360-Degree Video Analytics Framework for Firefighters Situational Awareness

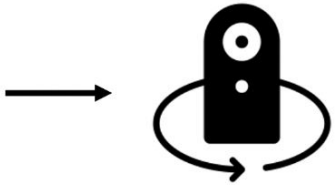
Jiaxi Li[1], Jingwei Liao[2], Aditi Tiwari[1]

[1] ILLINOIS

[2] GEORGE MASON UNIVERSITY

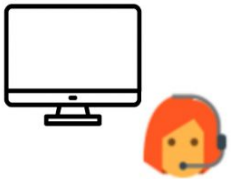
Framework Overview

Training Field



360-Degree Camera

Detection Model



Instructor

Objects & Actions

Object of Interest	Priority
Civilian	High
Fire	High
Smoke	High
Gas Mask	High
Firefighter	Low
Helmet	Low

Data based on an interview with a physical training instructor at Illinois Fire Service Institute (IFSI)

Example Actions

Viewing and Query Service

Video Controller

Start | Pause | Volume+ | Volume-

Select Video

Daylight

Labeling Object & Action

- [x] Firefighter
- [ ] Civilian
- [x] Ladder
- [ ] Oxygen Tank
- [ ] Firefighter Helmet
- [ ] Civilian Car
- [ ] Stairs
- [ ] Firefighter Mask

Object & Action Search

Firefighter Mask

Found 28 occurrences of Firefighter Mask

0.4 1.4 2.2 2.4 3.6 3.8 4.6 5
5.6 6.6 6.8 7.6 8.6 9.8 11.6 13.4
23 24.4 24.6 25 25.8 26.2 26.4
26.8 27.6 28 28.2 29.8

Background: 360-Degree Video

- What is 360-Degree video? What's the difference between normal 2D video and 360-Degree video? Why do we need 360-Degree video?

2D Video

360-Degree Video

Background: 360-Degree Video

- What is 360-Degree video? What's the difference between normal 2D video and 360-Degree video? Why do we need 360-Degree video?

2D Video

40°-70° field of view❌

360-Degree Video

360° field of view✅

Background: 360-Degree Video

- What is 360-Degree video? What's the difference between normal 2D video and 360-Degree video? Why do we need 360-Degree video?

2D Video

40°-70° field of view❌

Partial environment ❌

360-Degree Video

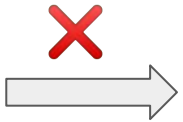
360° field of view✅

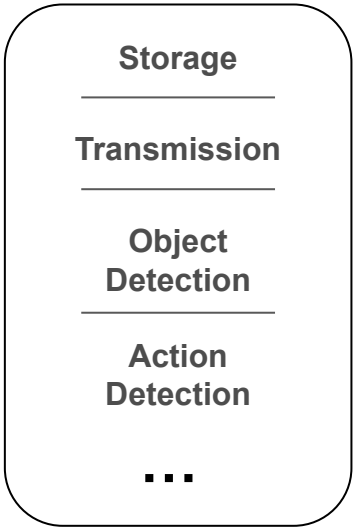
Entire environment ✅

Background: 360-Degree Video

- Preprocessing

360-Degree Video





Downstream Tasks

Most downstream tasks are not directly compatible with videos in 360-degree format

Background: 360-Degree Video

- Preprocessing

360-Degree Video

Project

Equirectangular Projection (ERP)

Storage

Transmission

Object Detection

Action Detection

...

Downstream Tasks

Background: 360-Degree Video

- Preprocessing

360-Degree Video

Project

Equirectangular Projection (ERP)

Storage ✅

Transmission ✅

Object Detection

Action Detection

...

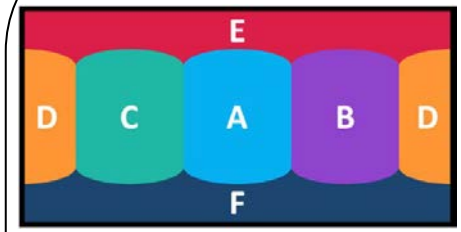
Downstream Tasks

Background: 360-Degree Video

- Preprocessing

360-Degree Video

Project



Equirectangular Projection (ERP)

Cubemap Projection (CMP)

...

Projection Methods

Storage

Transmission

Object Detection

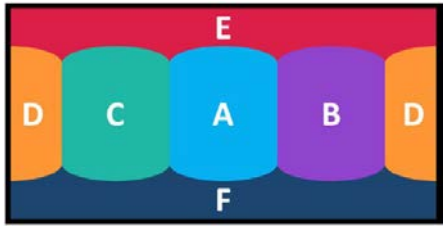
Motion Detection

...

Downstream Tasks

Background: 360-Degree Video

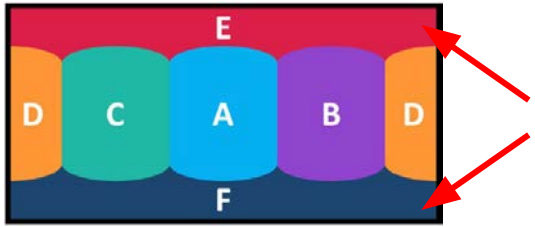
- Preprocessing



ERP

Background: 360-Degree Video

- Preprocessing

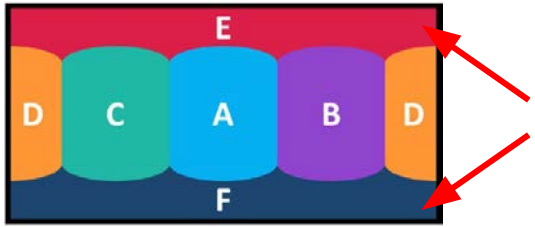


ERP

Distortion

Background: 360-Degree Video

- Preprocessing



ERP

Distortion

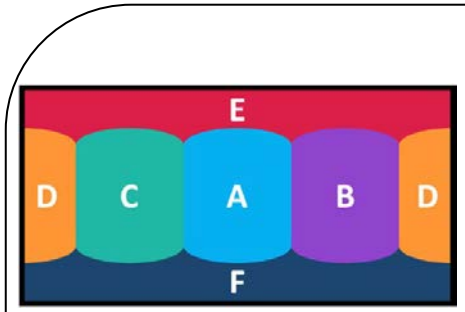
The straight line was distorted into a curve line

Background: 360-Degree Video

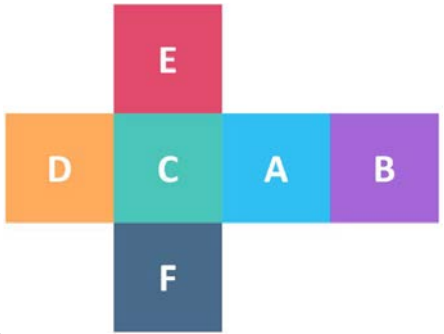
- Preprocessing

360-Degree Video

Project



Equirectangular Projection (ERP) ✅



Cubemap Projection (CMP)

...

Projection Methods

Storage ✅

Transmission ✅

Object Detection

Motion Detection

...

Downstream Tasks

Background: Object Detection

- Preparing the training and testing dataset
- Labeling these data
- Training and testing the task model

Background: Object Detection

- Preparing the training and testing dataset

We collected 25 360-degree videos at a IFSI

Outdoor Night

Outdoor Day

Indoor

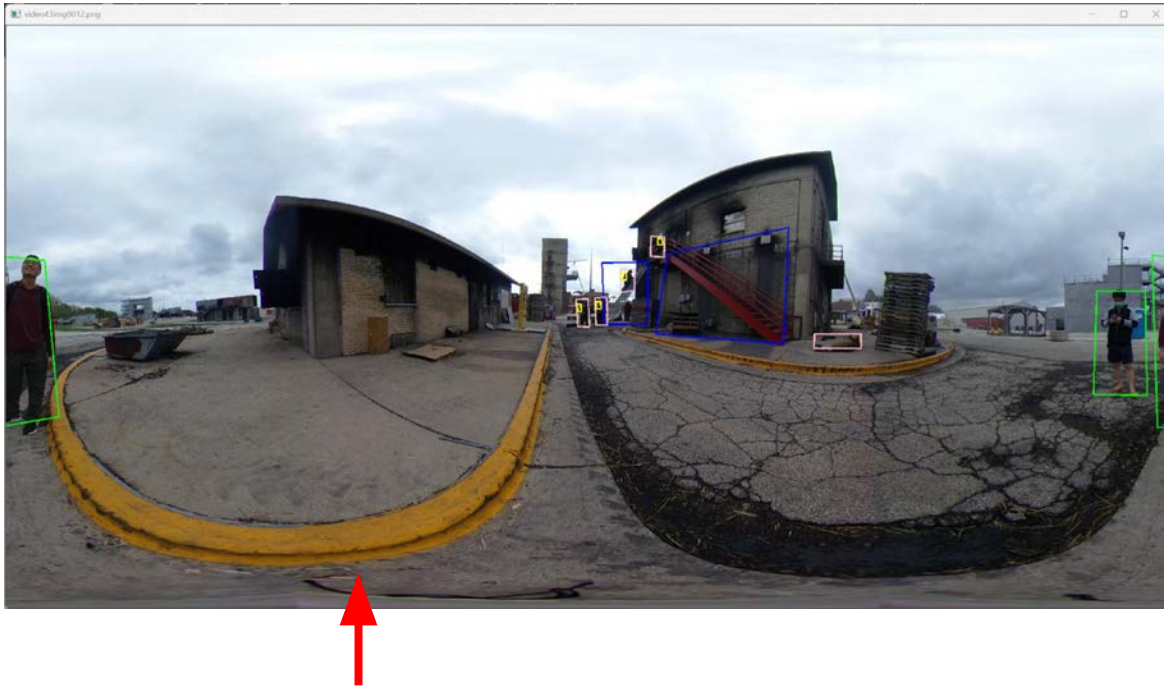
Background: Object Detection

- Labeling these dataset

The distortion will distort the geometry shape of an object

Background: Object Detection

- Labeling these dataset



The distortion will distort the geometry shape of an object

The straight line was distorted into a curve line

Background: Object Detection

- Labeling these dataset

An object annotation tool specially designed for 360-degree images in ERP format

Background: Object Detection

- Model Training

One stage object detector(YOLO Detector)

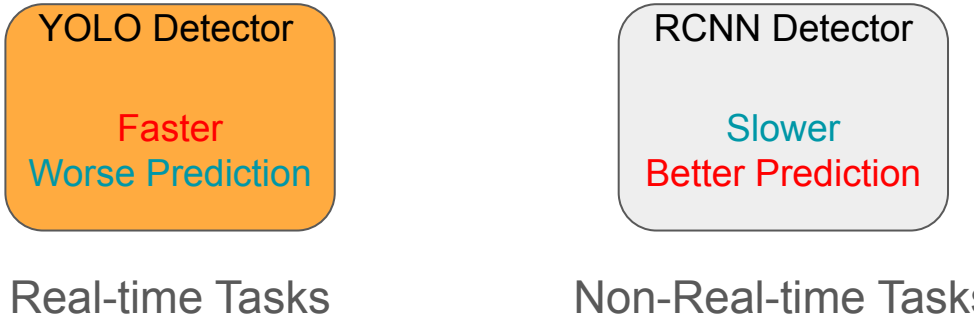
Two stage object detector(RCNN Detector)

Background: Object Detection

- Model Training

One stage object detector(YOLO Detector)

Two stage object detector(RCNN Detector)



360 Video Object Detection via 2D Object Detectors

Geometric Distortion

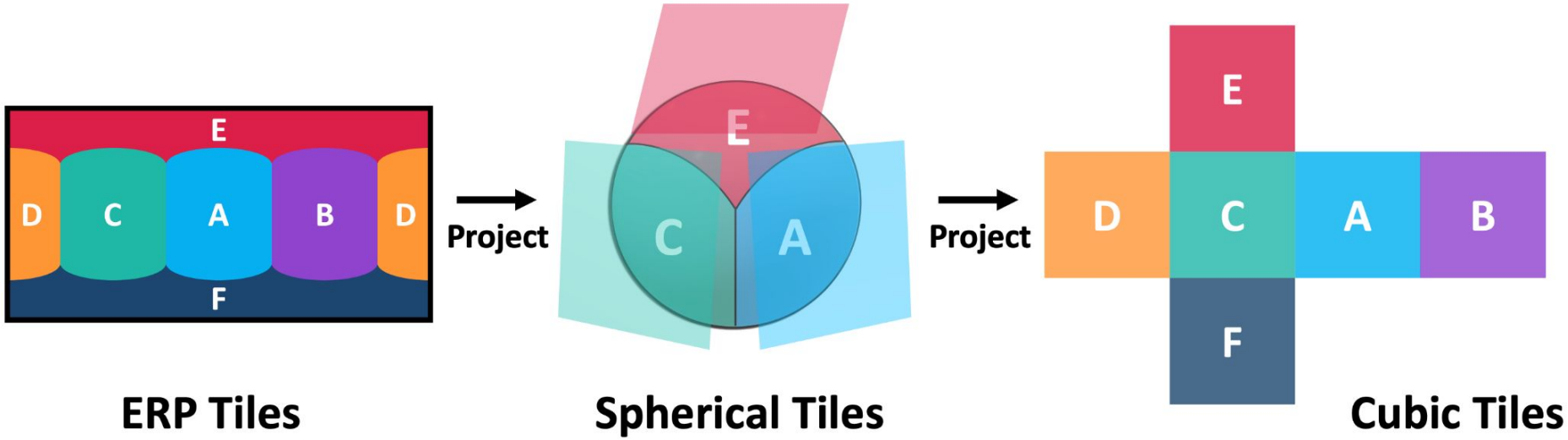
(Detection Error)

(Missed Detection)

Frame Size

- 4 times the number of pixels of a normal 1080p video

Dual-Projection Solution for 2D Object Detector Issues



Normal field of view

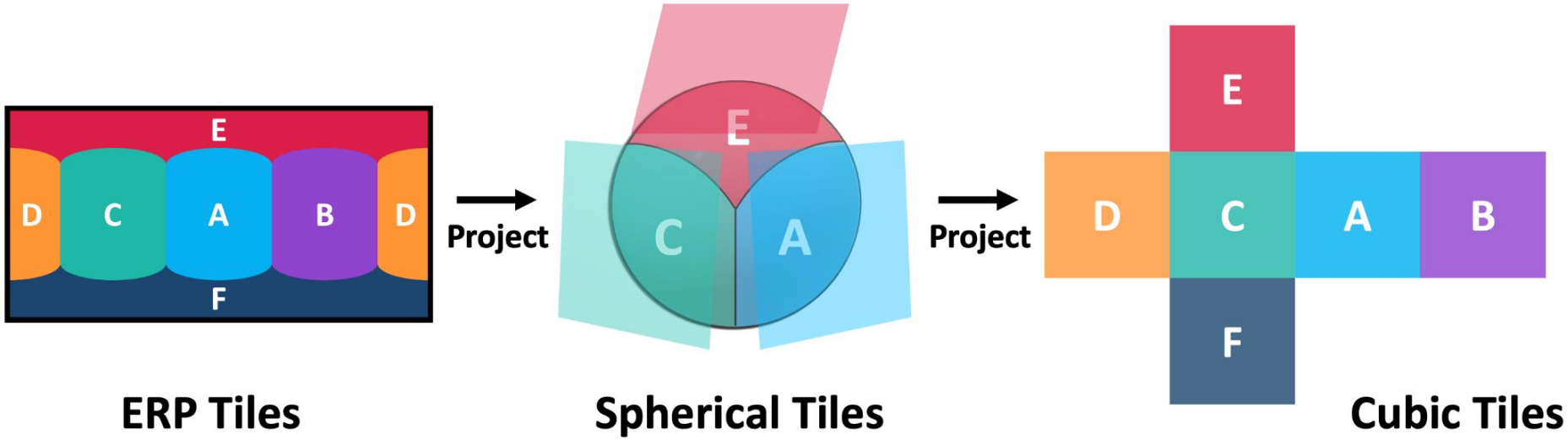
Dual-Projection Examples

ERP Tiles

Dual-Projection

Cubic Tiles

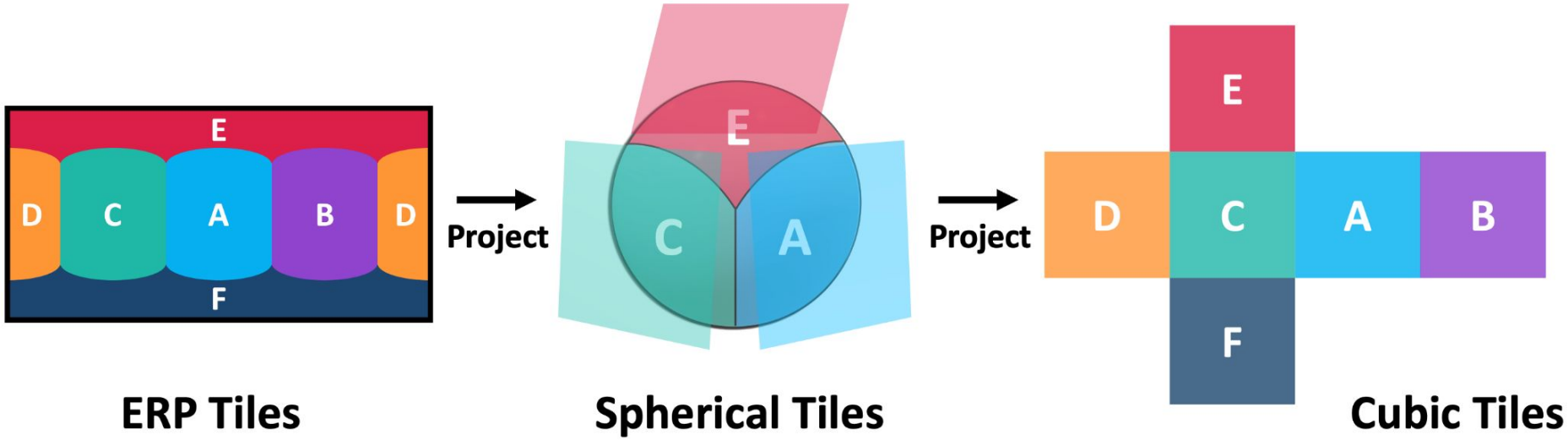
Dual-Projection Limitations



Leads to extra processing time in transforming ERP tiles to cubic tiles

E.g., >30% extra time on a 4-core CPU

Dual-Projection



Leads to **extra processing time in transforming ERP tiles to cubic tiles**

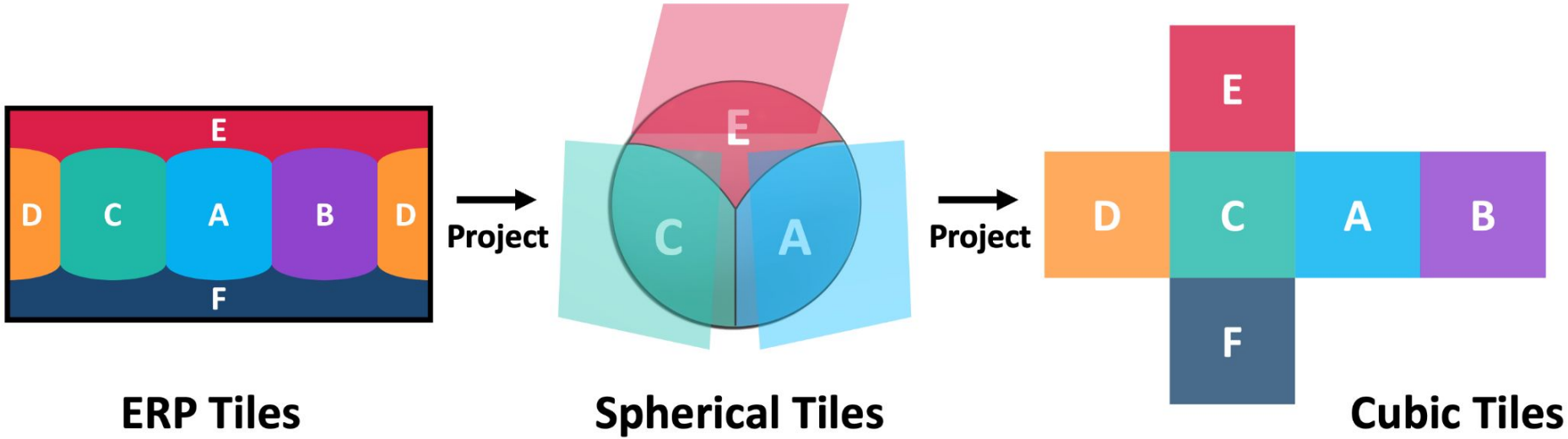
Dynamic Selection: **filters out unnecessary projection and detection.**

Dynamic Selection: Overview

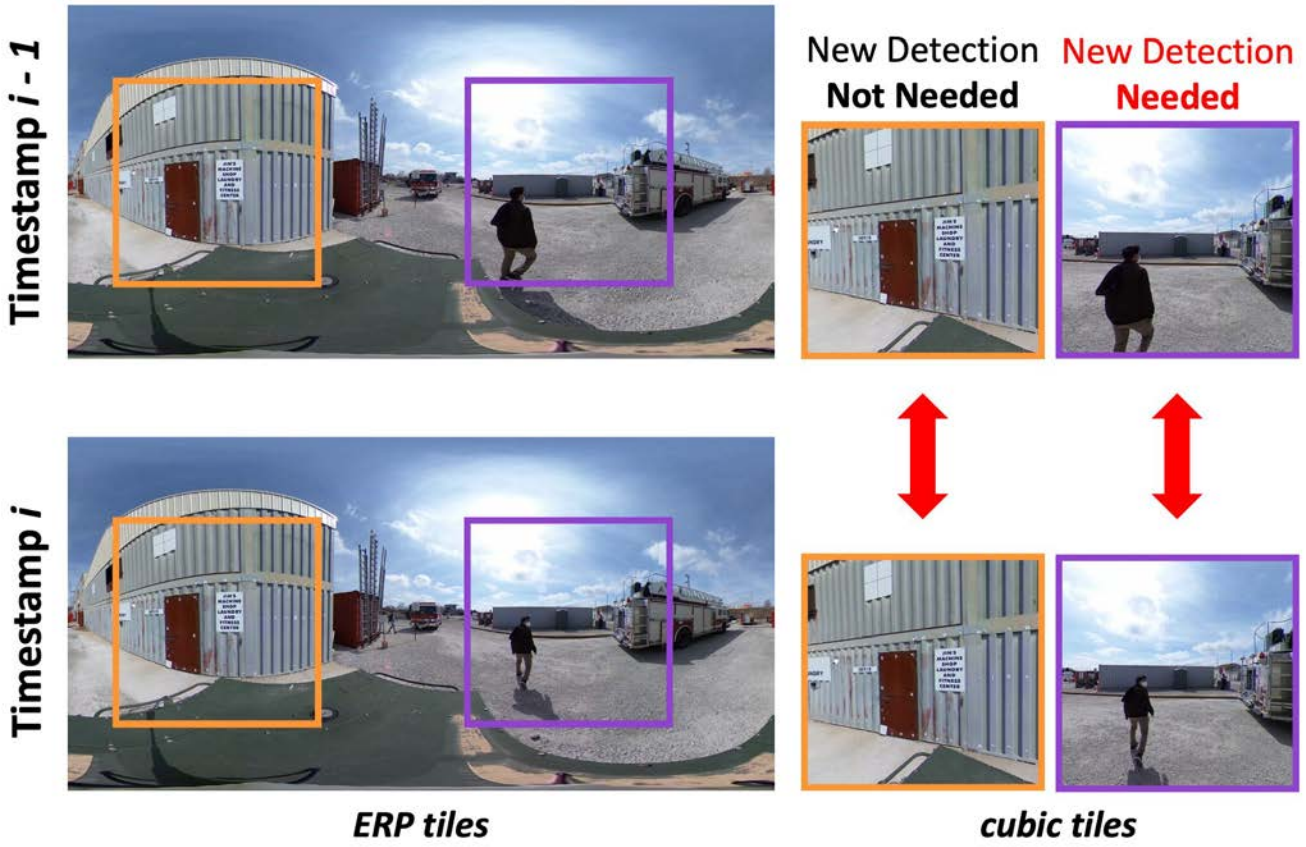
Detect on ***cubicTile***$_{i,j}$ (*cubic tile j* at timestamp *i*) only when

- Condition (1) ***cubicTile***$_{i,j}$ is structurally different from ***cubicTile***$_{i-1,j}$
- Condition (2) ***cubicTile***$_{i-1,j}$ contains a high *object cohesion*

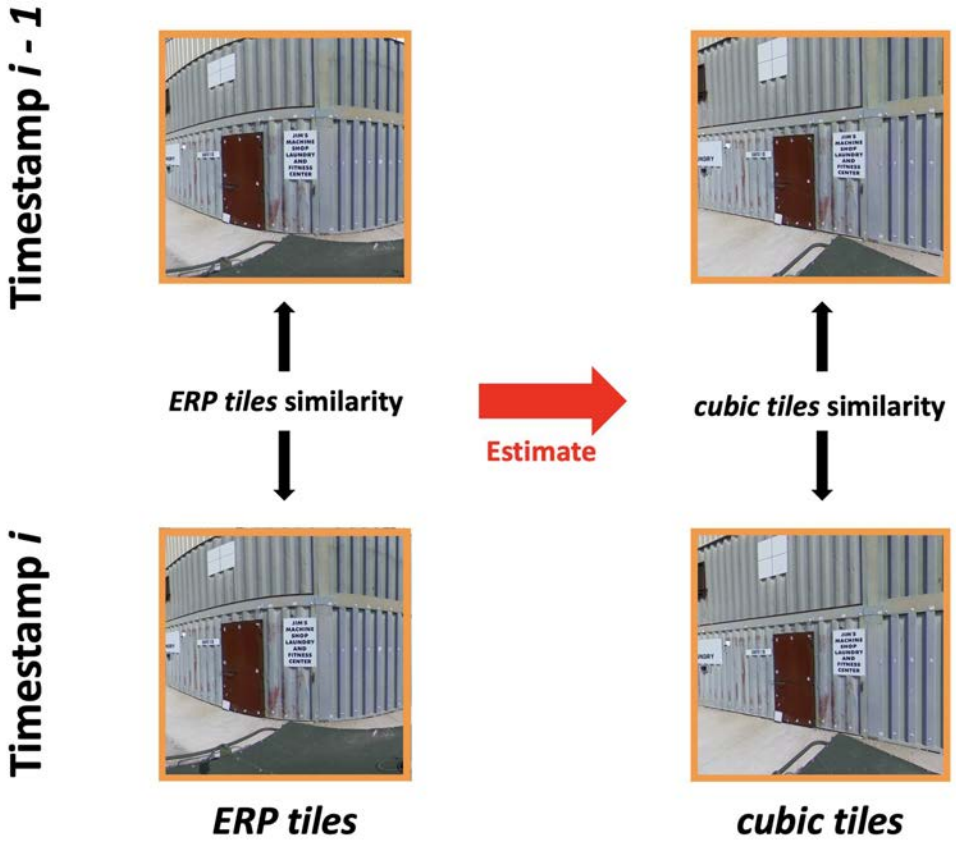
Otherwise, inference result of the previous timestamp is adopted for ***cubicTile***$_{i,j}$



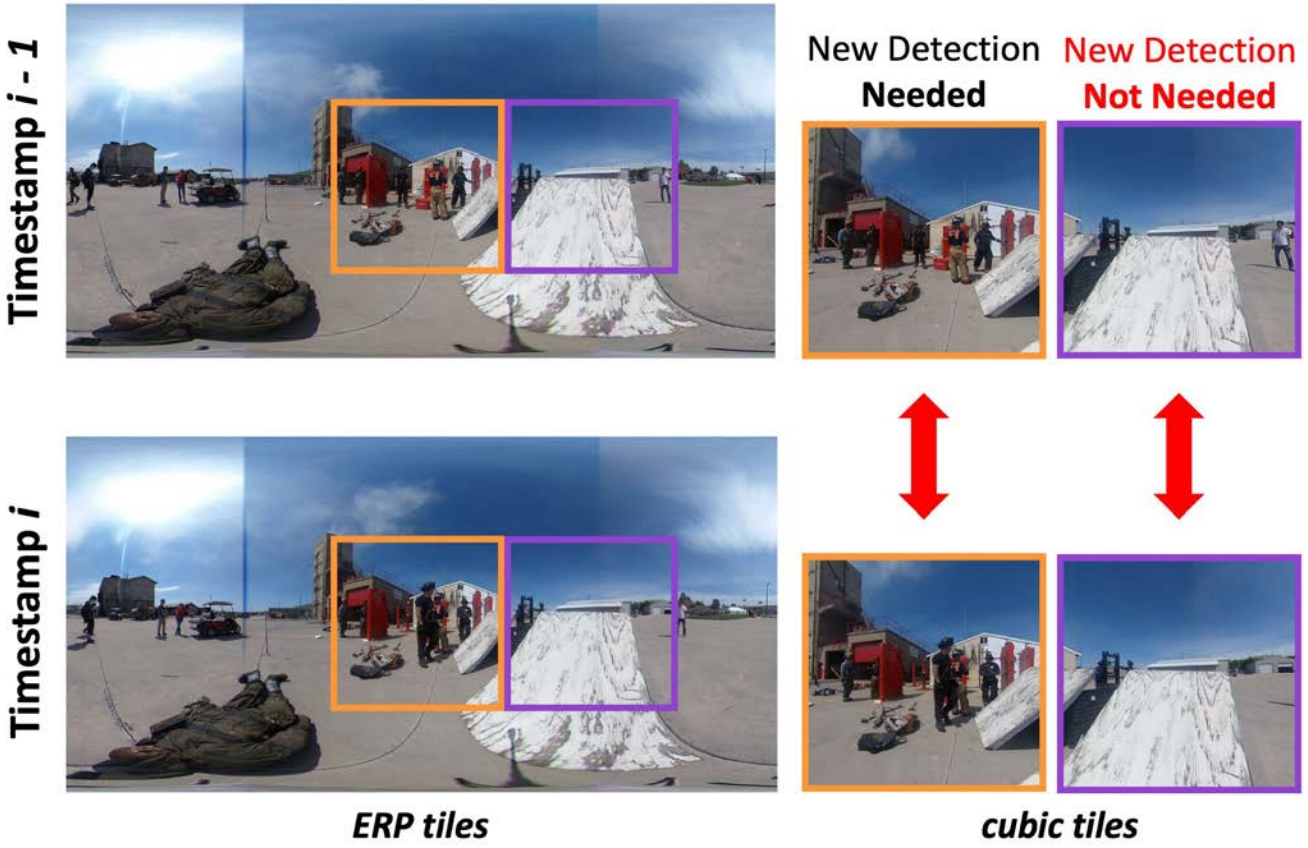
Inter-Frame Similarity (*sim*): Motivation



Inter-Frame Similarity (*sim*): Estimation



Object Cohesion (oc): Motivation

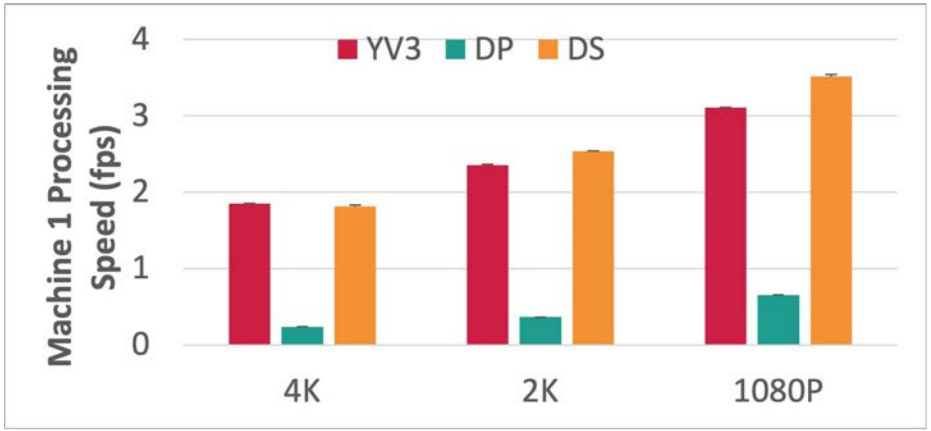


Models for Evaluation

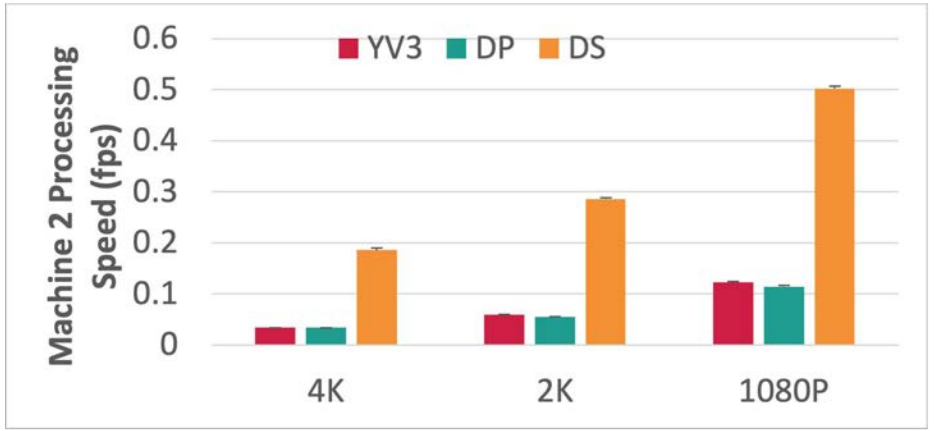
- **YV3**: Detecting using the **YOLOv3** model on **ERP** frames
- **DP**: Detecting using the **YOLOv3** model on cubic tiles generated by the **Dual-Projection** process
- **DS**: Our approach, detecting using the **YOLOv3** model on cubic tiles generated by the **Dynamic Selection** algorithm
- A dataset of 25 360-degree videos collected at Illinois Firefighter Service Institute (19 training; 6 testing)
- One 6-core CPU with a 12GB GBP and one 4-core CPU without a GPU

Evaluation of Processing Speed

NVIDIA GeForce RTX 3080 Ti



No GPU

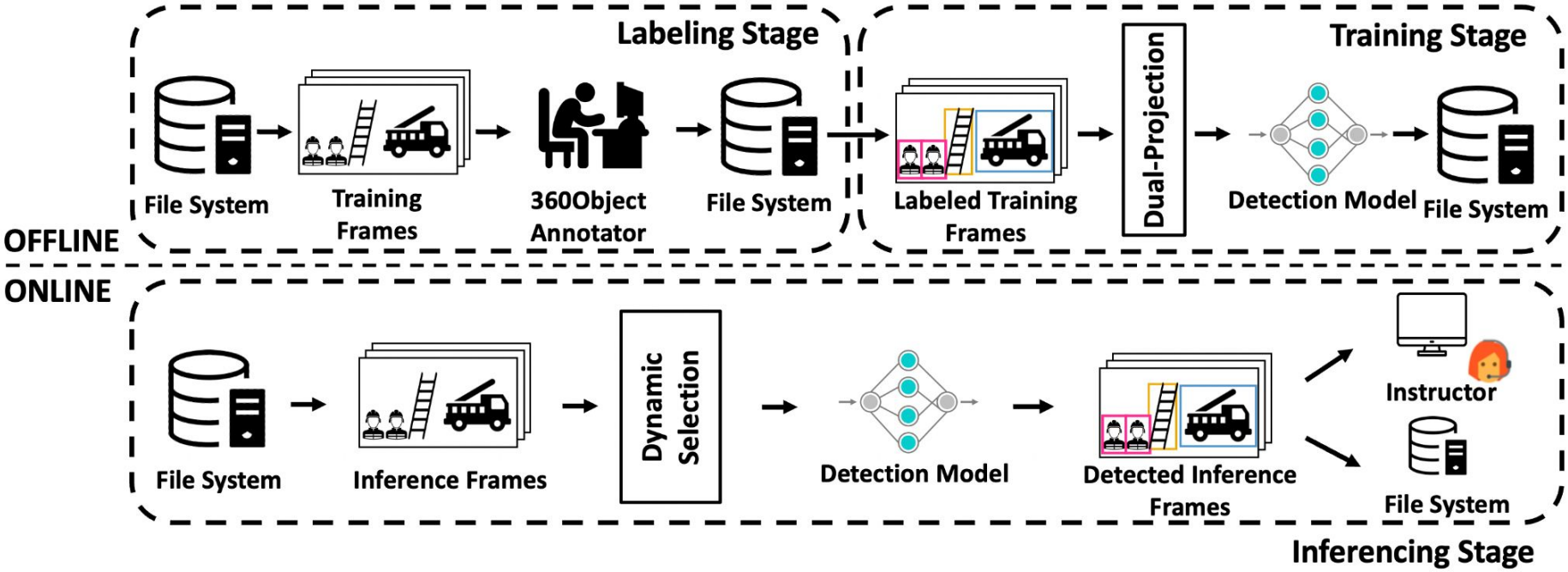


More than 4x speed up over dual projection on GPU and non-GPU machines

Evaluation of Detection Accuracy

Improved detection accuracy compared to YV3 (selected classes)

Framework Dataflow



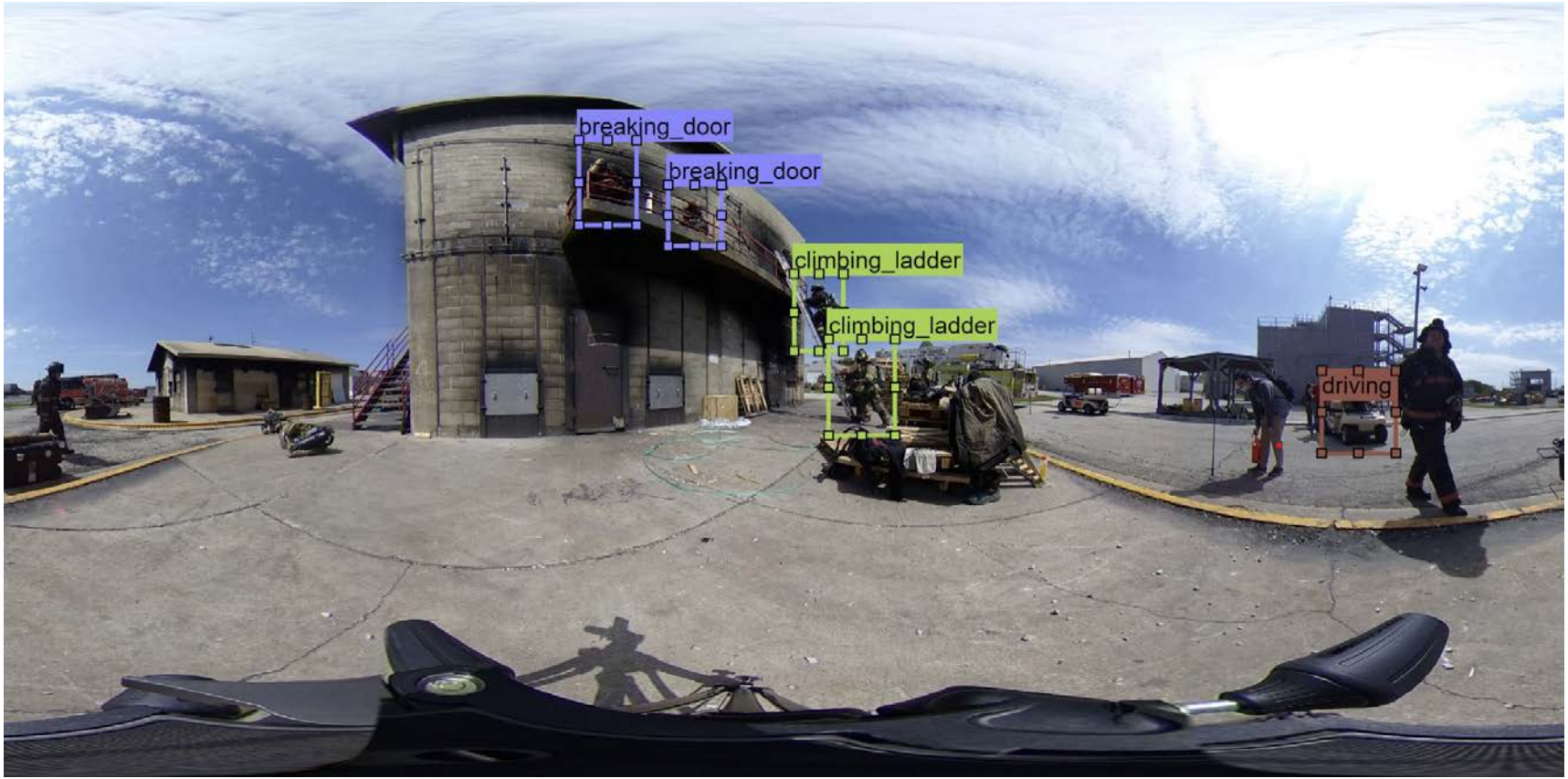
Action Detection

Action Detection: Why is it necessary?

Action detection is essential for firefighter training videos for several reasons:

1. Skill assessment
2. Safety evaluation
3. Procedure validation
4. Feedback and improvement
5. Training resource optimization

Action Detection results on ERP



Major Challenges

1. Field of view
2. Distortions and warping
3. Lack of frame-of-reference
4. Computational complexity
5. Limited labeled data

Identifiable Actions

Actions of Interest :

- Climbing up/down the ladder
- Carrying the body of civilians
- Dressing up in the firefighter gear
- Driving a vehicle
- Breaking a door or a window to enter the scene

Labeled dataset: 8 Training videos, 5 Testing videos

Data Labeling

Process differs from that of labeling the objects.

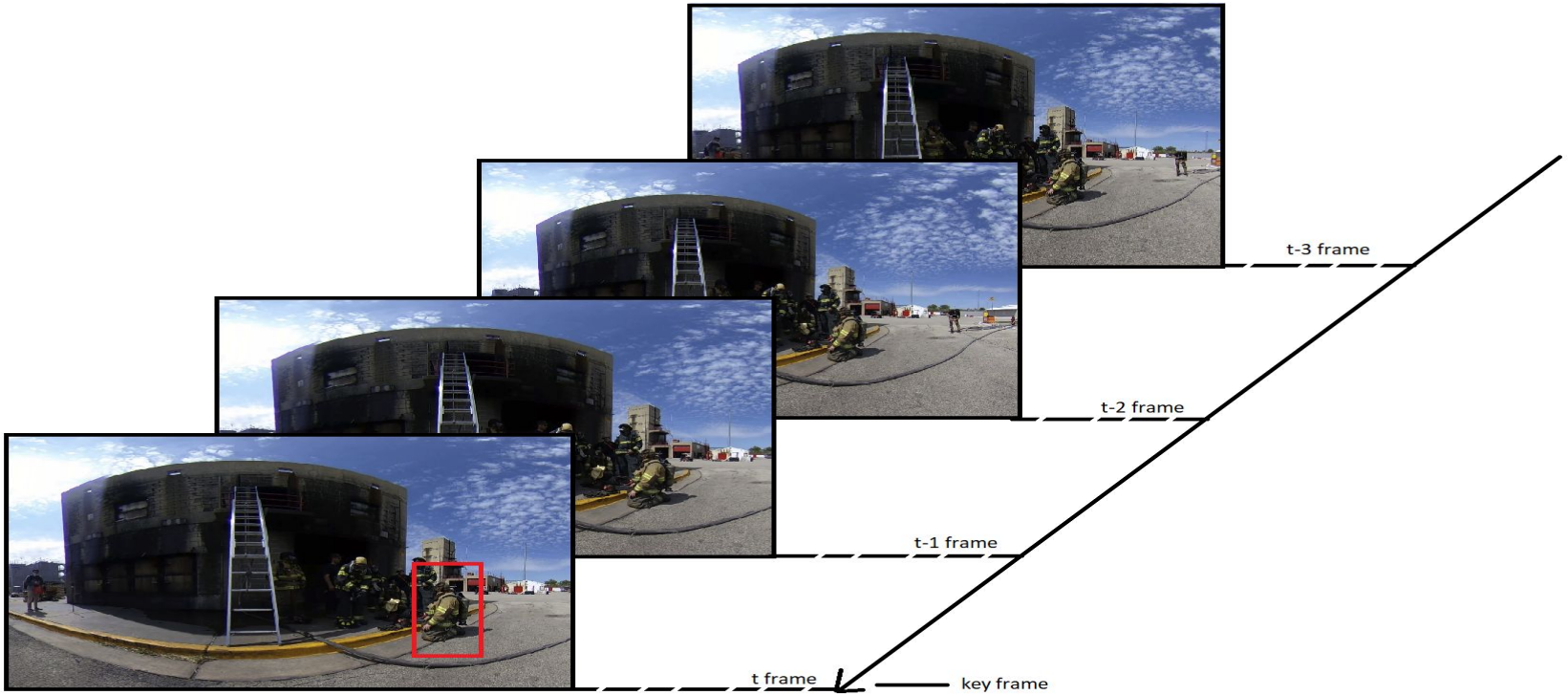
Labeling approach: MATLAB Script

Algorithm Used:

1. Region of Interest Algorithm (ROI)
2. Point Tracker Algorithm

Visualizing Data

Action Model Architecture



Actions can not be detected in a single frame.

Action Model Pipeline

Input Frames

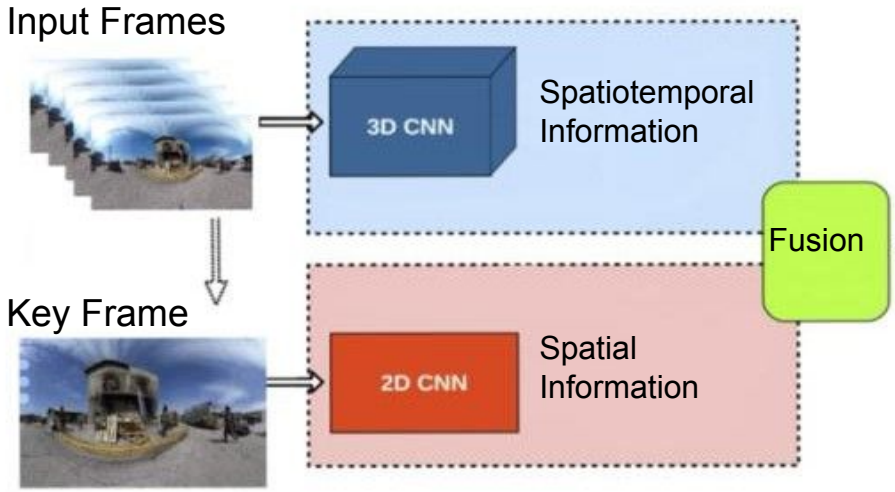
Key Frame

Model: YOWO (You only watch once)

Steps:

1. Input and Key Frame Extraction
2. 2D Feature Extraction
3. Temporal Encoding
4. Action Classification

Action Model Pipeline

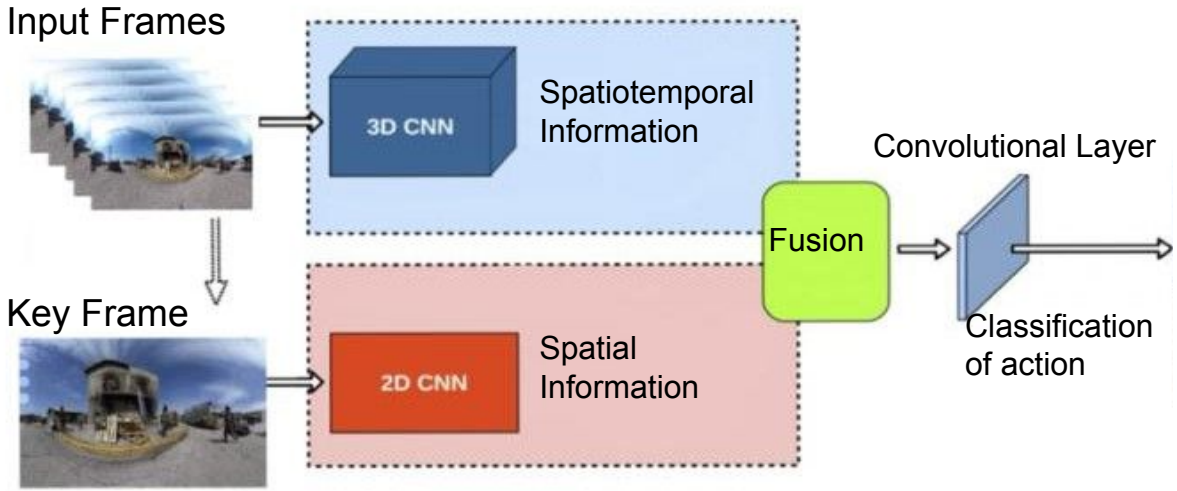


Model: YOWO (You only watch once)

Steps:

1. Input and Key Frame Extraction
2. 2D Feature Extraction
3. Temporal Encoding
4. Action Classification

Action Model Pipeline

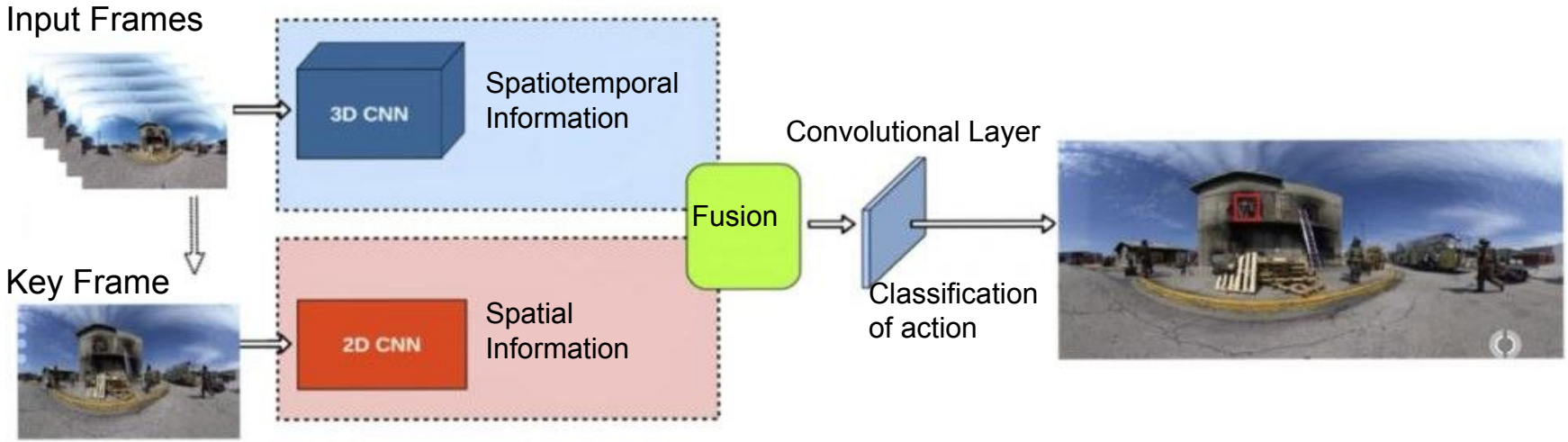


Model: YOWO (You only watch once)

Steps:

1. Input and Key Frame Extraction
2. 2D Feature Extraction
3. Temporal Encoding
4. Action Classification

Action Model Pipeline

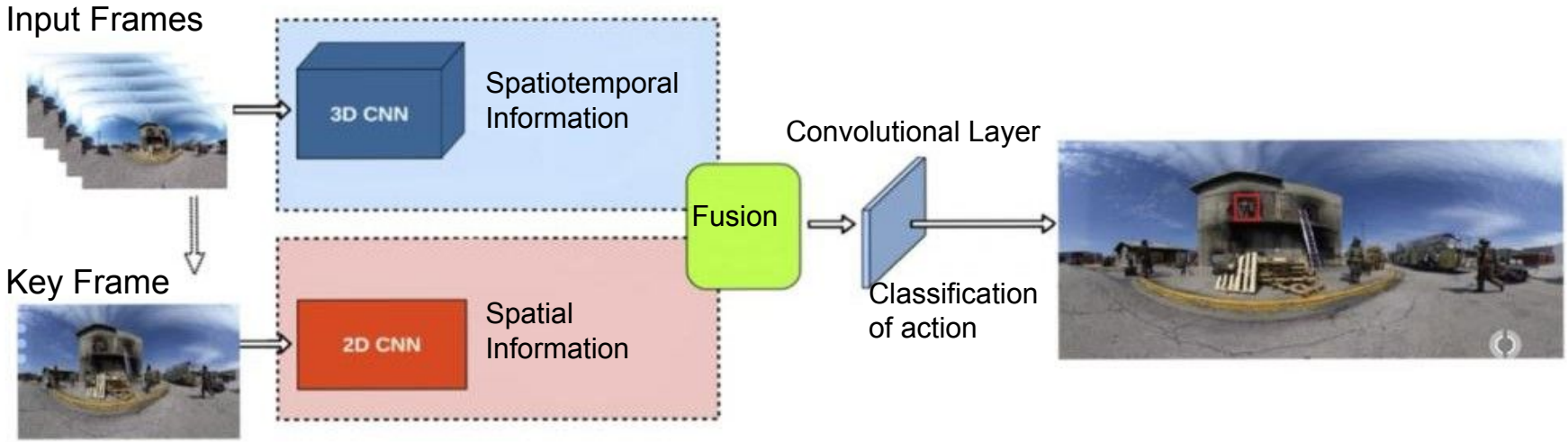


Model: YOWO (You only watch once)

Steps:

1. Input and Key Frame Extraction
2. 2D Feature Extraction
3. Temporal Encoding
4. Action Classification

Action Model Pipeline

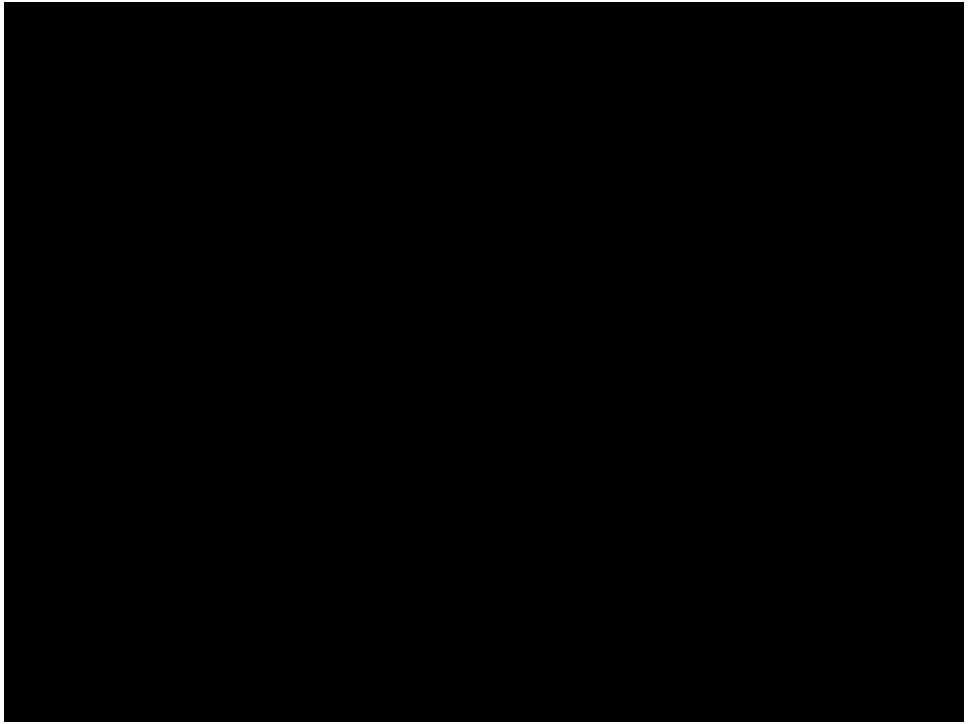


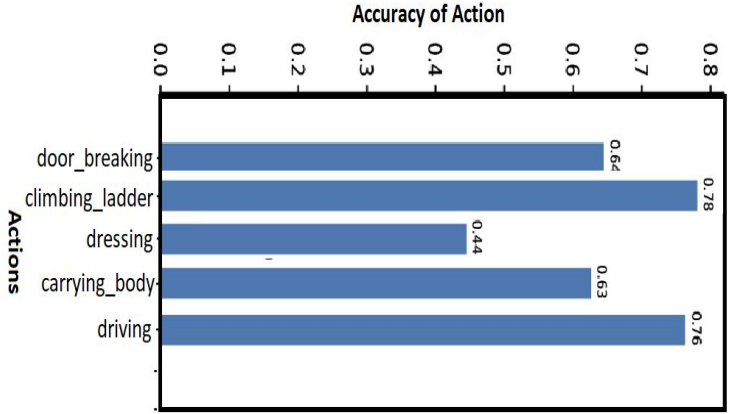
Model: YOWO (You only watch once)

Steps:

1. Input and Key Frame Extraction
2. 2D Feature Extraction
3. Temporal Encoding
4. Action Classification

Action Prediction Evaluation





Demo

http://10.195.70.26:3000

VQS: Current Supported Objects & Events

Objects

- Firefighter
- Civilian
- Ladder
- Fire
- Window
- Oxygen Tank
- Door
- Fire Truck
- Firefighter Helmet
- Civilian Car
- Firefighter Mask
- Smoke

Events

- Breaking Door
- Climbing Ladder
- Driving
- Dressing Firefighter
- Carrying Body

Conclusion

- **Dynamic Selection**
  - >4X Speedup
  - >25% memory usage reduction
  - Improved detection accuracy

- **Viewing and Query Service**
  - Object Labeling
  - Object Search

Acknowledgments

- Illinois Fire Service Institute (IFSI)
- National Science Foundation (NSF)

Contact Information

- Presenters:
  - **Jiaxi Li** (jiaxili3@illinois.edu)
  - **Jingwei Liao**(jliao2@gmu.edu)
  - **Aditi Tiwari** (aditit5@illinois.edu)