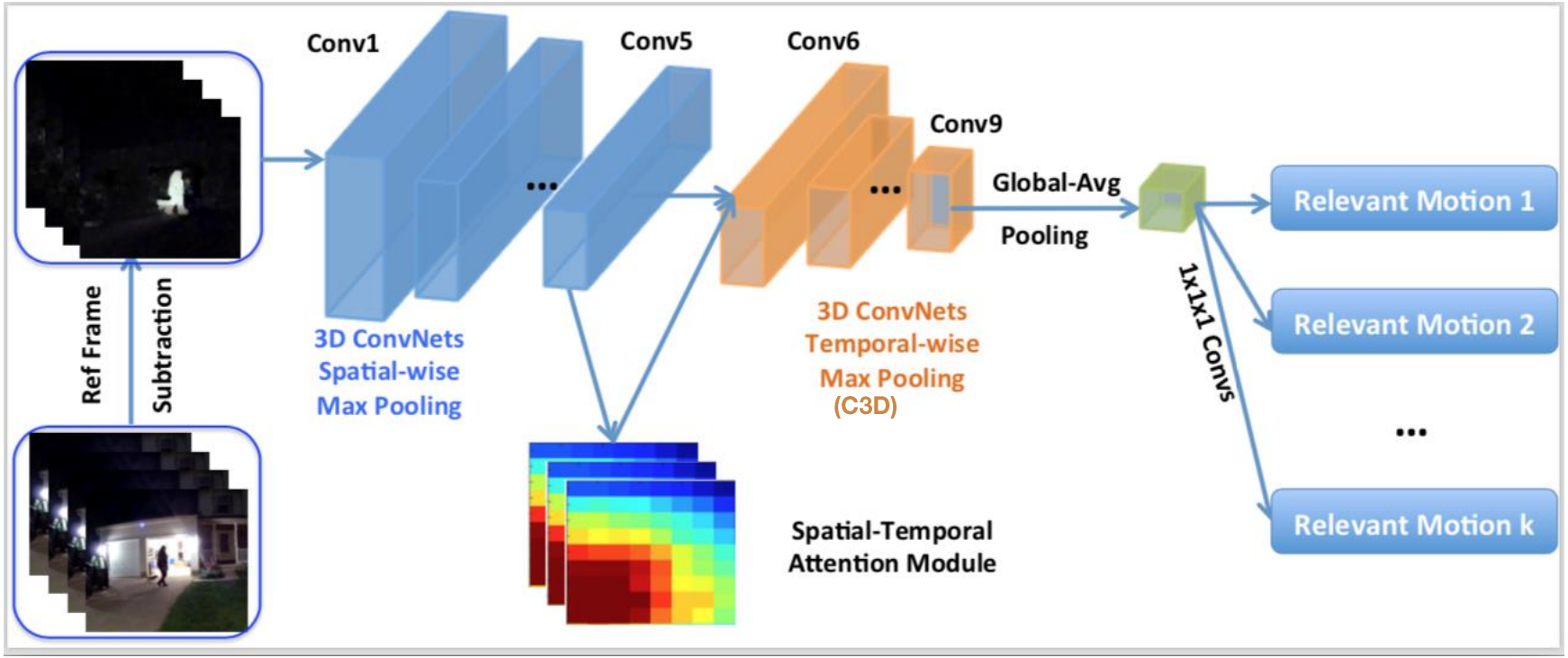
* All time is based on running 6 threads on GPU machine with 6 CPU cores

Issues With the Detection-Based Approach



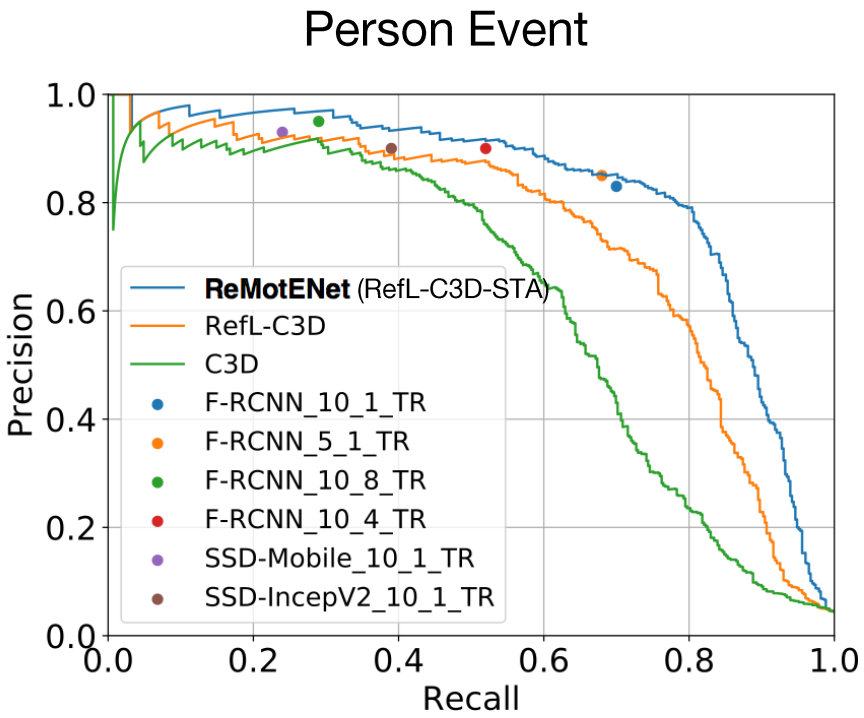
- Not efficient
  - Brute force: takes 15 - 75s / 15s video clip, 100 - 400 MB model
  - Optimized: ~100 ms/event
- Not robust
  - False detections and missed detections
- Parameters tuning
  - At least one parameter is associated with each stage of the pipeline

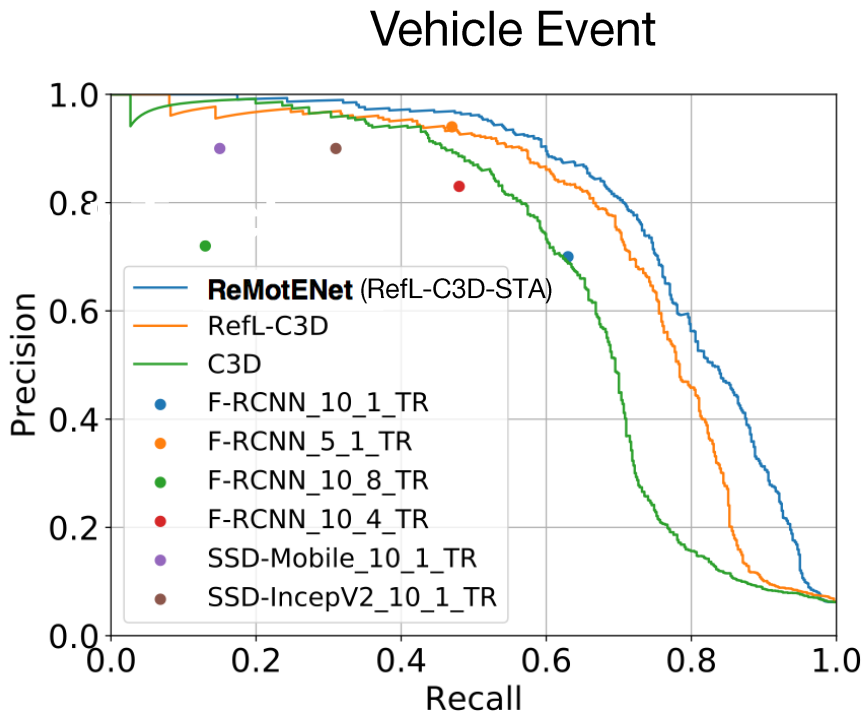
ReMotENet: A Relevant Motion Detection Network



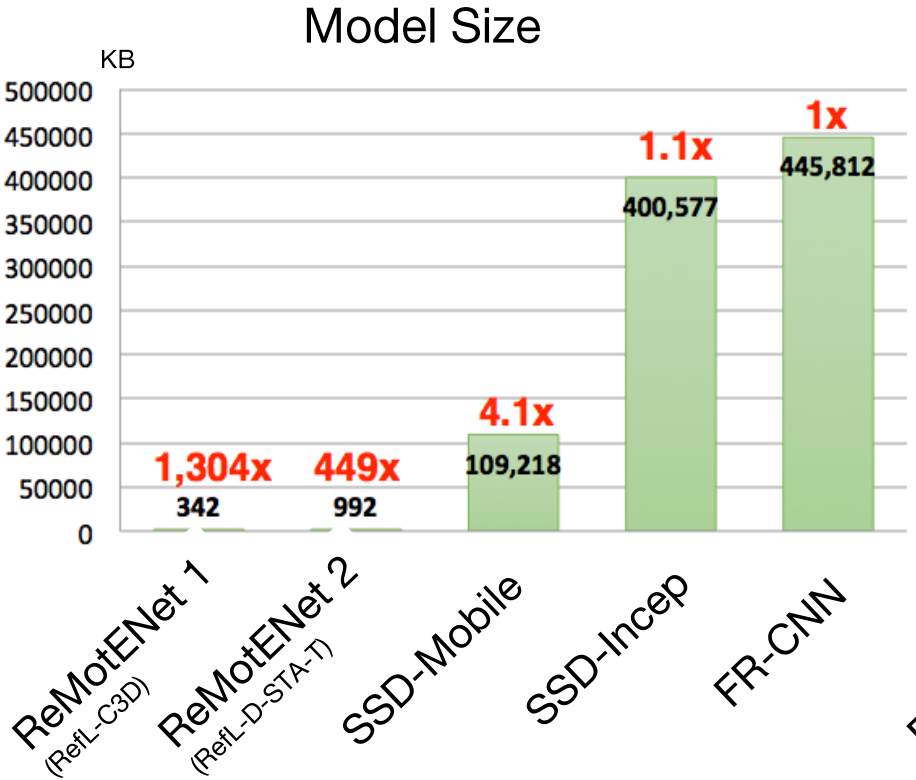
Ruichi Yu, Hongcheng Wang, Larry S. Davis, **ReMotENet: Efficient Relevant Motion Event Detection for Large-scale Home Surveillance Videos, IEEE WACV 2018**

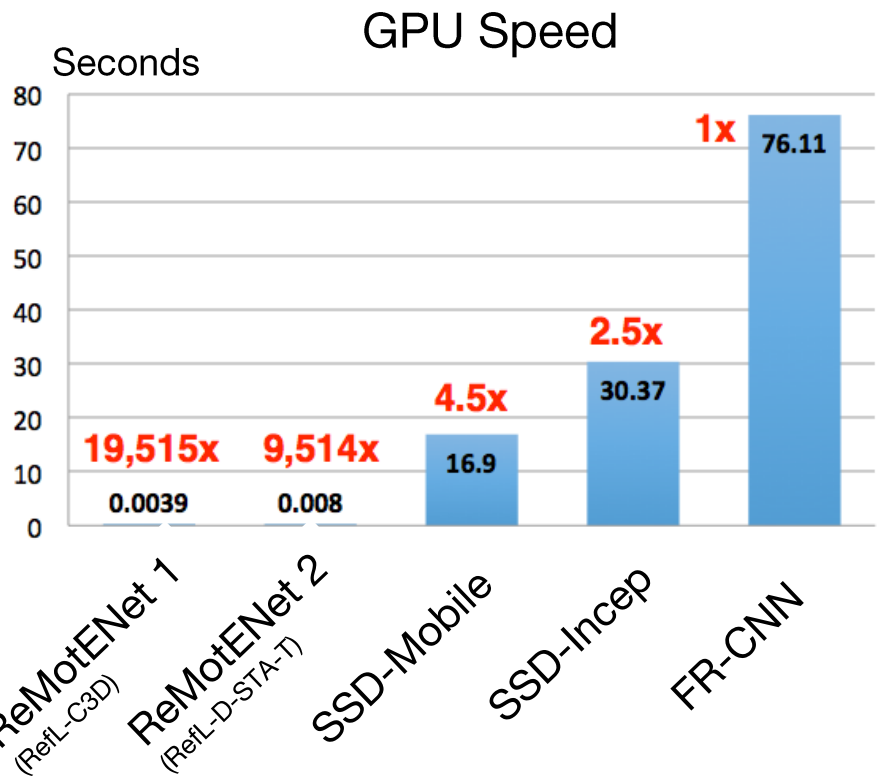
ReMotENet Improves Relevant Event Detection





ReMotENet Has Smaller Model Size and Faster Speed





Currently running on cloud, but has the potential for edge implementation

ReMotENet: Some Results

- Failure cases by the traditional object detection-based method

Falsely detected car event

Missed-the-person event

- Failure cases of ReMotENet

Missed-the-person event

Falsely detected person event

Summary

- Cost-Effective, Robust, and Scalable Analytics are critical to deploy deep learning video analytics algorithms for millions of cameras
- A cost-aware and robustness-aware hybrid architecture has been implemented
  - On-edge computing with relevant and high-quality metadata output
  - GPU/CPU optimization on the cloud
  - Efficient and compact deep learning model on the cloud (Implementation on edge in progress)

We Are Hiring!

Senior/Lead/Principal Computer Vision and Deep Learning Researchers

More information:

Applied AI group: http://dclabs.comcast.com/jobs

or email us at: **applied-ai-jobs@comcast.com**

BACKUP

Computer Vision @ Comcast Applied AI Research

Smart Home (IoT, xFi)

Smart TV

Computer Vision + Deep Learning

+

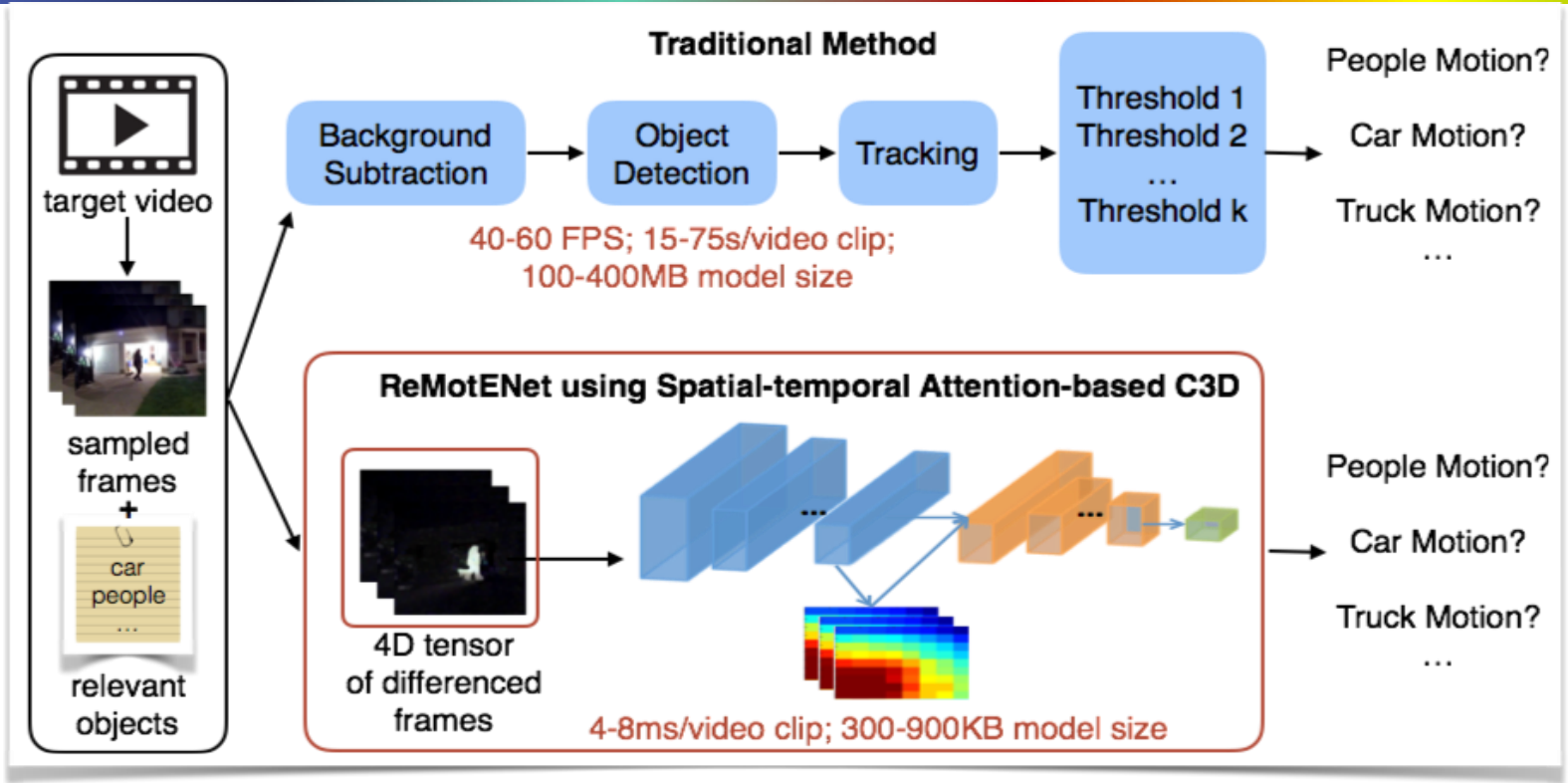
- Big Data: Sensor & Trickplay Data (Motion, door, window, wifi, mobile, etc.)
- Audio: Voice / Audio Analytics
- Text: NLP, CC / Semantic Query

ReMotENet: Some Results

- Failure cases by the traditional object-detection-based method

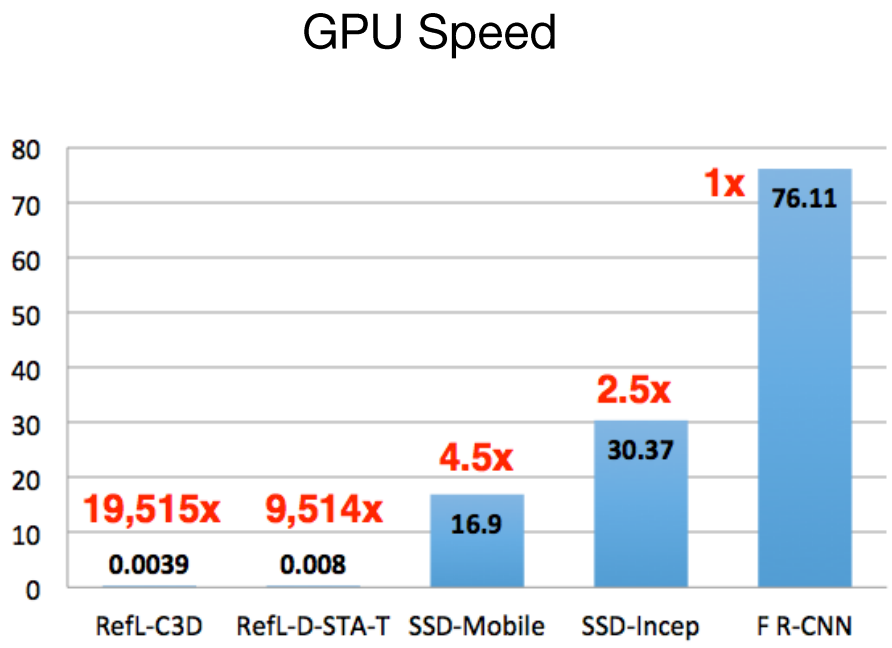
- Failure cases of ReMotENet

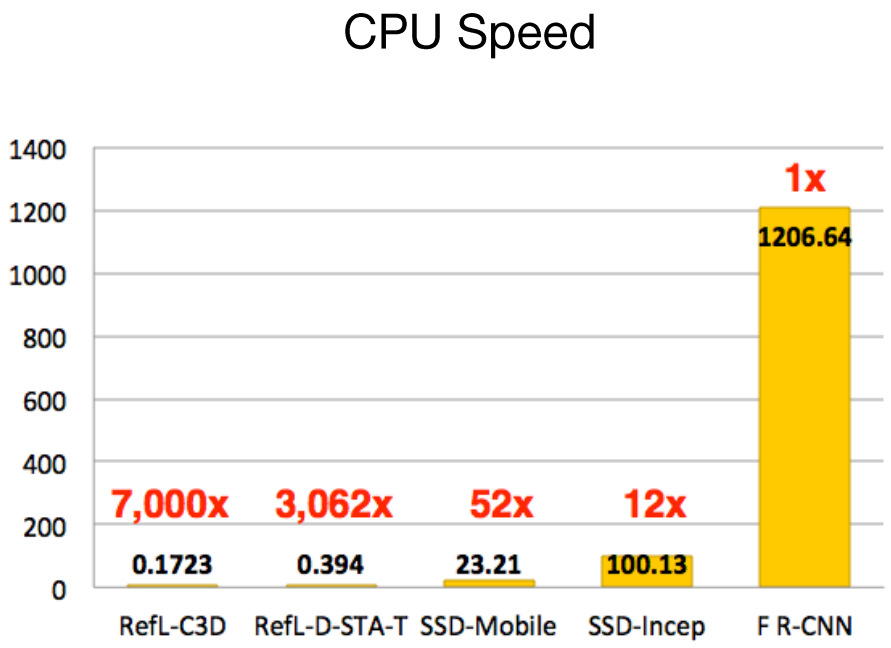
ReMotENet: A Relevant Motion Detection Network



Ruichi Yu, Hongcheng Wang, Larry S. Davis, **ReMotENet: Efficient Relevant Motion Event Detectio for Large-scale Home Surveillance Videos, IEEE WACV 2018**

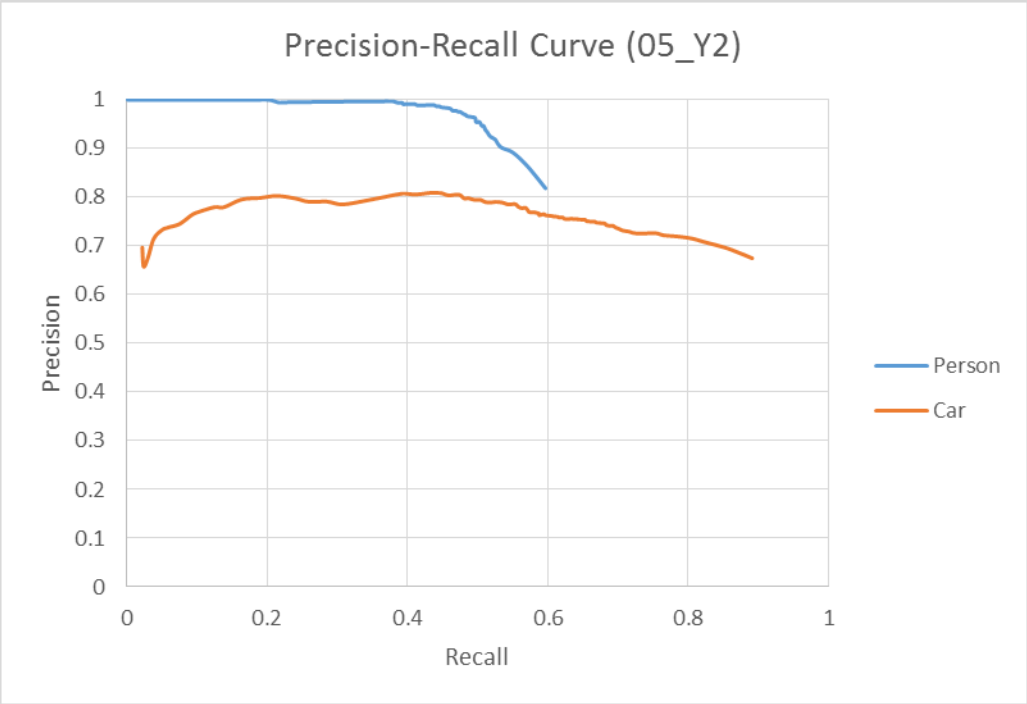
ReMotENet: A Relevant Motion Detection Network





seconds

Precision-Recall Curve



- Person has high precision and low recall
  - Many cases of person standing/sitting there but not moving much, which leads to more missed detections (lower recall)
- Car has lower precision and high recall
  - Car is usually larger
  - Other moving objects may classify stationary car as moving, which leads to more false detections (lower precision)

Architecture Details