ACTION Detection & Anticipation

Snehashis MAJHI

Email: snehashis.majhi@inria.fr

Ph.D. Candidate @STARS Team INRIA

Collaboration with TOYOTA Motor Europe

Action Classification

→ Video Classifier →

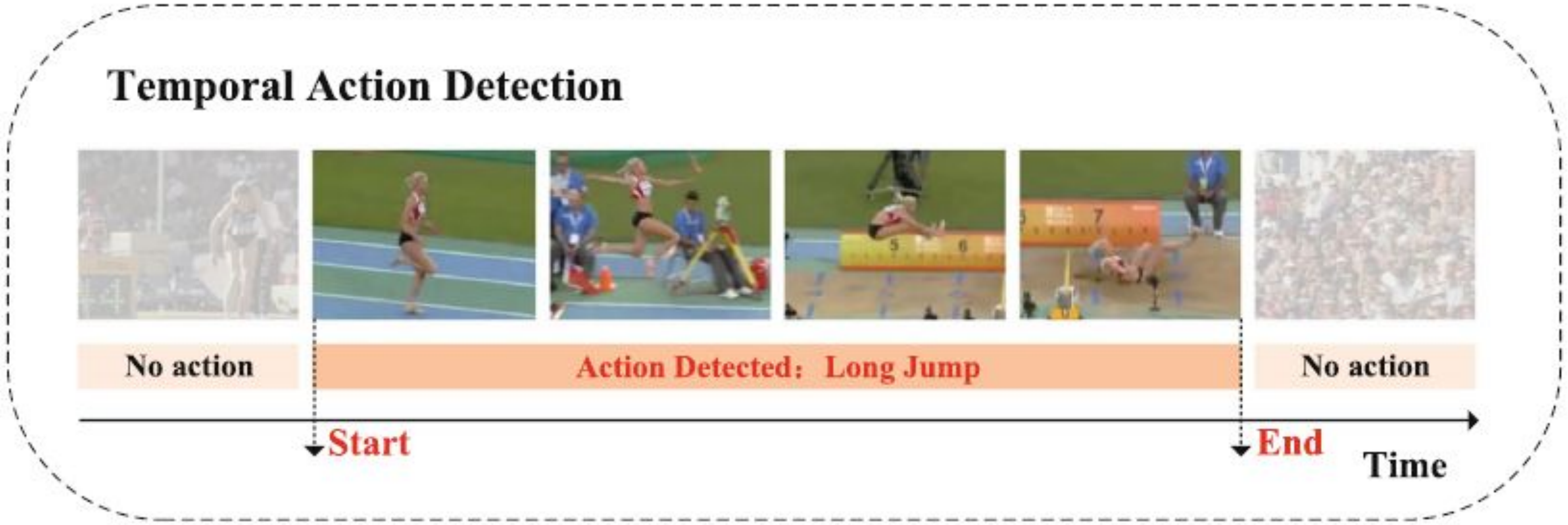
CLASS	SCORE
Swing Dancing	0.52
Salsa Dancing	0.39
Holding Hands	0.03
Walking	0.01
....	...

Action Classification

→ Video Classifier →

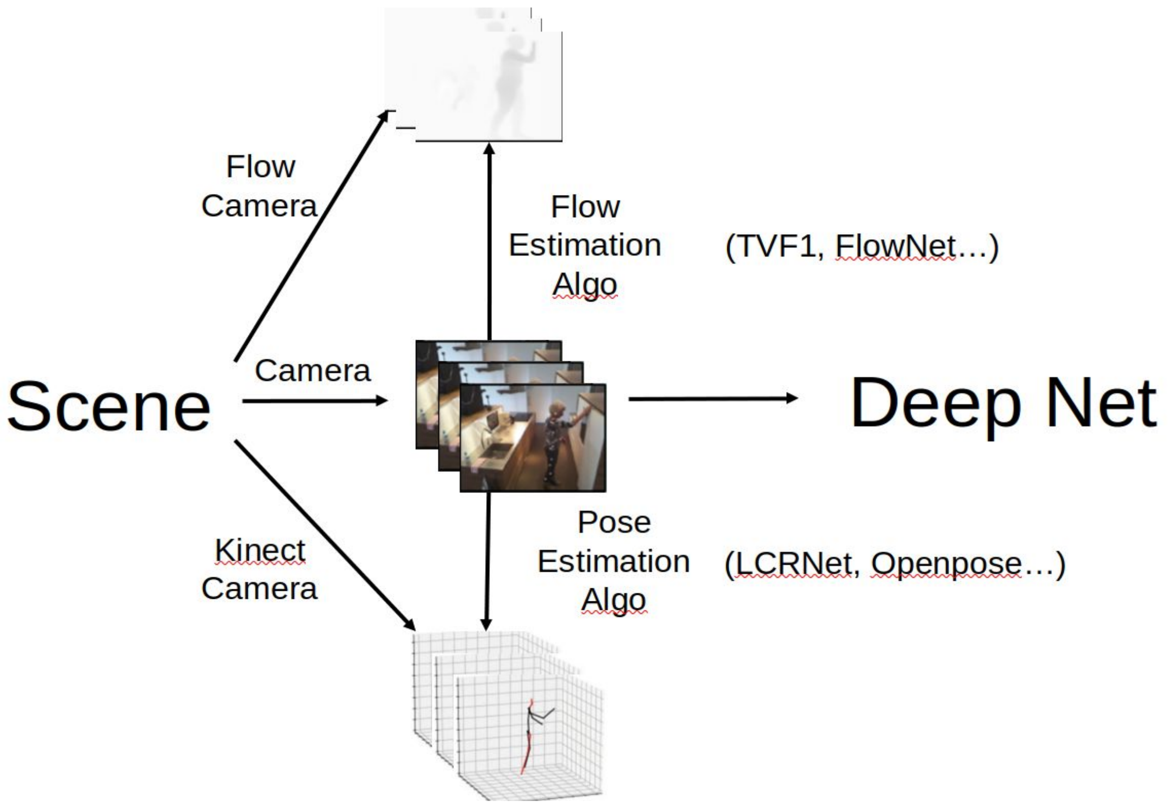
CLASS	SCORE
Swing Dancing	0.52
Salsa Dancing	0.39
Holding Hands	0.03
Walking	0.01
....	...

Action Detection

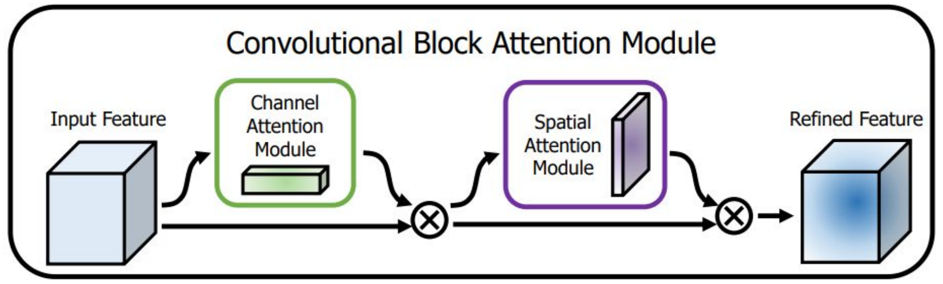


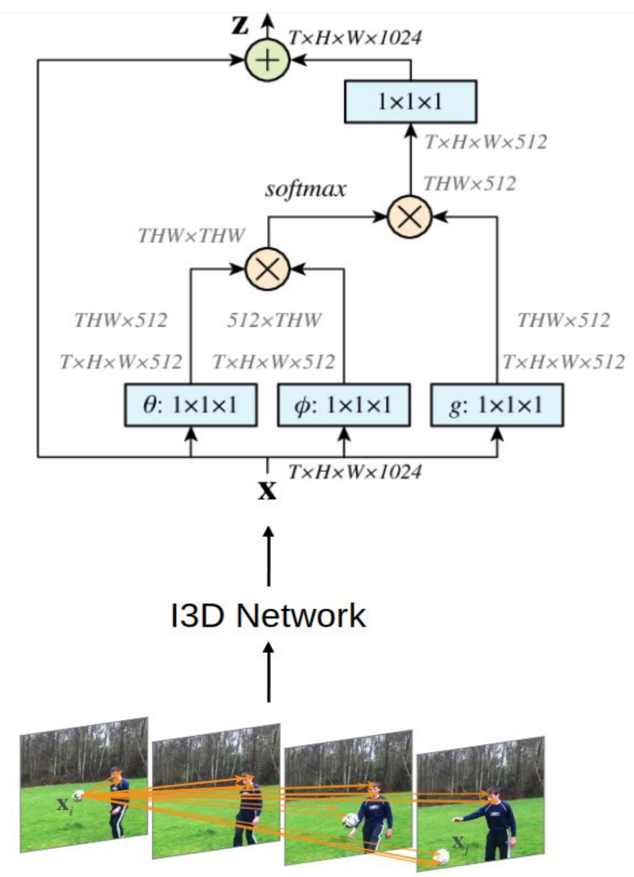
Recap:

Combining Multiple Modalities for HAR

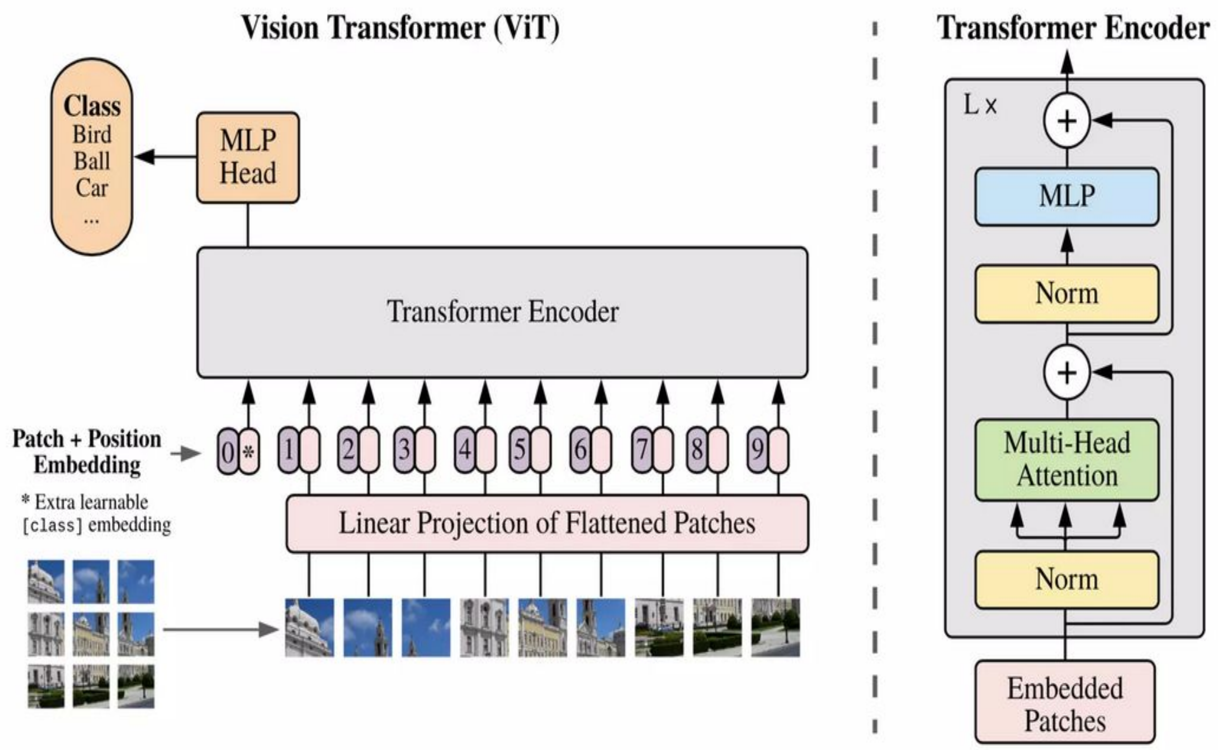


Attention Mechanism

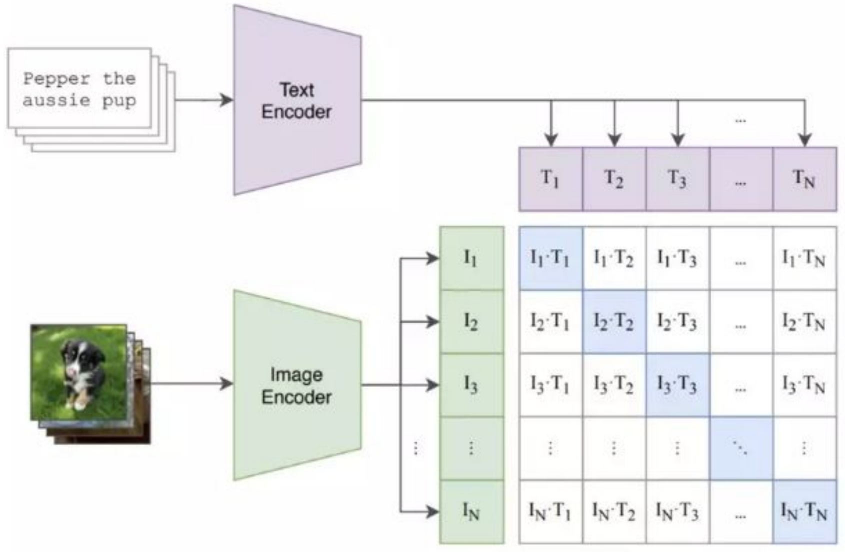




Transformer Models



CLIP:Vision-language

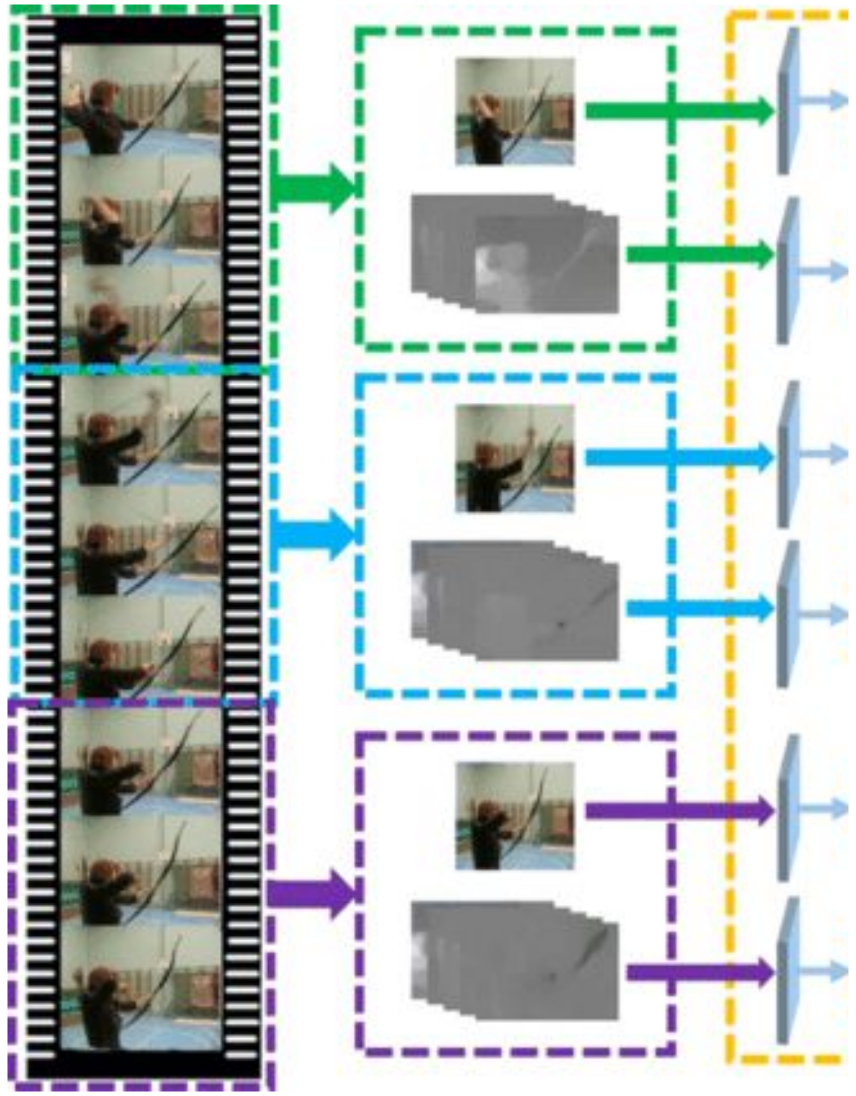


STEPS in Action Detection

Step-1

Feature Extraction:

-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.
-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.



STEPS in Action Detection

Step-1

Feature Extraction:

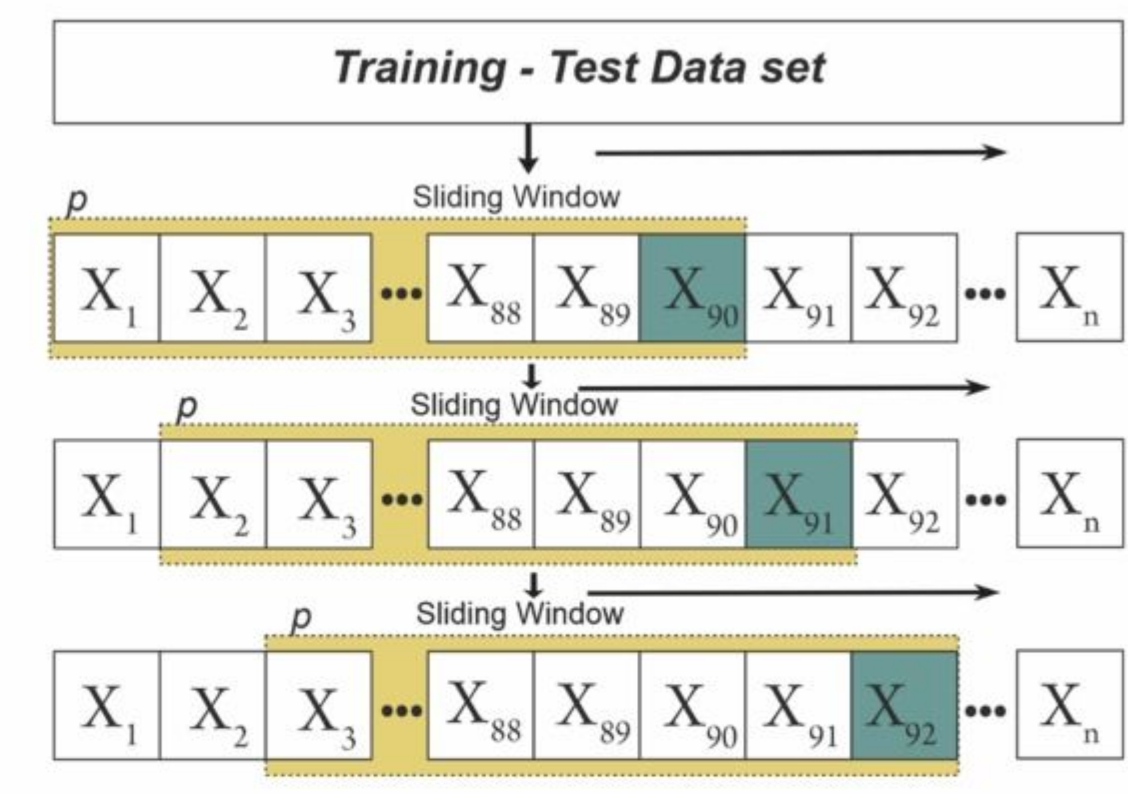
-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.
-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.

Step-2

Proposal Generation:
-- **Sliding Windows**: video is divided into **overlapping or non-overlapping windows**.

Each window is treated as a candidate for containing an action.

These are usually fixed-length windows (8,or 16,or 64 frames).



STEPS in Action Detection

Step-1

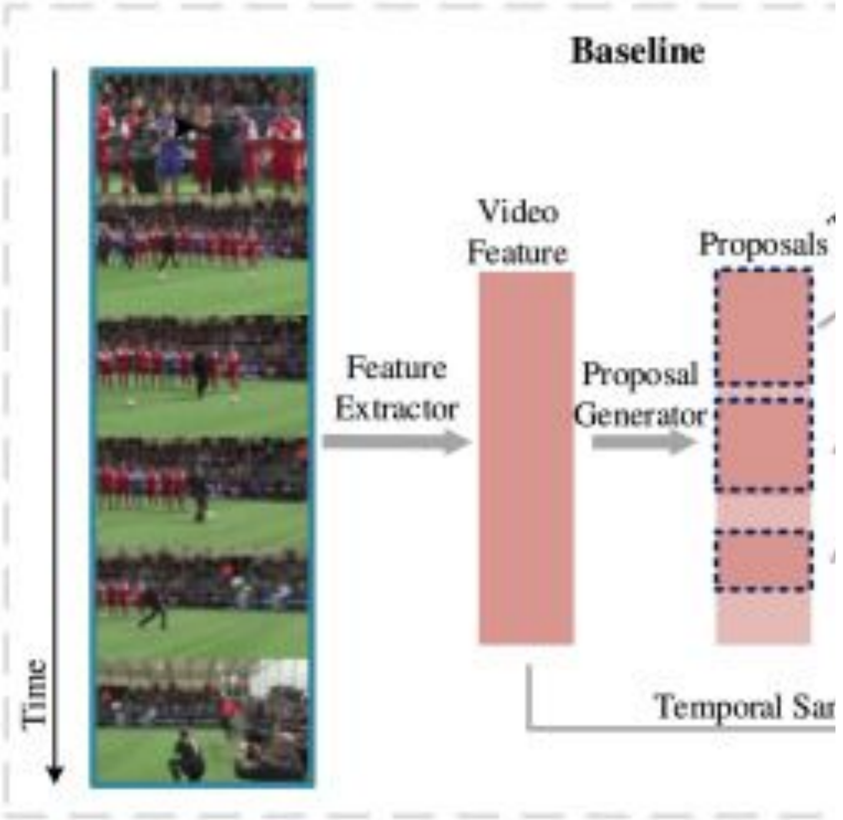
Feature Extraction:

-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.
-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.

Step-2

Proposal Generation:
-- **Temporal Proposal or Anchors**: can predict temporal regions where actions might occur.

These are trained using labeled data and learn to identify proposals based on patterns in the video.



STEPS in Action Detection

Step-1

Feature Extraction:

-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.
-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.

Step-2

Proposal Generation:

-- **Temporal Proposal or Anchors**: can predict temporal regions where actions might occur.

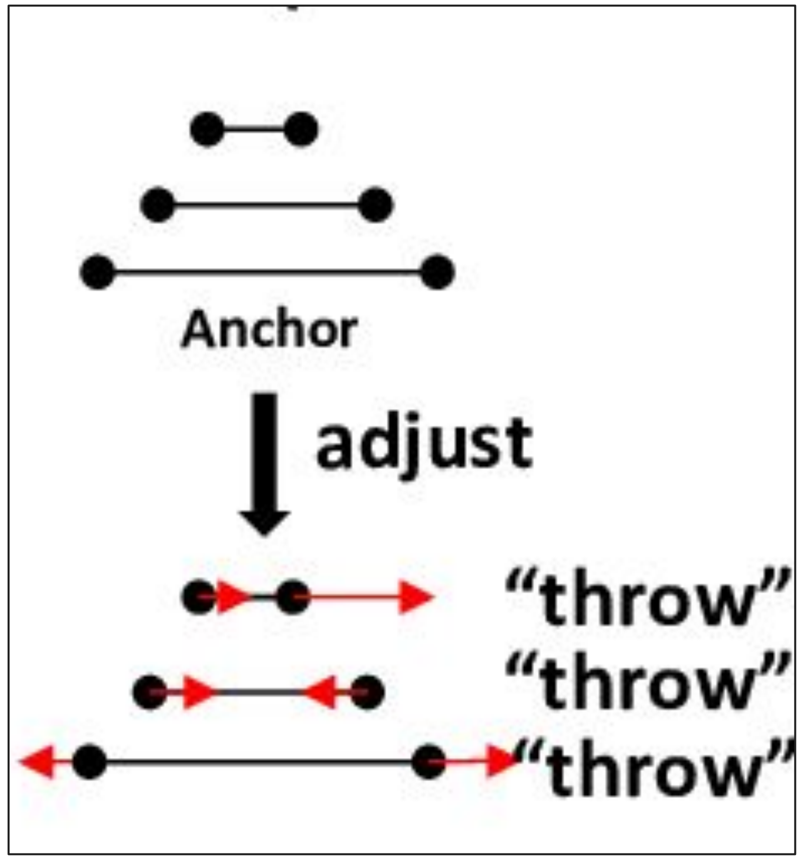
These are trained using labeled data and learn to identify proposals based on patterns in the video.

Step-3

Temporal Localization:

-- **Boundary Refinement**
Proposals often need adjustment to **match the ground-truth start and end times.**

-- **Classification**
Assign each proposal an action label using classifiers like fully connected layer

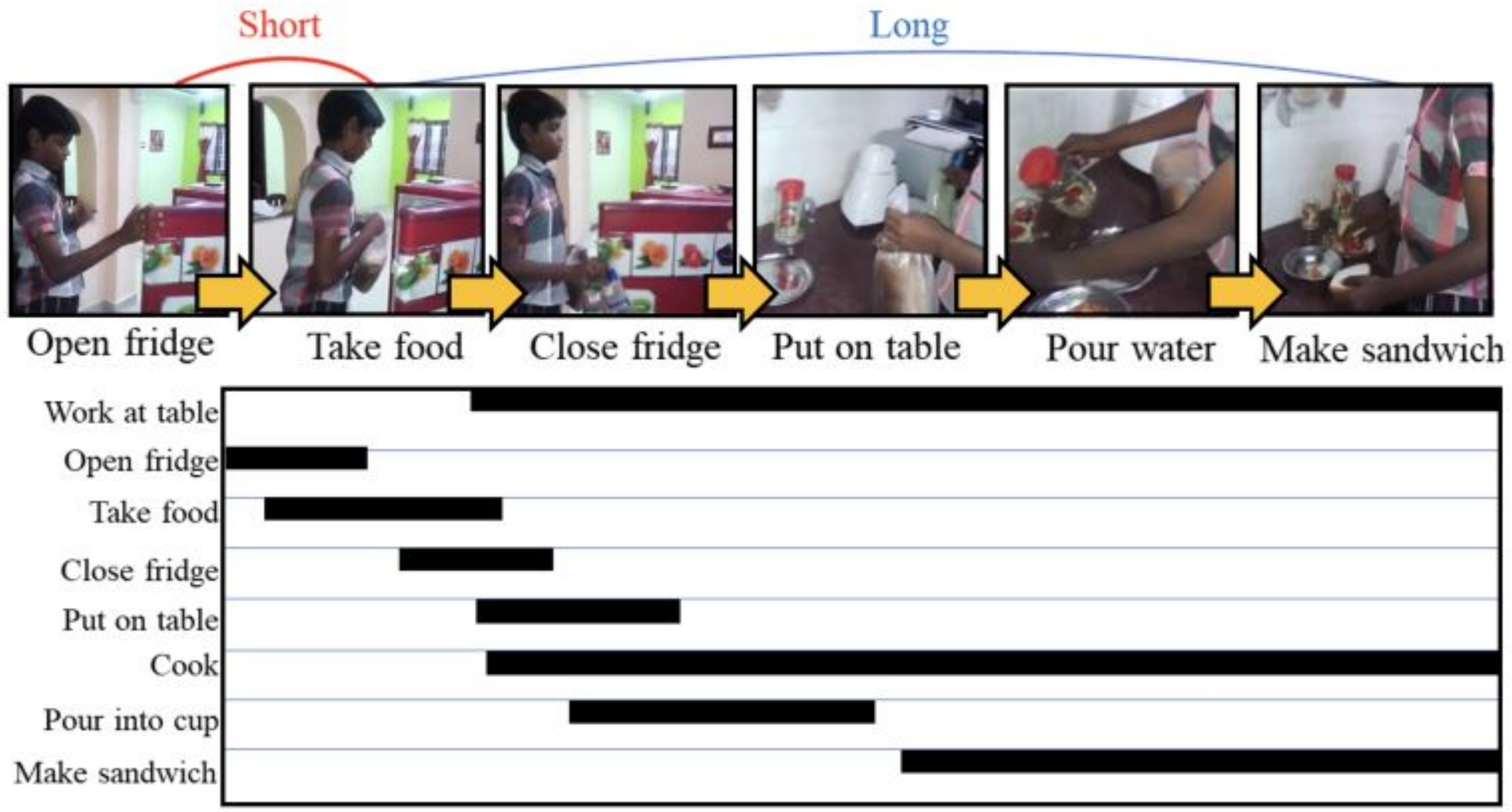


Important STEPS in Action Detection

Multi-scale Temporal Modeling :

- Actions occur at different durations (short gestures vs. long activities). Multi-scale features or temporal pyramid networks can handle this variability.

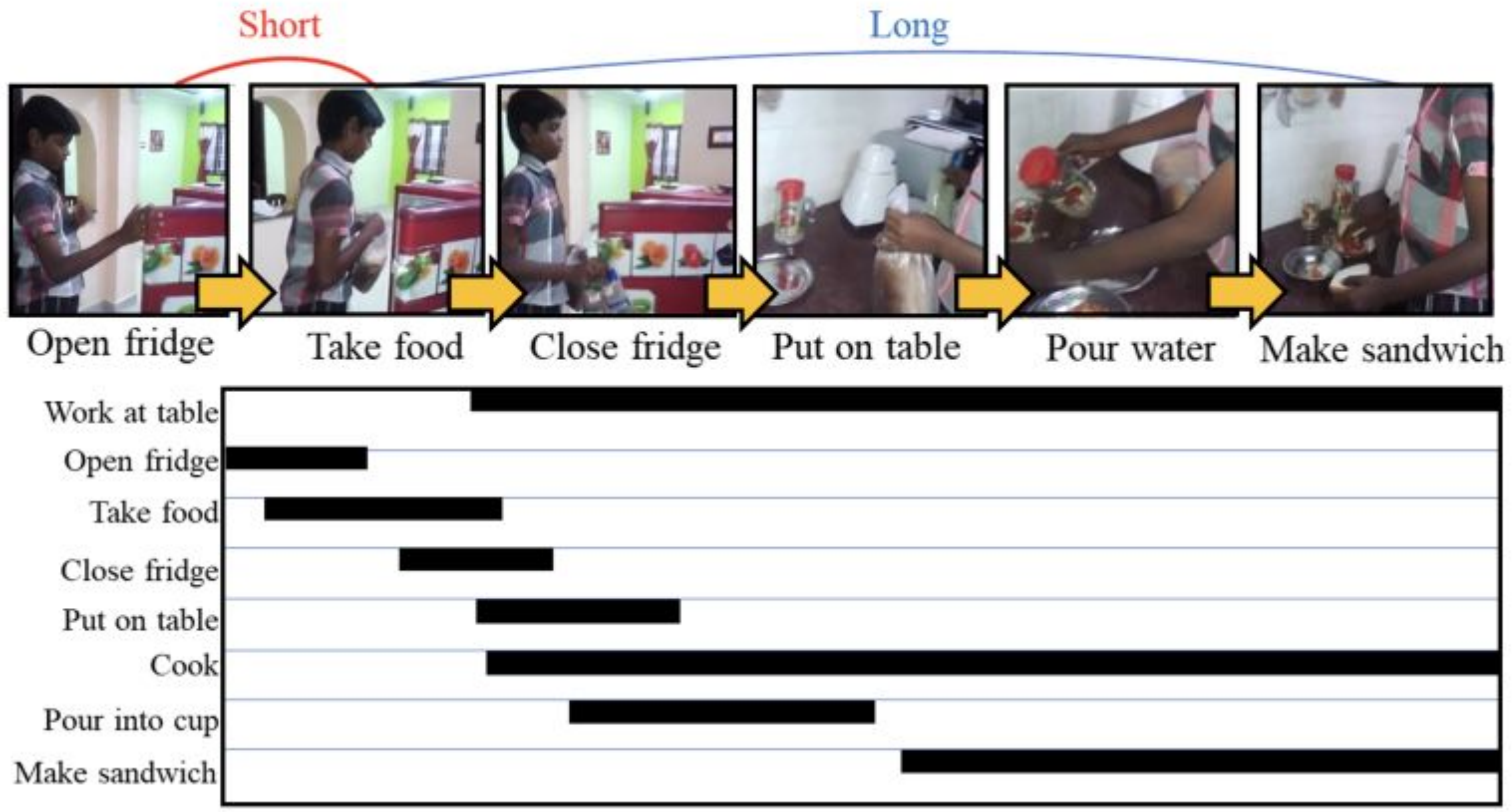
- Capture dependencies and relationships across video frames to better classify actions.



Important STEPS in Action Detection

Temporal Attention :

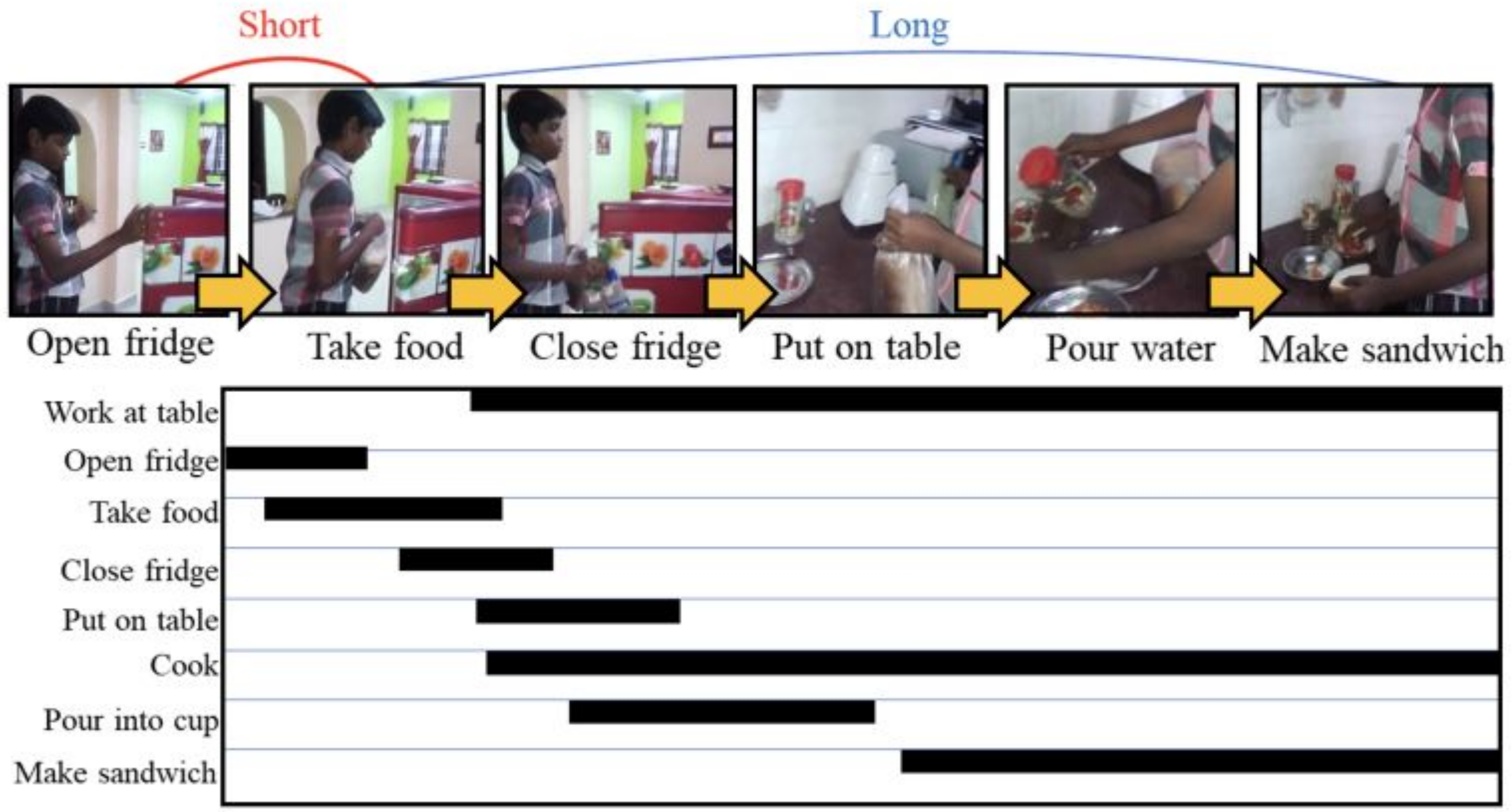
- Focus on discriminative parts of the video (important frames) using attention mechanisms to improve classification accuracy.



Important STEPS in Action Detection

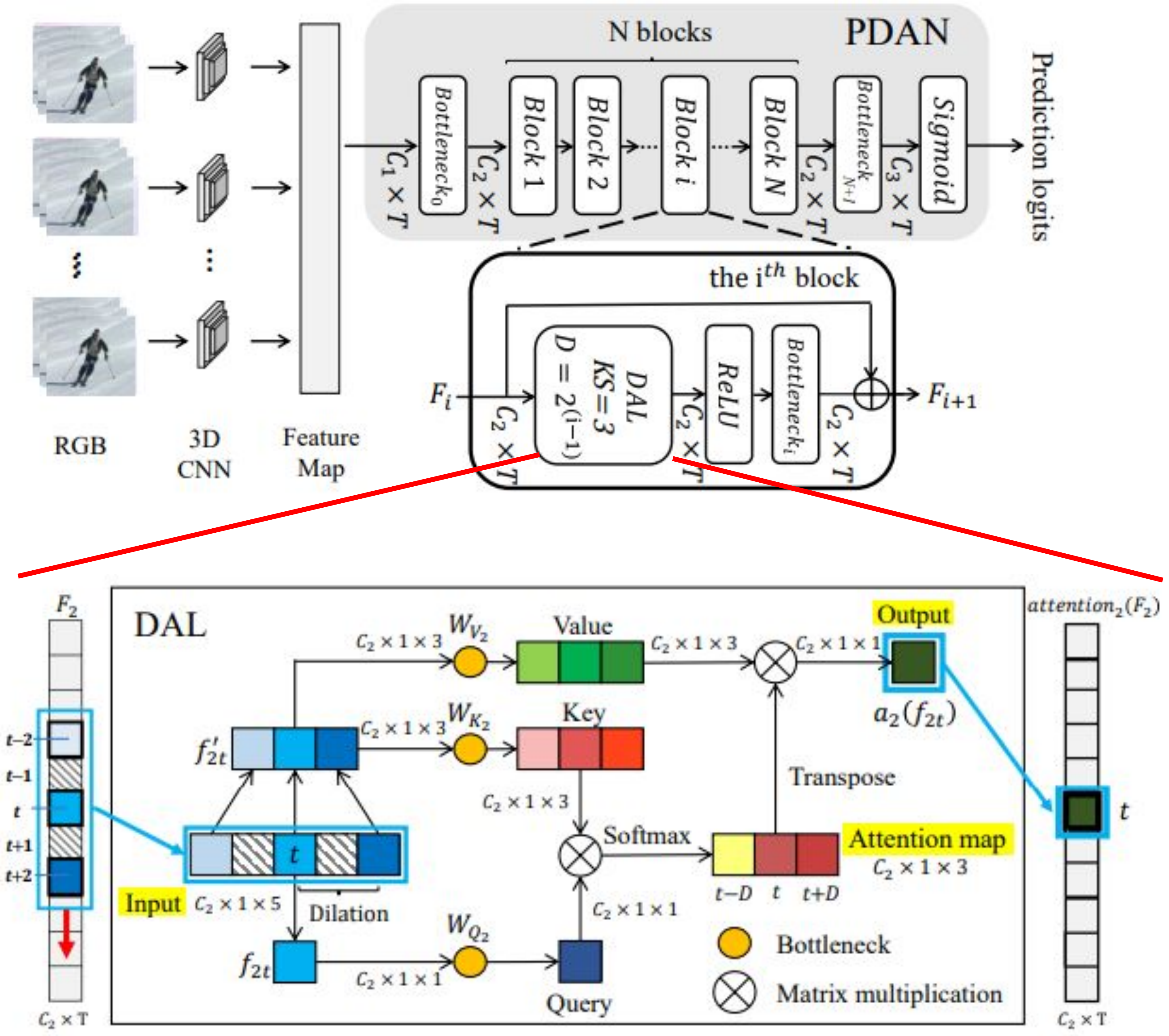
Temporal Attention :

- Focus on discriminative parts of the video (important frames) using attention mechanisms to improve classification accuracy.



Let's See Few TOP Action Detection Methods , Excited !

PDAN:

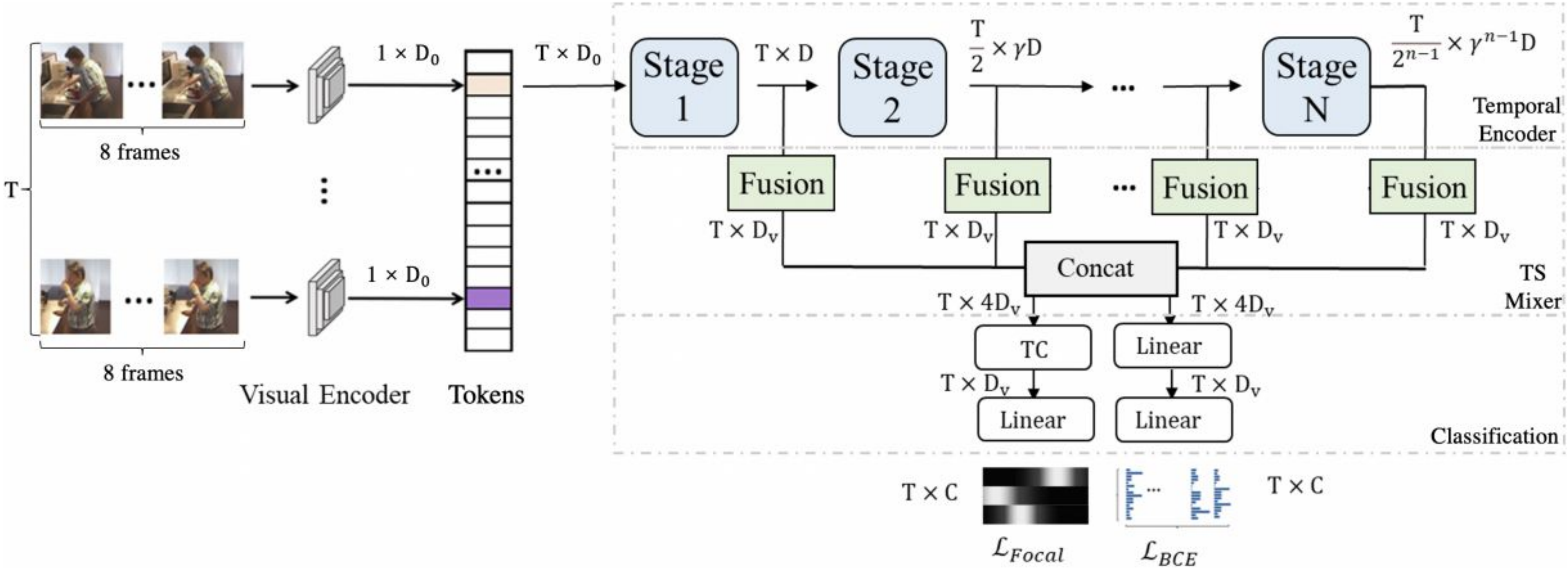


- **Leverages dilated convolutions** and **attention mechanisms** to handle varying action durations and refine predictions.
- Pyramid dilated convolutions are used to capture features at multiple temporal scales.

- By varying the dilation rates in convolutional layers, PDAN can effectively expand the receptive field, ensuring that both short-term and long-term temporal dependencies are captured.
- This helps in detecting actions of varying durations (e.g., short gestures vs. prolonged activities).

- PDAN integrates attention mechanisms to focus on important temporal regions within the video.
- **Global Attention**: Identifies key frames across the entire video
- **Local Attention**: Focuses on refining details within action boundaries

MS-TCT:



- Integrates **multi-scale temporal modeling** and **transformer-based attention mechanisms** to enhance action localization and classification. It builds on the strengths of temporal convolutional networks (TCNs) and transformers, combining them in a unified architecture.

MS-TCT: Key Components

Multi-Scale Temporal Convolutions:

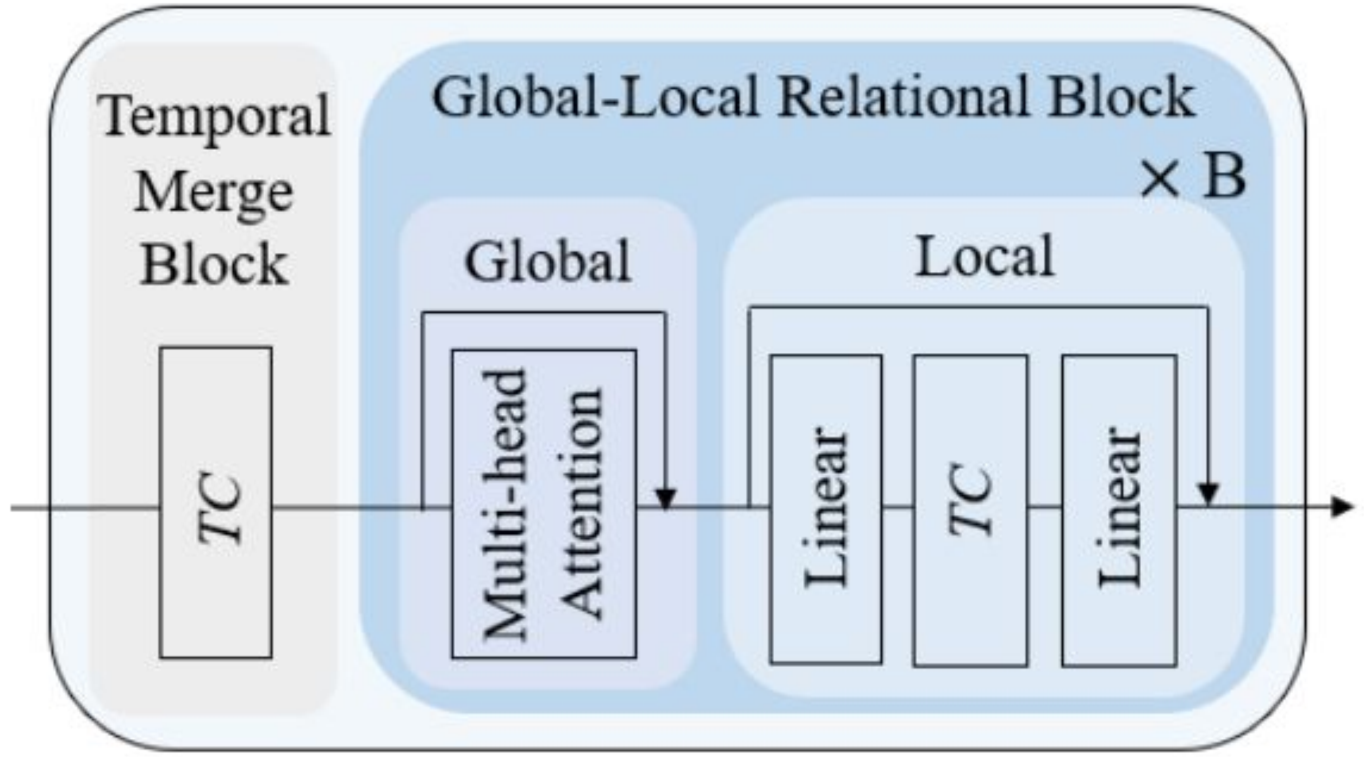
- Uses layers with increasing kernel sizes and dilation rates to extract temporal features at various resolutions.
- Efficiently captures actions of different durations without significantly increasing computational cost.
- Inspired by the success of **TCNs** in temporal modeling.

Temporal Transformers:

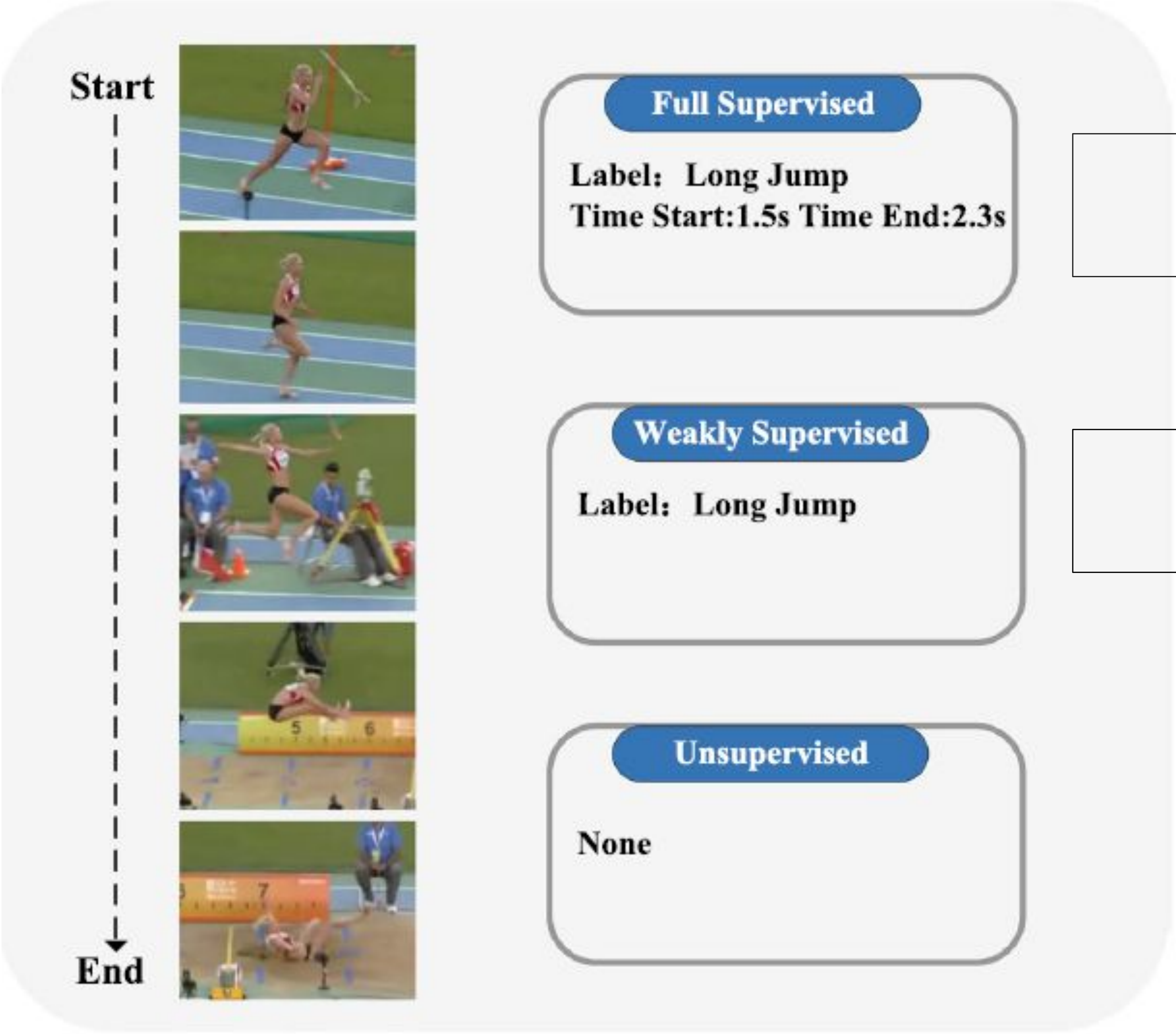
- Employs self-attention to model global temporal dependencies.
- Allows the framework to capture long-range contextual information, which is critical for detecting complex or overlapping actions.

Multi-Scale Feature Aggregation:

- Aggregates features from different temporal scales to create a unified representation.
- Ensures that both short-term and long-term patterns are included in the final predictions.



Learning Methods:



Frame-by-Frame Annotations required
- Time Consuming
- Boundary Region Could be prone to error

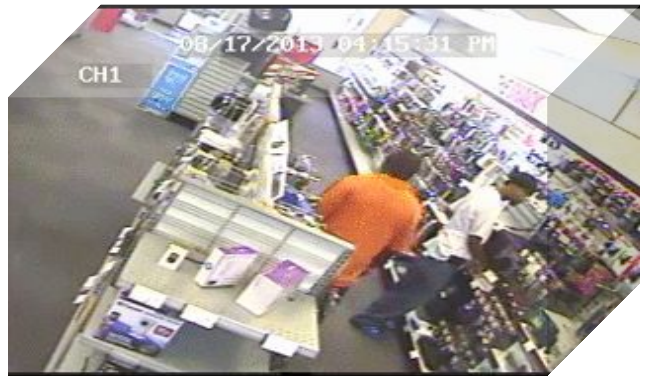
Video-level Annotations required
- Easy to Obtain
- Less mistakes in annotation

Anomaly Detection ?

Real-world Anomalies?

Abuse | Arrest | Arson | Assault | Fighting

Burglary | Shoplifting | Shooting | Robbery | RoadAccident



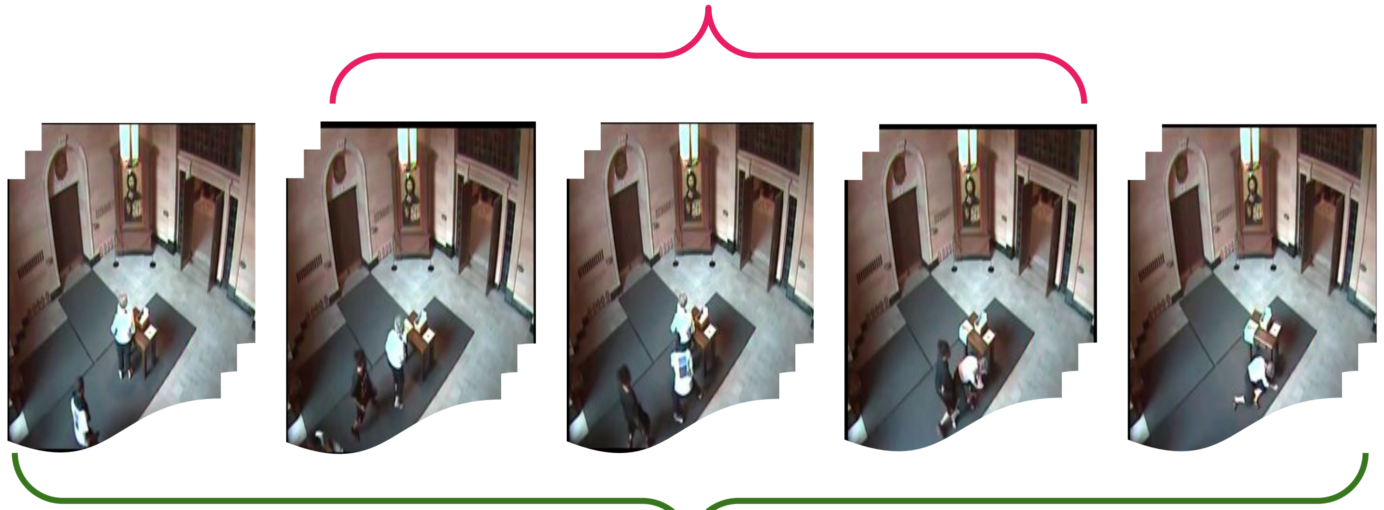
Stealing | Vandalism | Explosion

Source = CCTV

Is it that easy to detect real-world anomalies?

- No Temporal Annotation in Videos

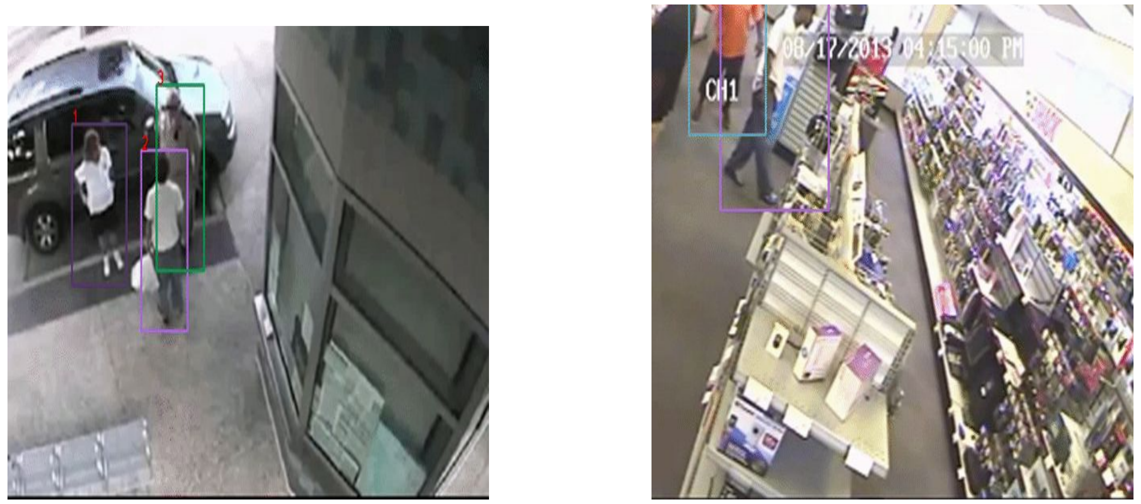
[Supervised]
Temporal Annotations



Video-level Annotations
[Weakly Supervised]

- Sparsity of Anomaly

- Human Centric fine-grained Anomalies



- Long and Short Duration Anomalies

Long Short

Our Two Recent Works

CVPR 2025 Submission #11647. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Just Dance with π!
A Poly-modal Inductor for Weakly-supervised Video Anomaly Detection

Anonymous CVPR submission

Under review as a conference paper at ICLR 2025

MIXTURE OF EXPERTS GUIDED BY GAUSSIAN SPLATTERS MATTERS: A NEW APPROACH TO WEAKLY-SUPERVISED VIDEO ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

CVPR'25 ACCEPTED

Just Dance with π!

A Poly-modal Inductor for Weakly-supervised Video Anomaly Detection

Motivation

- RGB only features are not sufficiently distinctive enough to distinguish **complex anomalies like shoplifting and visually similar normal events**.

- **Towards robust complex real-world anomaly detection, it is essential to augment RGB with additional modalities.**

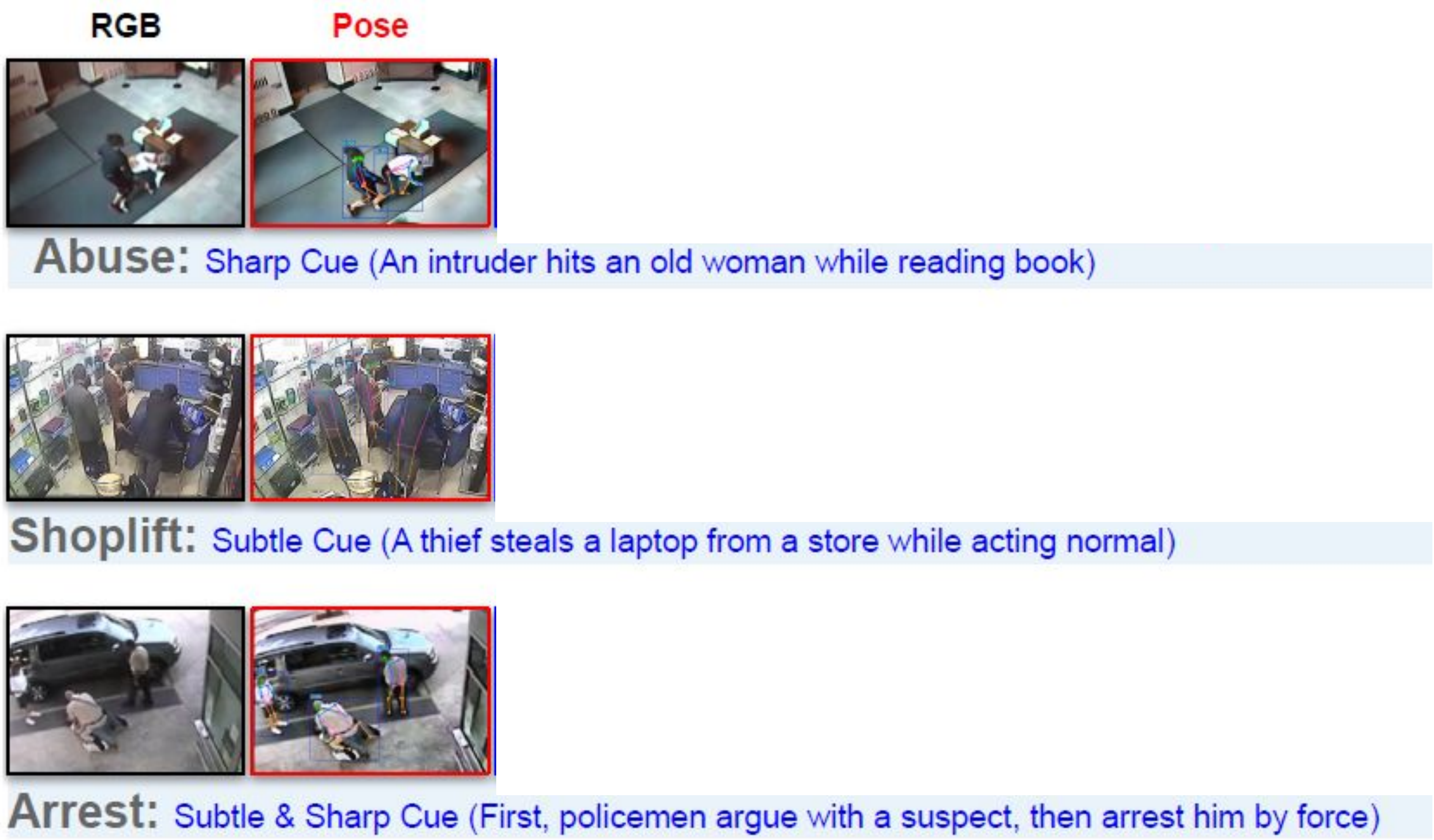
- **But how many additional modalities?**

(a) Complex real-world anomalies with multi-modal saliencies

Motivation

- **RGB only features** are not sufficiently distinctive enough to distinguish **complex anomalies like shoplifting and visually similar normal events**.

- **Towards robust complex real-world anomaly detection, it is essential to augment RGB with additional modalities.**



(a) Complex real-world anomalies with multi-modal saliencies

- **But how many additional modalities?**
  - **ONE**

Motivation

- RGB only features are not sufficiently distinctive enough to distinguish **complex anomalies like shoplifting and visually similar normal events**.

- **Towards robust complex real-world anomaly detection, it is essential to augment RGB with additional modalities.**

(a) Complex real-world anomalies with multi-modal saliencies

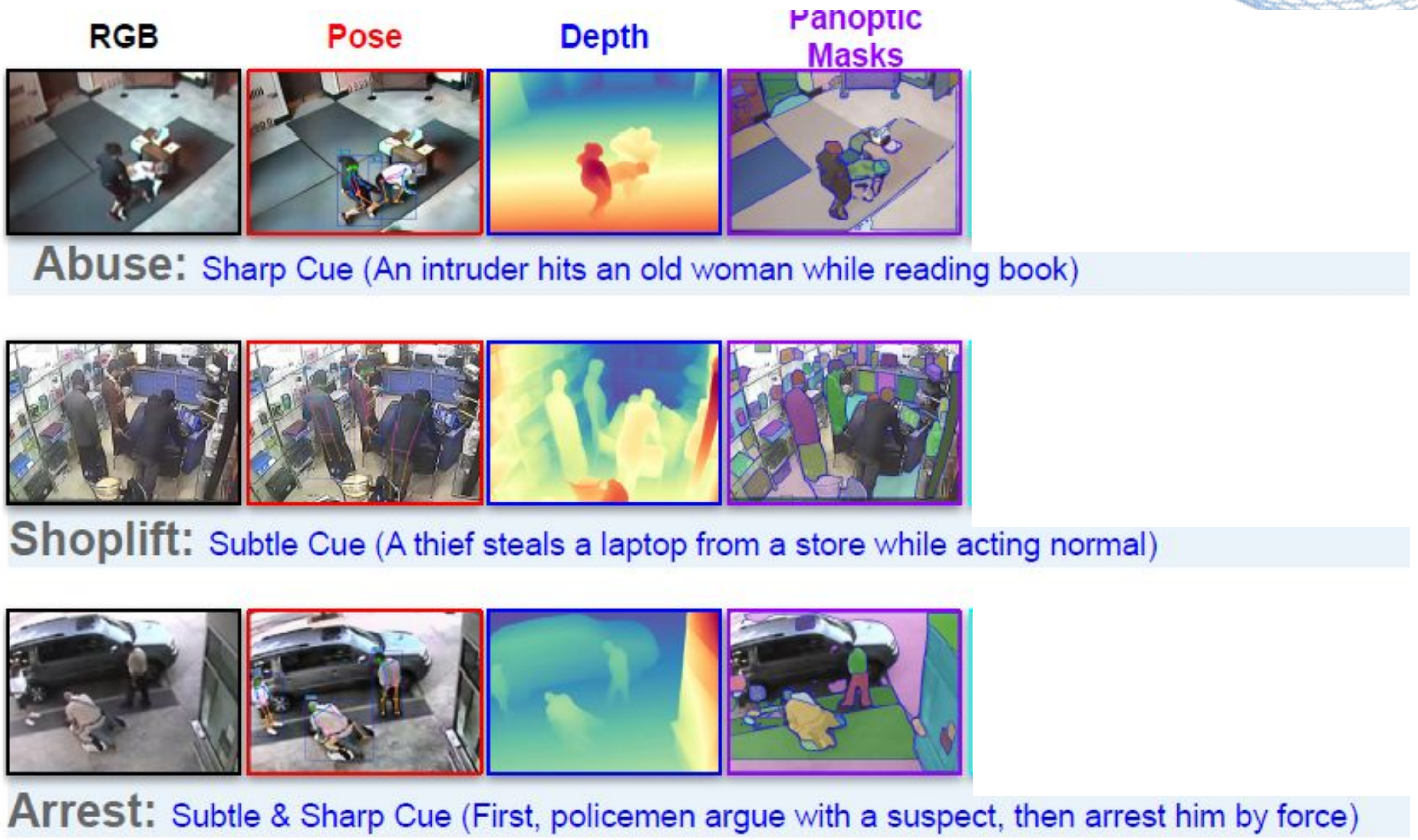
- **But how many additional modalities?**
  - **TWO**

Motivation

- RGB only features are not sufficiently distinctive enough to distinguish **complex anomalies like shoplifting and visually similar normal events**.

- **Towards robust complex real-world anomaly detection, it is essential to augment RGB with additional modalities.**

- **But how many additional modalities?**
  - **THREE**



(a) Complex real-world anomalies with multi-modal saliencies

Motivation

- RGB only features are not sufficiently distinctive enough to distinguish **complex anomalies like shoplifting and visually similar normal events**.

- **Towards robust complex real-world anomaly detection, it is essential to augment RGB with additional modalities.**

- **But how many additional modalities?**
  - **FOUR**

(a) Complex real-world anomalies with multi-modal saliencies

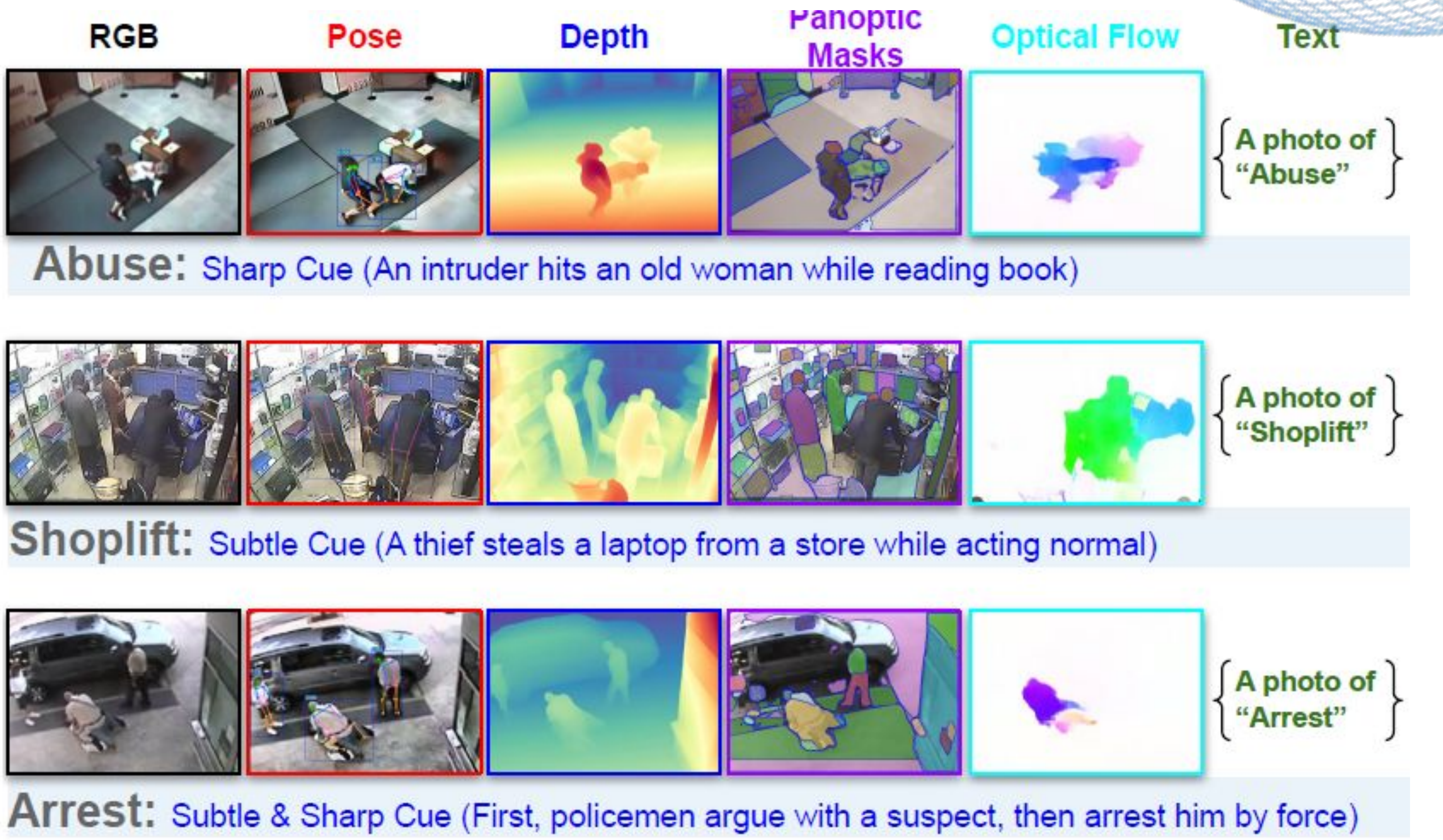
Motivation

- **RGB only features** are not sufficiently distinctive enough to distinguish **complex anomalies like shoplifting and visually similar normal events**.

- **Towards robust complex real-world anomaly detection, it is essential to augment RGB with additional modalities.**

- **But how many additional modalities?**
  - **FIVE**

We will see in this work !!



(a) Complex real-world anomalies with multi-modal saliencies

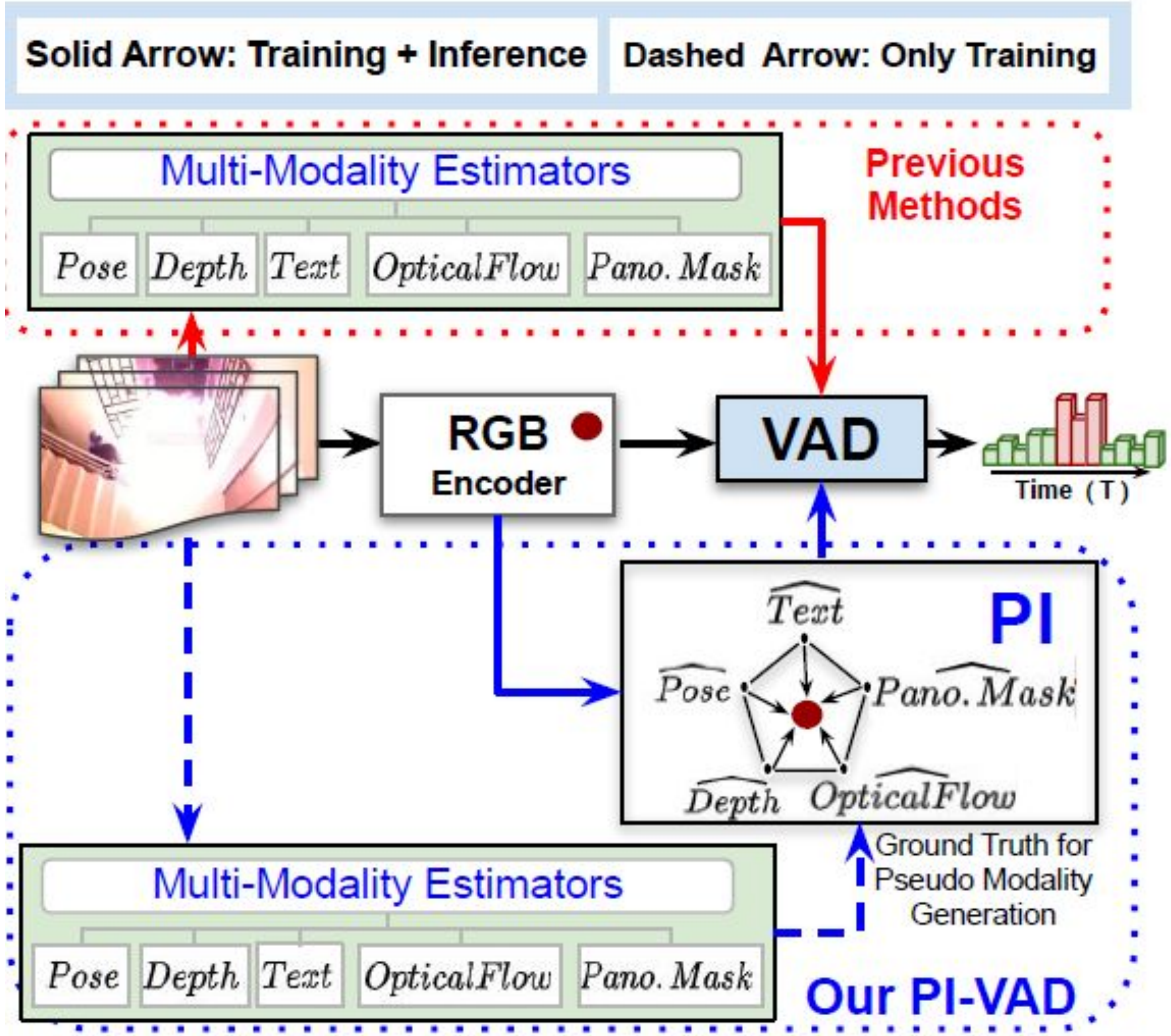
Where is the DIfficulties?

IMAGEBIND: One Embedding Space To Bind Them All

Rohit Girdhar* Alaaeldin El-Nouby* Zhuang Liu Mannat Singh
Kalyan Vasudev Alwala Armand Joulin Ishan Misra*
FAIR, Meta AI
https://facebookresearch.github.io/ImageBind

- **Difficulties arises due to:**
  - **Limited Data, Limited Supervision**
  - **Disparity Among Modalities**
  - **Noise and Redundant Information**
  - **Increased Inference Overhead**

What is our IDEA?



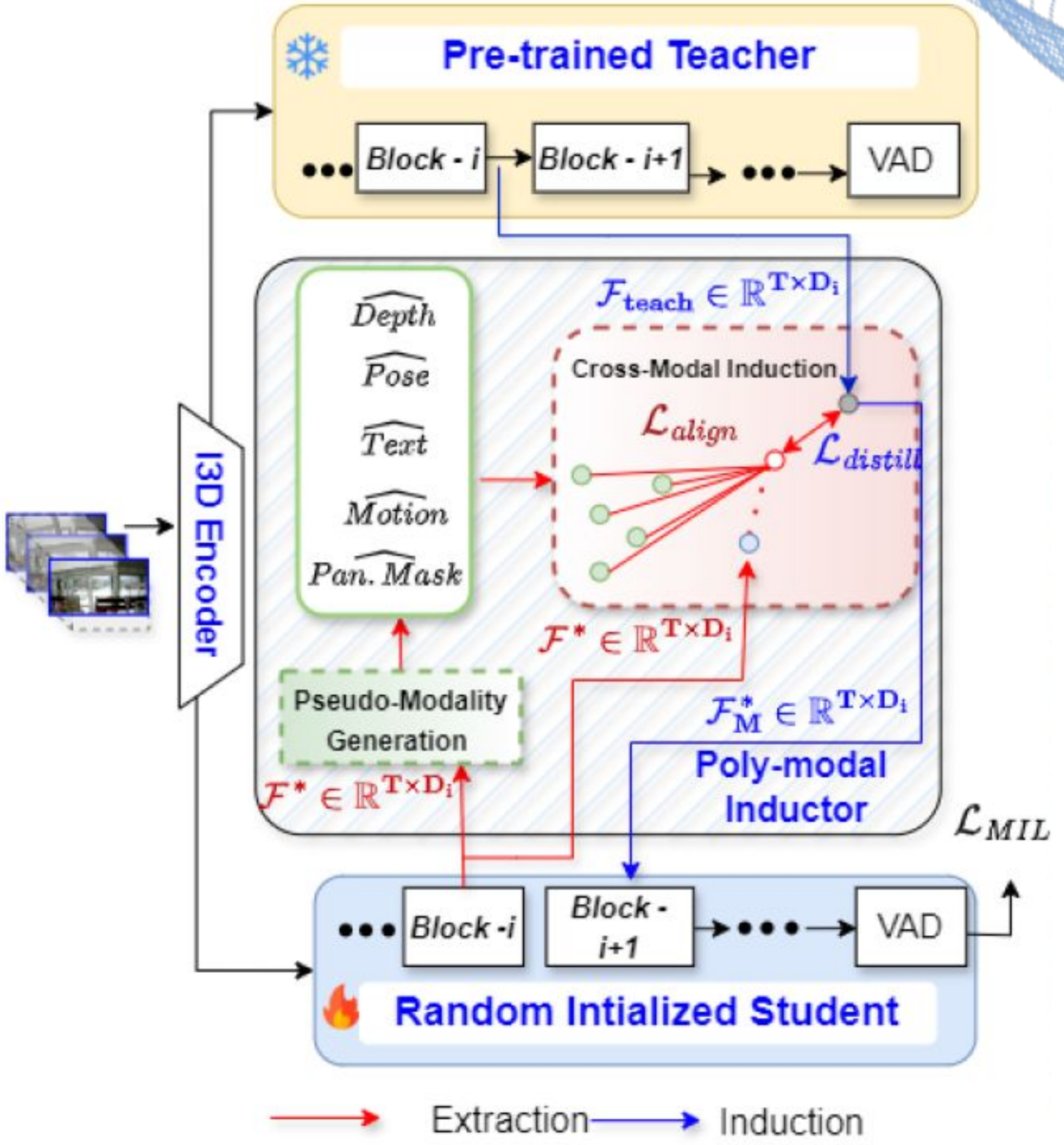
(b) Key difference between previous methods and our PI-VAD

What is the best strategy to overcome:

- **Inference Overhead: Generate Pseudo Modalities**
- **Noise and Redundancy: Task aware Generation is Necessary**
- **Disparity Among Modalities: Dissociatively binding each modality to RGB via a contrastive loss.**
- **Limited Data and Supervision: Follow a Teacher and Student Paradigm (Teacher supervise student network via Pseudo label)**

Proposed PI-VAD

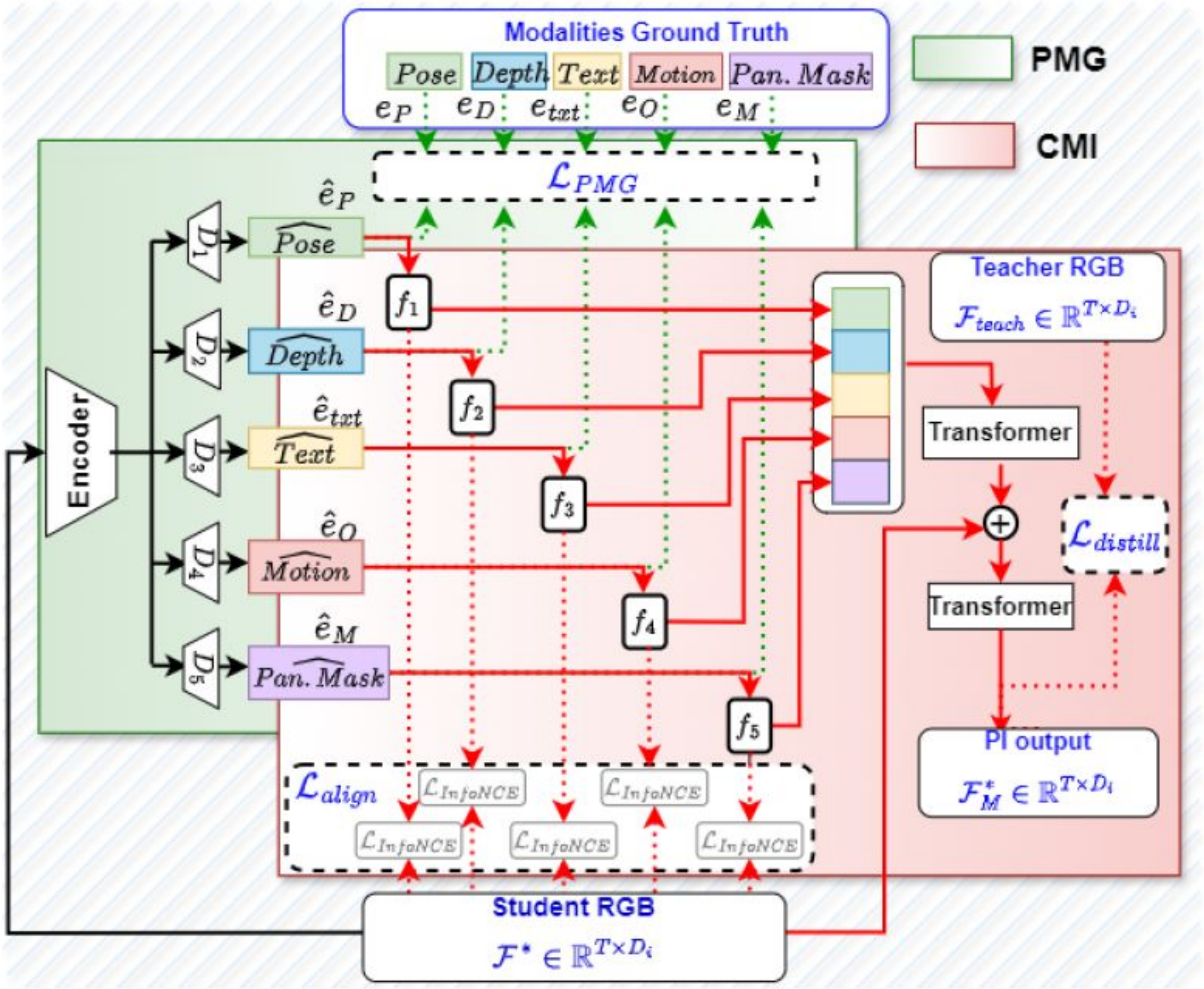
- Core of PI-VAD is a **Poly-modal Inductor**
- **PI-VAD follows Teacher-student design paradigm**
  - Teacher and Student network has Identical Functional blocks, just that
  - Teacher is Pre-trained RGB Backbone
  - Student is Random initialized
- **Poly-modal Inductor operates between the Teacher and Student, can be included at any Stage**
  - Early Stage
  - Later Stage
- **Teacher guides the poly-modal Inductor** by providing the coarse anomaly representation
- Thanks to Poly-modal inductor **Student learn the fine-grained anomaly Representation**



(a) Overview of π-VAD

Poly-modal Inductor

- Two Functional Modules of Poly-modal Inductor
  - **PMG (Pseudo Modality Generation)**
  - **CMI (Cross Modality Induction)**
- **PMG generates modality specific prototype embeddings directly from latent RGB embedding.**
- **PMG learns the anomaly relevant synthetic approximation of actual modalities.**
- **CMI aligns uncoupled modalities within a unified, RGB-anchored embedding space.**
- **CMI facilitates the semantic alignment between the multi-modal encodings from PMG and the RGB embeddings of the student while ensuring that the alignment is pertinent to WSVAD.**



(b) Poly-modal Inductor (PI)

PI-VAD Optimization

$$\mathcal{L}_{PMG} = \sum_{j=1}^{5} \frac{1}{d_j} \sum_{k=1}^{d_j} (\hat{e_{j,k}} - e_{j,k})^2, \text{where } j \in \{\text{P, D, M, O, txt}\}.$$

$$\mathcal{L}_{InfoNCE} = -\frac{1}{T} \sum_{i=1}^{T} \log \frac{\exp\left(\frac{sim(\mathcal{F}^*(T_i), \hat{e_j}(T_i))}{\tau}\right)}{\sum_{k=1, i \neq k}^{T} \exp\left(\frac{sim(\mathcal{F}^*(T_i), \hat{e_j}(T_k))}{\tau}\right)} \tag{2}$$

$$\mathcal{L}_{align} = \sum_{i=1}^{5} \mathcal{L}_{InfoNCE}, \quad i \in \{\text{P, D, M, O, txt}\} \tag{3}$$

$$\mathcal{L}_{distill} = \frac{1}{D_i} \sum_{k=1}^{D_i} (\mathcal{F}^*_{Mk} - \mathcal{F}_{teachk})^2.$$

Enough Action/Anomaly Detection !! Let’s See the Future Action/Anomaly

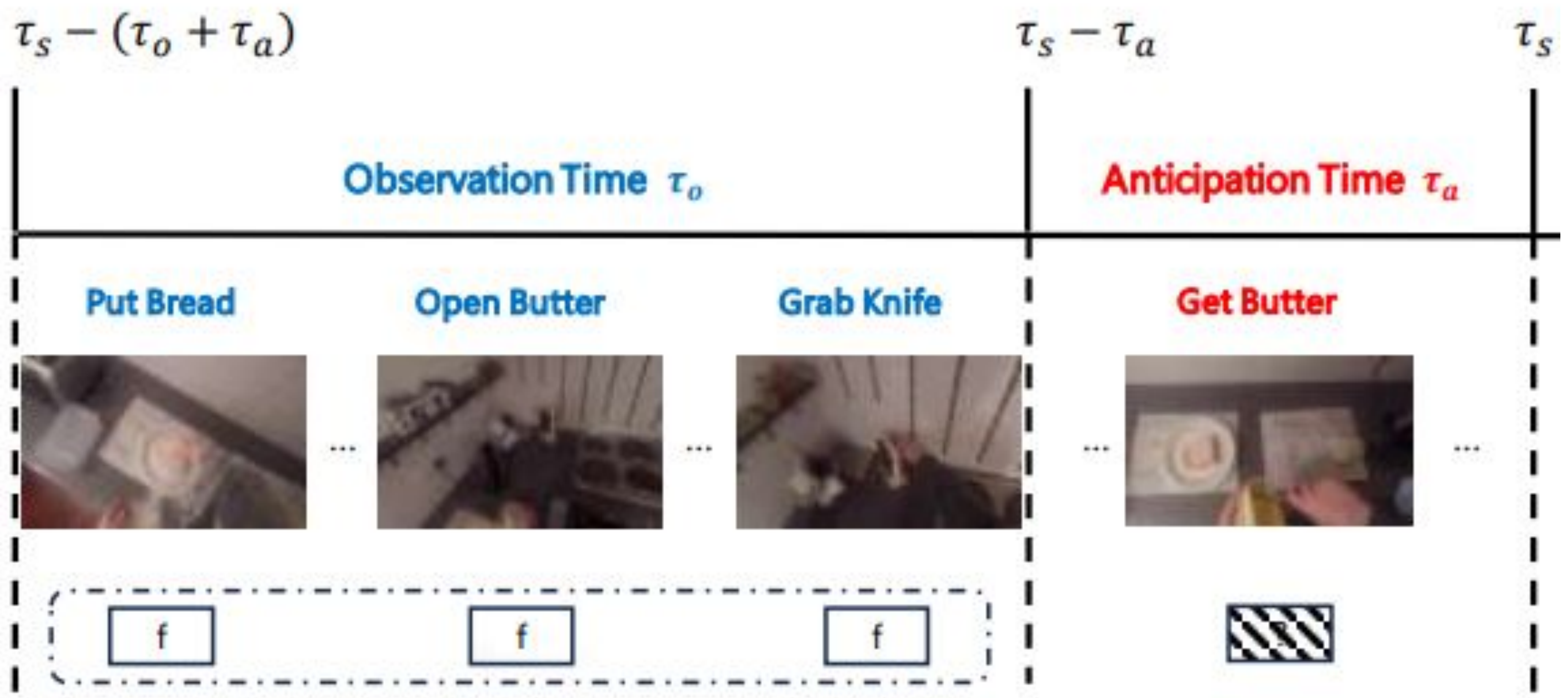
Action Anticipation

What is Action Anticipation?

- Predicting future human actions from partial or ongoing observations.

Why is it challenging?

- Partial observations.
- Ambiguity in actions.
- Temporal variability.
- Uncertainty between observation and future event



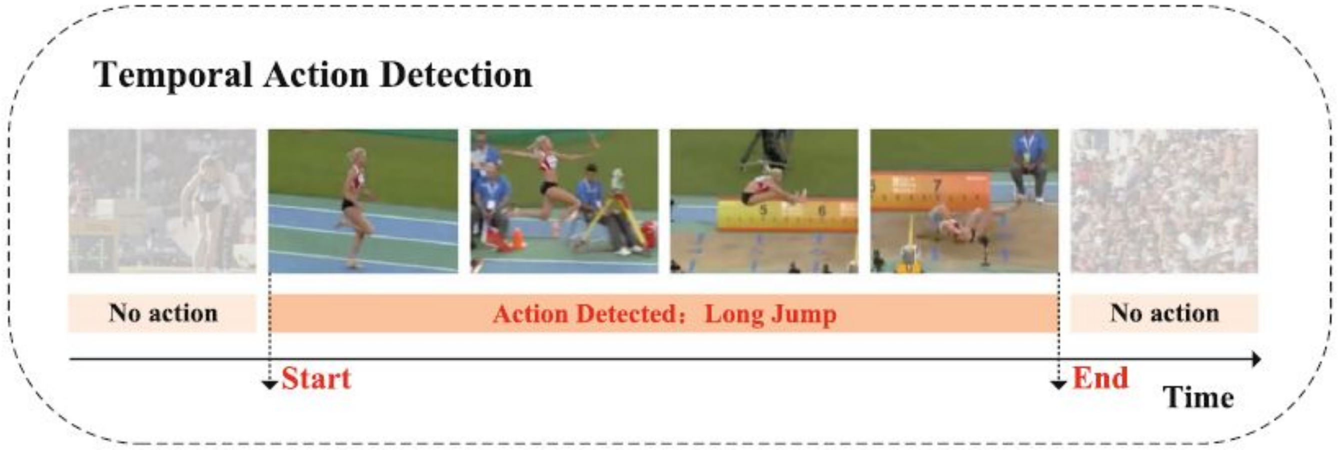
Applications

- Autonomous vehicles: Predict pedestrian or driver behavior.
- Surveillance: Identify potentially harmful actions early.
- Healthcare: Anticipate falls or movements in elderly care.
- Human-robot interaction: React to human actions in collaborative settings.

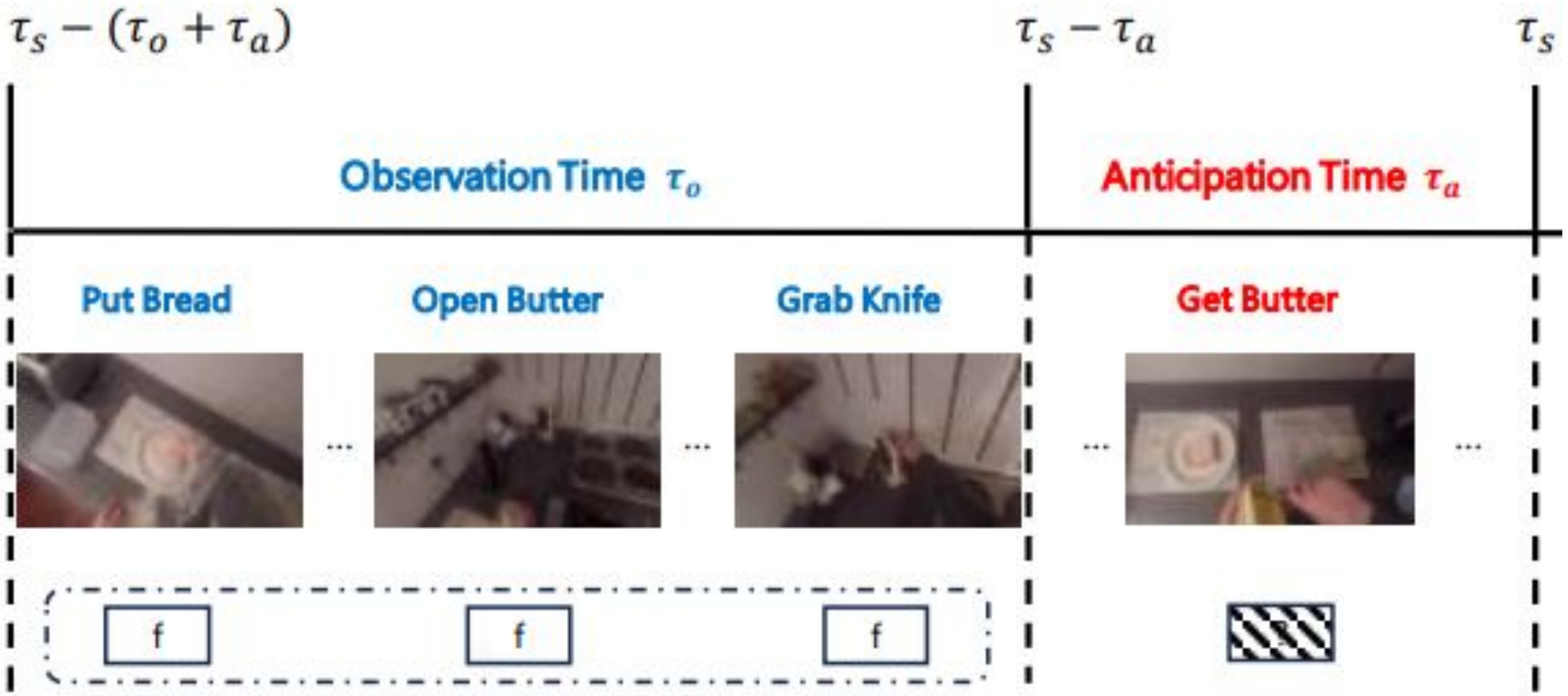
Action Classification

CLASS	SCORE
Swing Dancing	0.52
Salsa Dancing	0.39
Holding Hands	0.03
Walking	0.01
....	...

Action Detection



Action Anticipation



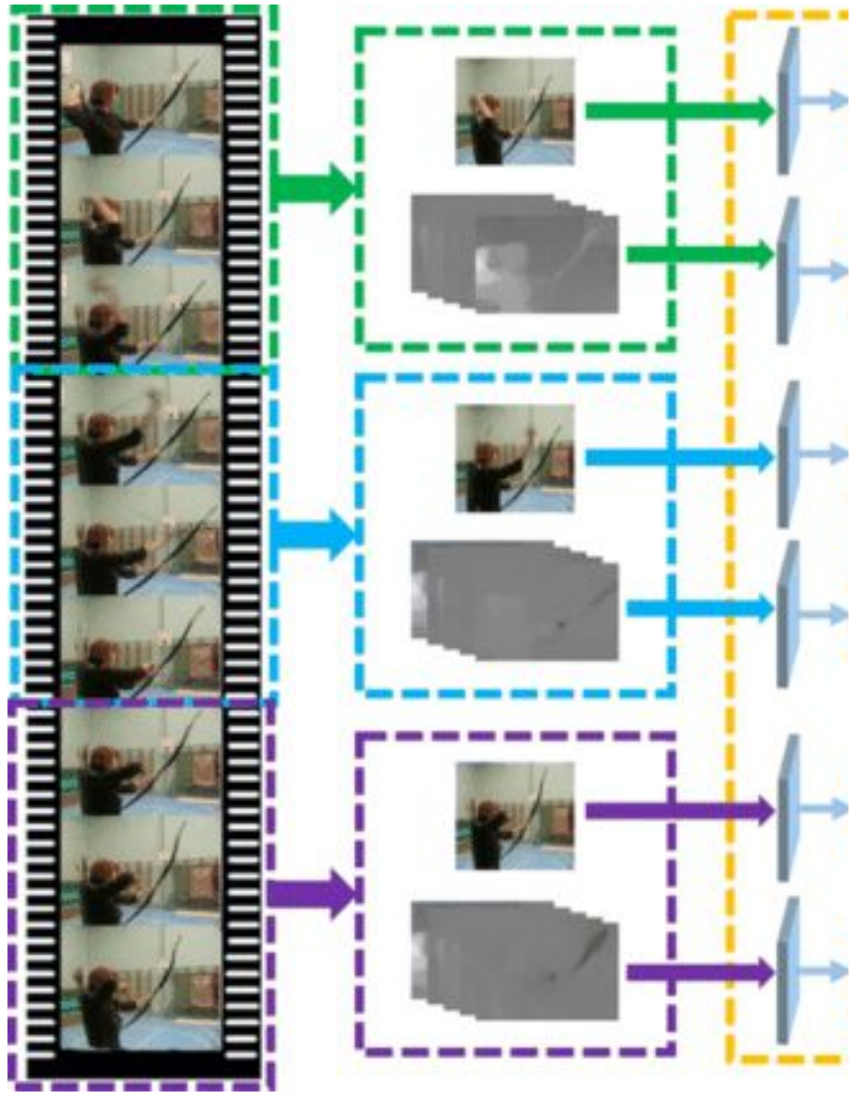
STEPS in Action Anticipation

Step-1

Observation Frames Feature Extraction:

-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.

-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.



STEPS in Action Anticipation

Step-1

Observation Frames Feature Extraction:

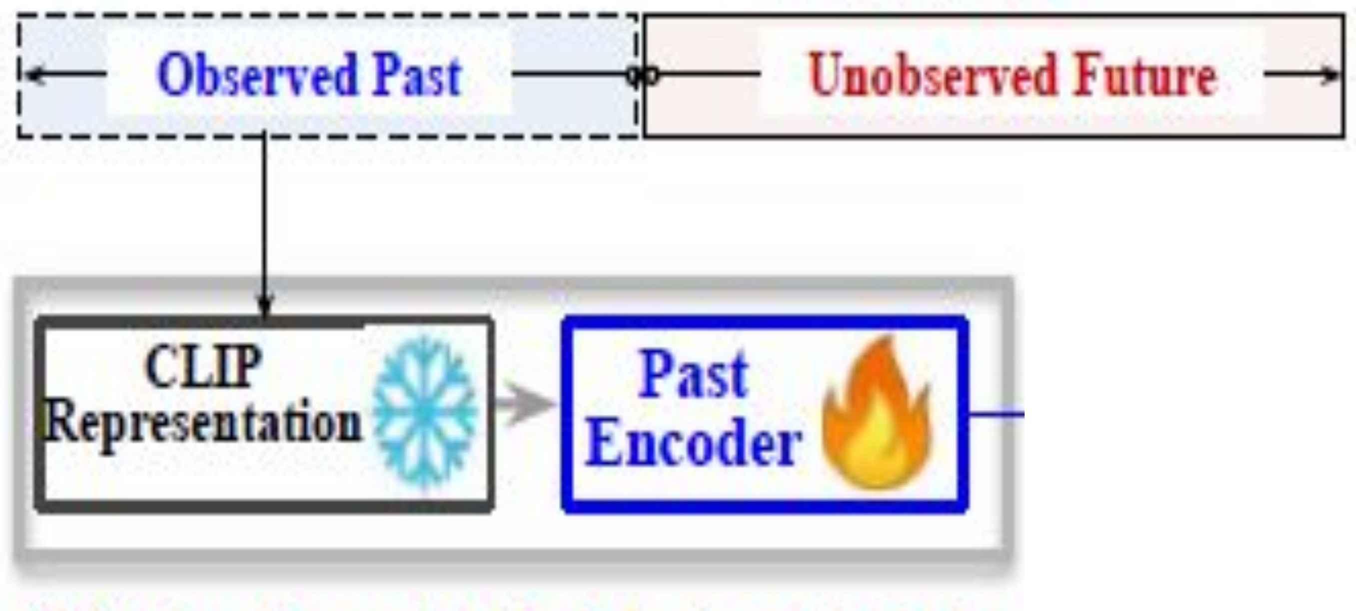
-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.

-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.

Step-2

Past Encoder:

--With LSTM, GRU, TCN or Transformers encode the spatio-temporal dynamics of the past observation.



STEPS in Action Anticipation

Step-1

Observation Frames Feature Extraction:

-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.

-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.

Step-2

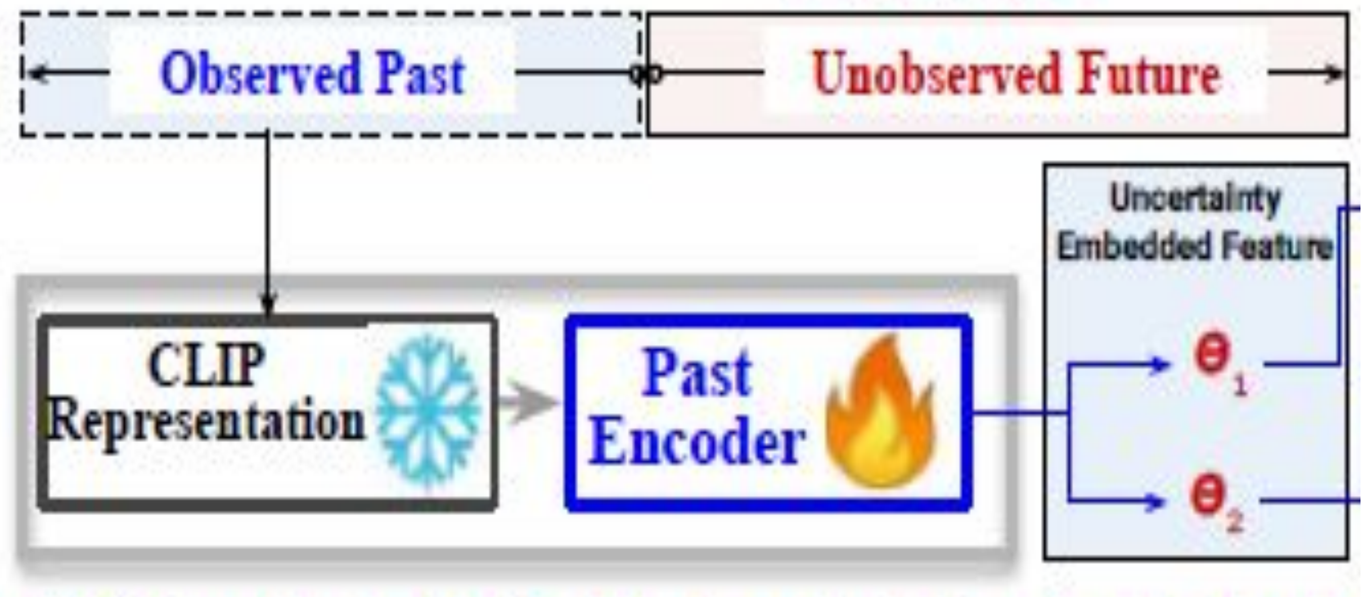
Past Encoder:

--With LSTM, GRU, TCN or Transformers encode the spatio-temporal dynamics of the past observation.

Step-3

Uncertainty Encoder:

--[Optional] Modify the latent space of observation to embed the uncertainty associated with the future



STEPS in Action Anticipation

Step-1

Observation Frames Feature Extraction:

-- Use pre-trained models (e.g., ResNet, I3D, or SlowFast networks) to extract meaningful spatio-temporal features.

-- Features can include RGB (appearance) and multiple modalities like optical flow, pose, depth etc.

Step-2

Past Encoder:

--With LSTM, GRU, TCN or Transformers encode the spatio-temporal dynamics of the past observation.

Step-3

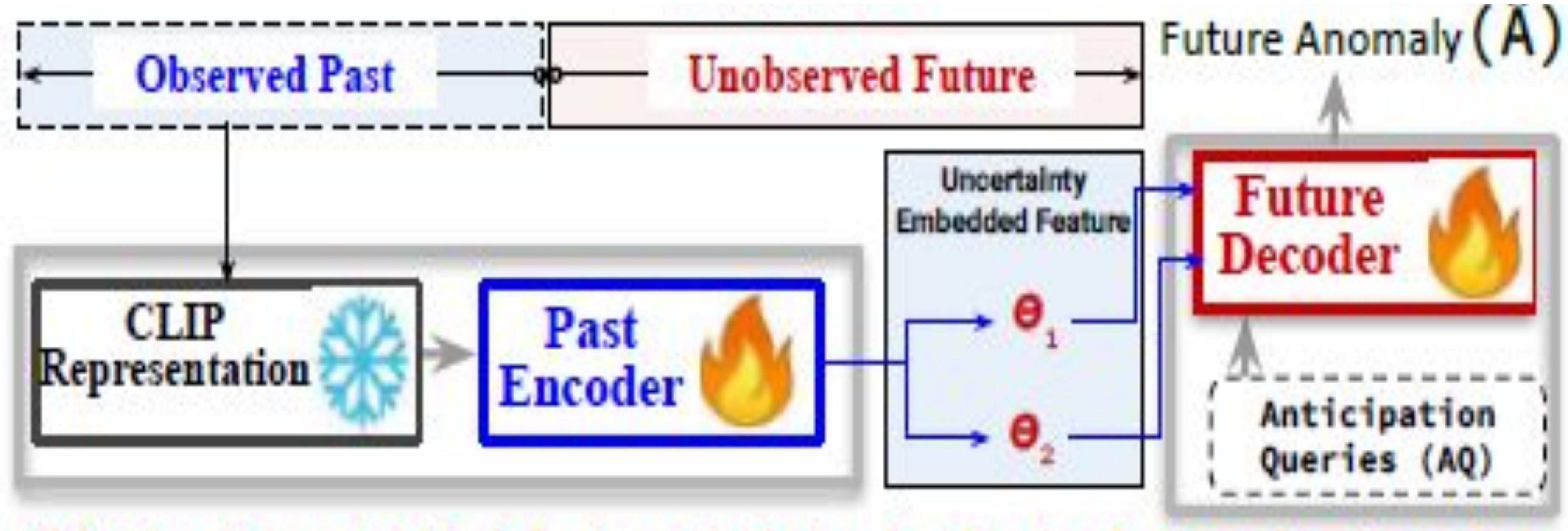
Uncertainty Encoder:

--[Optional] Modify the latent space of observation to embed the uncertainty associated with the future

Step-4

Future Decoder:

-- Predict the next action based on observed features.
-- Predict intermediate states (e.g., subactions). First predict intermediate steps, then the final action.



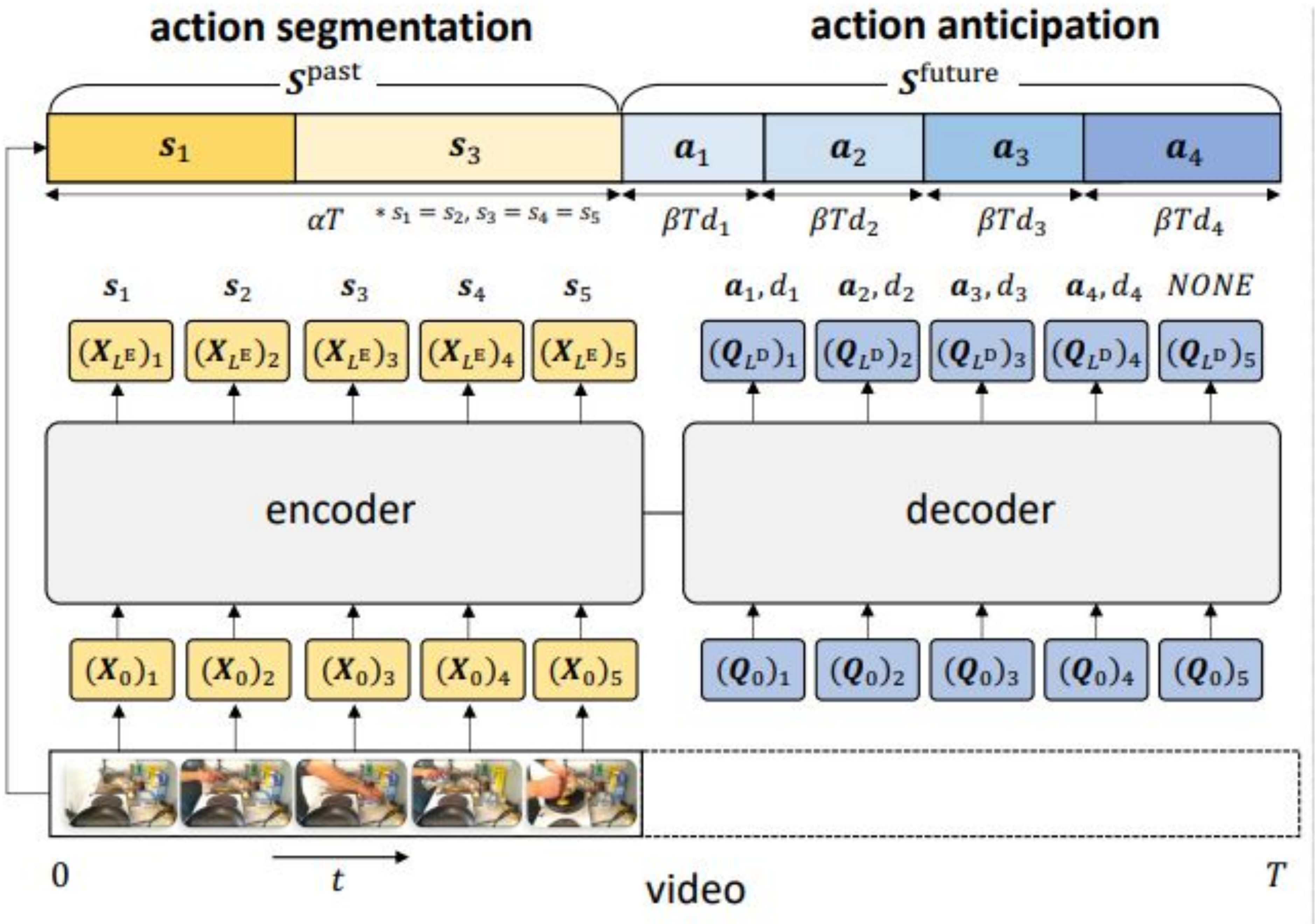
Let’s See Few TOP Action Anticipation Methods , Excited !

FUTR:

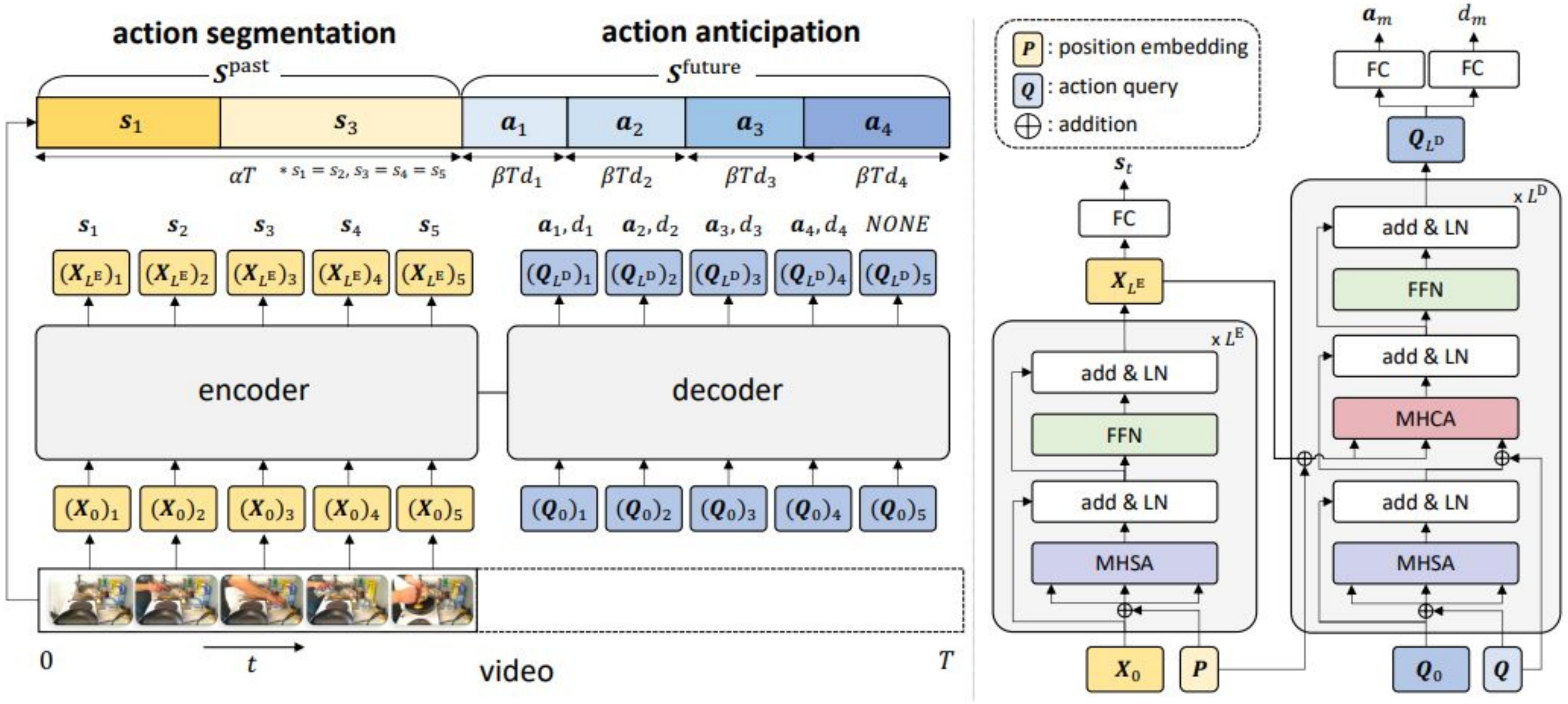
- It is an an end-to-end attention neural network to anticipate actions in parallel decoding, leveraging global interactions between past and future actions for long-term anticipation.

- FUTR is composed of an encoder and a decoder; each classifies action labels of past frames (action segmentation) and anticipates future action labels and corresponding durations (action anticipation), respectively.

- The encoder learns distinctive feature representation from past actions via self-attention, and the decoder learns long-term relations between past and future actions via self-attention and cross-attention.



FUTR:



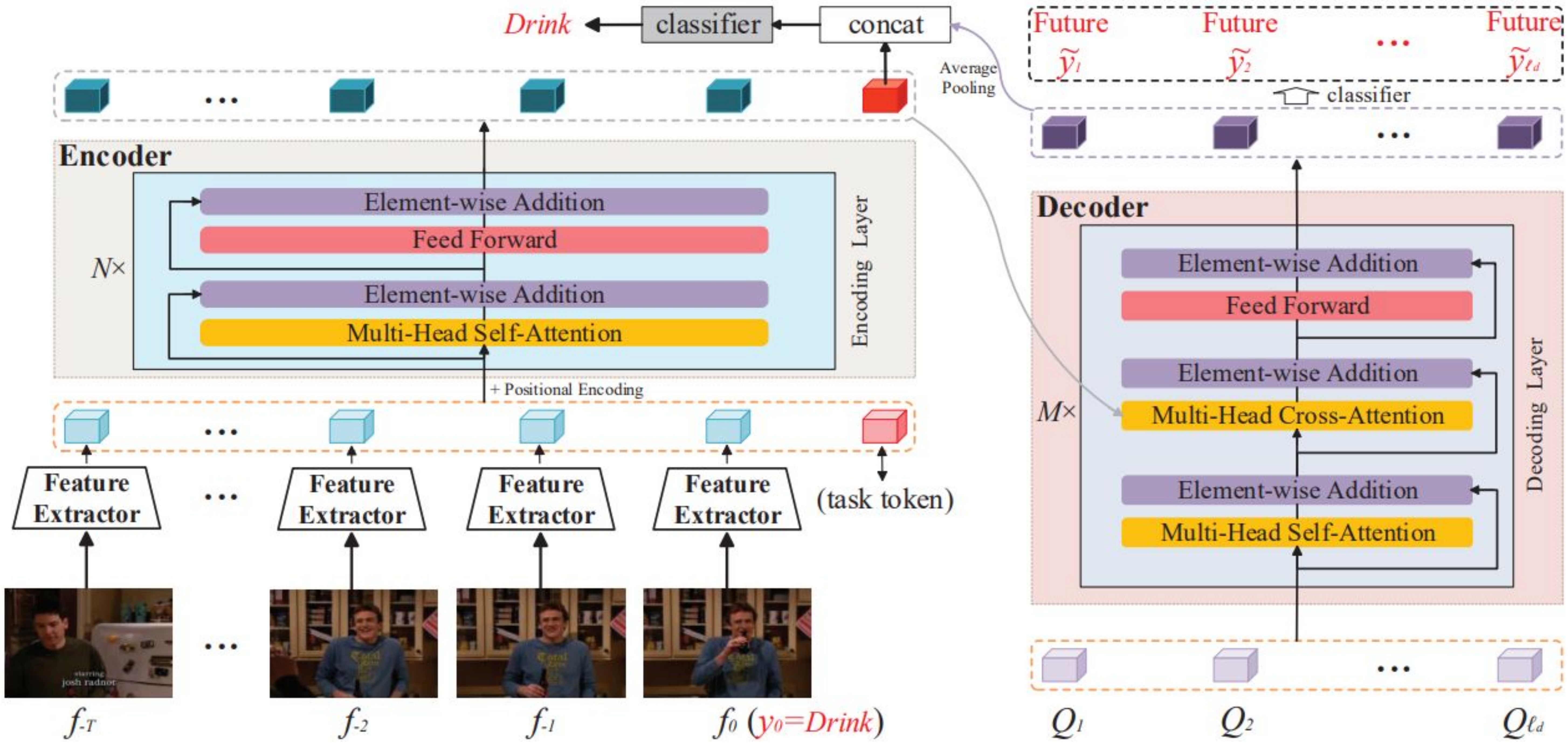
Key Contributions:

1. **End-to-End Attention Network:** FUTR employs a Transformer-based architecture that captures fine-grained temporal relations among observed frames, facilitating effective long-term action anticipation.

2. **Parallel Decoding:** Unlike traditional autoregressive models that predict future actions sequentially, FUTR predicts the entire sequence of future actions in parallel. This parallel decoding approach enhances both the accuracy and speed of inference, mitigating potential error accumulation inherent in sequential predictions.

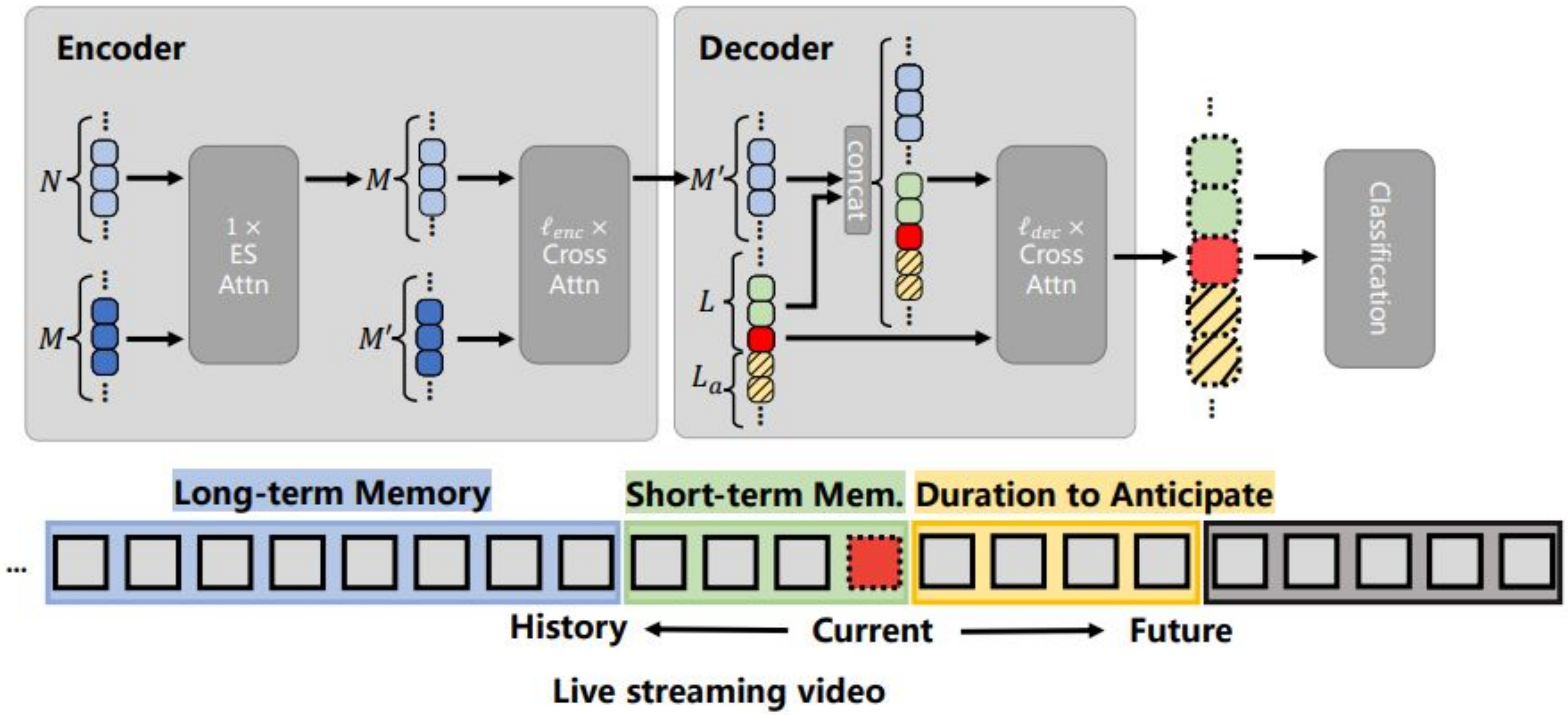
3. **Integrated Action Segmentation Loss:** The model incorporates an action segmentation loss during training to learn distinctive feature representations in the encoder. This integration ensures that the encoder captures meaningful temporal features, improving the overall anticipation performance.

OADTR:



Drink
classifier
concat
Average Pooling
Future
Future
Future
classifier
Encoder
Element-wise Addition
Feed Forward
N×
Element-wise Addition
Multi-Head Self-Attention
Encoding Layer
+ Positional Encoding
Feature Extractor
Feature Extractor
Feature Extractor
Feature Extractor
(task token)
f_{-T}
f_{-2}
f_{-1}
f_0 (y_0=Drink)
Decoder
Element-wise Addition
Feed Forward
Element-wise Addition
Multi-Head Cross-Attention
M×
Element-wise Addition
Multi-Head Self-Attention
Decoding Layer
Q_1
Q_2
Q_{ℓ_d}

TesTra: (Memory-based)



Separation of Long and Short-Term Memories:

- By explicitly dividing the entire history into long-term and short-term memories, it effectively captures temporal relations over prolonged sequences while retaining fine granularity of events.

- **Encoder:** Compresses and abstracts long-term memory by processing an extended temporal window (e.g., 2048 frames spanning up to 8 minutes), capturing coarse-scale historical information.

- **Decoder:** Focuses on a short-term memory window (e.g., 32 frames spanning 8 seconds), modeling fine-scale characteristics through self-attention and cross-attention mechanisms.

WACV'25 [ORAL] ACCEPTED

Let's See our Work

Guess Future Anomalies from Normalcy: Forecasting Abnormal Behavior in Real-World Videos

Snehashis Majhi[1,2,*], Mohammed Guermal[1,2,*], Antitza Dantcheva[1,2], Quan Kong[3], Lorenzo Garattoni[4], Gianpiero Francesca[4], François Brémond[1,2]
[1] INRIA [2] Côte d'Azur University [3] Woven by Toyota [4] Toyota Motor Europe
* Joint first authors.

Motivation

- **Offline or Online Anomaly Detection:** **Provides Investigative or Timely Intervention**
- **Future Anomaly Prediction:** **Provides Anomaly Preventive Measures (High Societal Impact)**
- **Is it possible to predict all future anomalies? NOT ALL** but yes for Abnormal Human Behaviour
- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.
- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**
- **But how far can you predict the future?**
  - **next 1 second (30 frames)**

(a) Notice the various interaction in Abnormal Human behaviours

Motivation

- **Offline or Online Anomaly Detection: Provides Investigative or Timely Intervention**

- **Future Anomaly Prediction: Provides Anomaly Preventive Measures (High Societal Impact)**

- **Is it possible to predict all future anomalies? NOT ALL** but yes for Abnormal Human Behaviour

- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.

- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**

- **But how far can you predict the future?**
  - **next 2 seconds (60 frames)**

(a) Notice the various interaction in Abnormal Human behaviours

Motivation

- **Offline or Online Anomaly Detection:** **Provides Investigative or Timely Intervention**

- **Future Anomaly Prediction:** **Provides Anomaly Preventive Measures (High Societal Impact)**

- **Is it possible to predict all future anomalies? NOT ALL** but yes for Abnormal Human Behaviour

- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.

- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**

- **But how far can you predict the future?**
  - **next 3 seconds (90 frames)**

(a) Notice the various interaction in Abnormal Human behaviours

Motivation

- **Offline or Online Anomaly Detection:** **Provides Investigative or Timely Intervention**

- **Future Anomaly Prediction:** **Provides Anomaly Preventive Measures (High Societal Impact)**

- **Is it possible to predict all future anomalies?** **NOT ALL** but yes for Abnormal Human Behaviour

- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.

- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**

- **But how far can you predict the future?**
  - **next 4 seconds (120 frames)**

(a) Notice the various interaction in Abnormal Human behaviours

Motivation

- **Offline or Online Anomaly Detection: Provides Investigative or Timely Intervention**

- **Future Anomaly Prediction: Provides Anomaly Preventive Measures (High Societal Impact)**

- **Is it possible to predict all future anomalies? NOT ALL** but yes for Abnormal Human Behaviour

- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.

- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**

- **But how far can you predict the future?**
  - **next 5 seconds (150 frames)**

(a) Notice the various interaction in Abnormal Human behaviours

Motivation

- **Offline or Online Anomaly Detection: Provides Investigative or Timely Intervention**
- **Future Anomaly Prediction: Provides Anomaly Preventive Measures (High Societal Impact)**

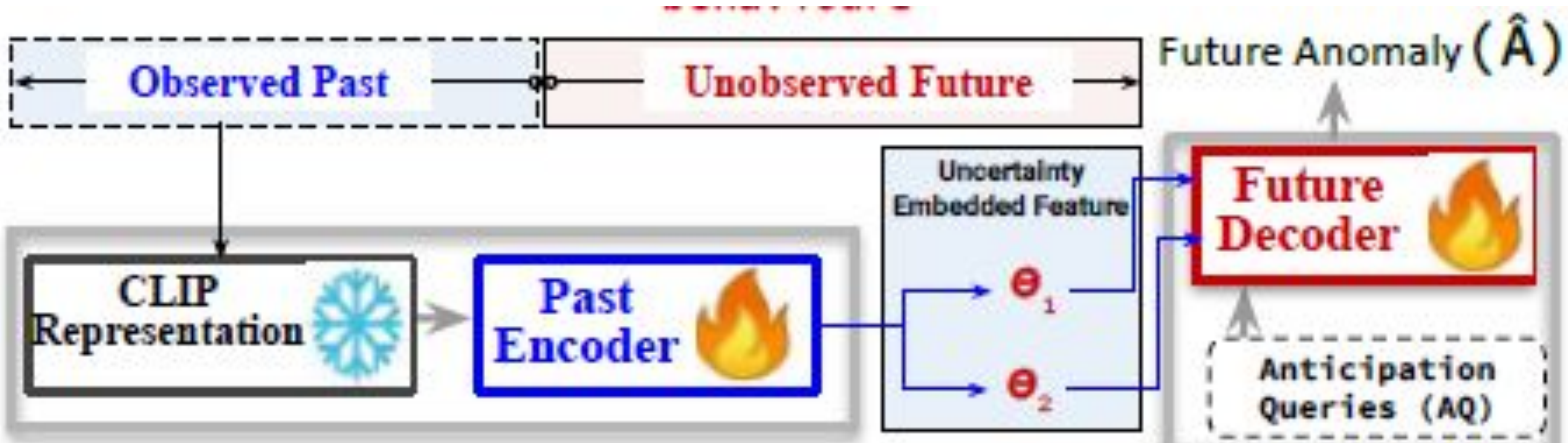
- **Is it possible to predict all future anomalies? NOT ALL** but yes for Abnormal Human Behaviour
- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.
- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**
- **But how far can you predict the future?**
  - **next 8 seconds (240 frames)**

(a) Notice the various interaction in Abnormal Human behaviours

Motivation

- **Offline or Online Anomaly Detection: Provides Investigative or Timely Intervention**
- **Future Anomaly Prediction: Provides Anomaly Preventive Measures (High Societal Impact)**

- **Is it possible to predict all future anomalies? NOT ALL** but yes for Abnormal Human Behaviour
- **Why Abnormal Human Behaviour?** Bcz human interacts with the surrounding agents like objects , other living entity etc.
- **So predicting future abnormal human behaviour is possible by carefully analysing the early trends of their interactions.**
- What about **Uncertainty** between observation and future event?



(b) Our Idea to Model Uncertainty in Typical Encoder-Decoder Framework

Detection Vs. Anticipation

Anomaly Anticipation can Answer Questions like:

- **Whether an anomaly will occur in the near future? (Short Anticipation)**
- **If yes, What kind of anomaly is likely to occur? (Anomaly Class: One can guess the seriousness of anomaly)**
- **Is there a chance of re-occurrence of the same anomaly in a future time window? (Long Anticipation).**

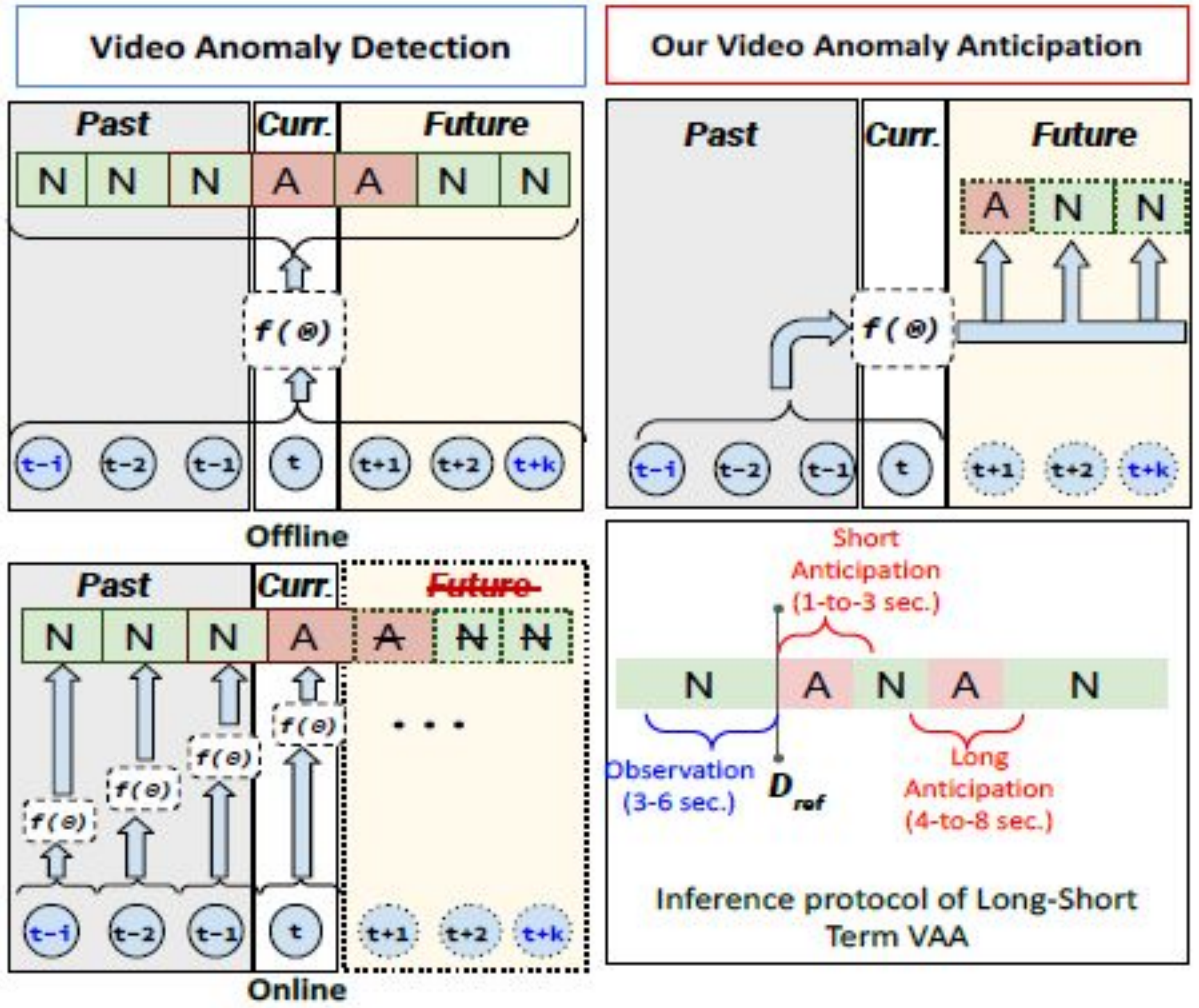


Figure 2. Illustration VAD Vs. VAA: Suppose the current time step is t. For **online VAD**, a parametrized model $f(\theta)$ can predict normal (N) or anomaly (A) for the current t based on observed time stamps $t-i \ldots t-1, t$, where i represents the observed duration. However, for **our VAA** we predict what kind of anomaly will occur in the future in a range of $[\,t+1,\ t+2, \ldots,\ t+k]$ where k represents anticipation duration. Further, we comprehend the short and long-term anticipation to identify the potential re-occurrence of an anomaly in the long future.

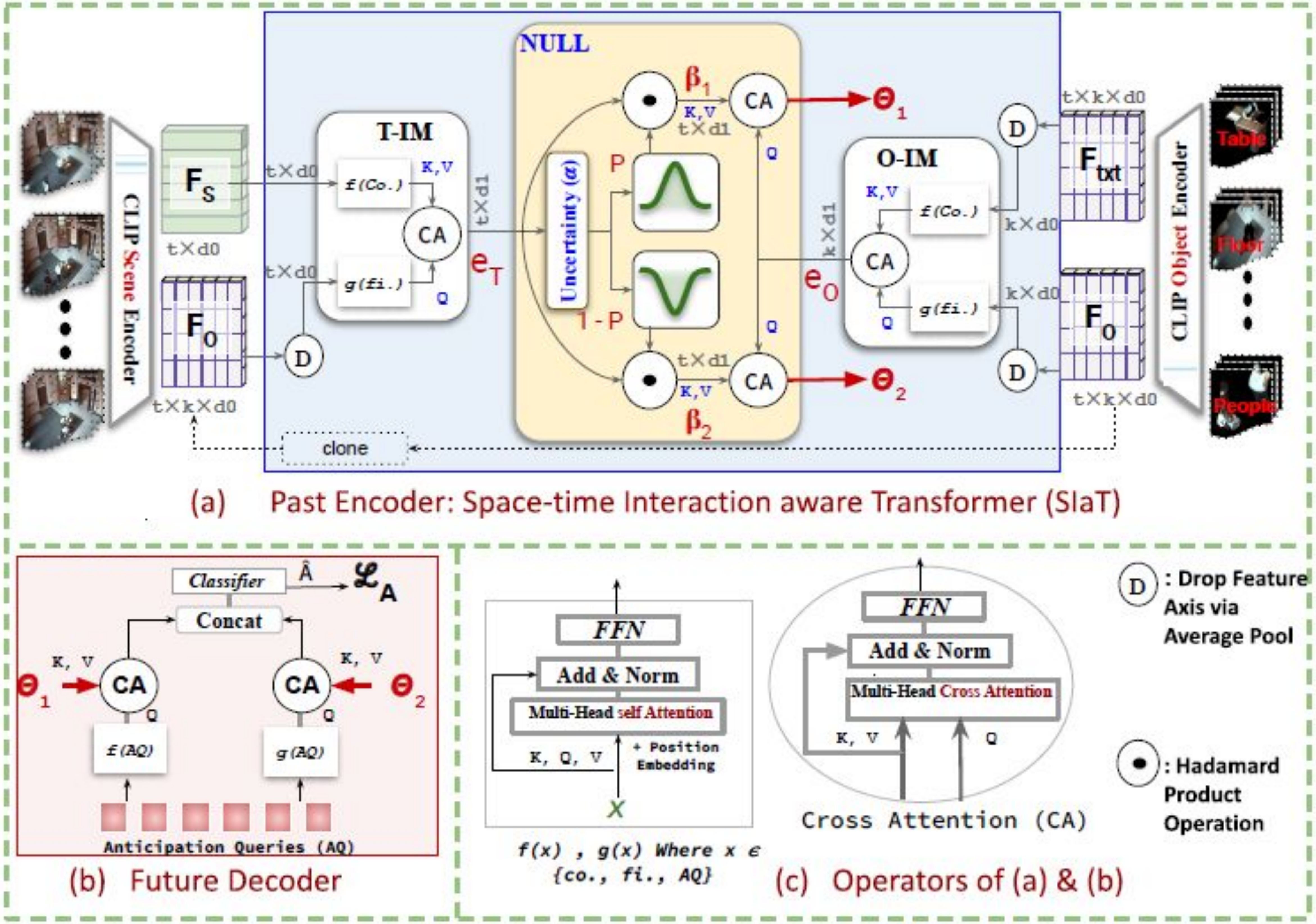
SlaT:

- Two Key Modules of SlaT:
  - **Interaction Modules (T/O-IM)**
  - **Normalcy Uncertainty Latent Learner (NULL)**

- **T/O-IM constitutes two identical modules** with different functionalities,
  - **Temporal Interaction Module (TIM) and Object Interaction Module (OIM)** to dissociatively capture the scene-level global temporal interactions and object-level local spatial interaction.

- **NULL** associates the interaction encoded scene and object semantics by exploiting the inherent uncertainty associated with normal observation to future AHB.

- **NULL** adjusts the flow of information from the past encoder to the future decoder by learning latent features that are aligned with future predictions.

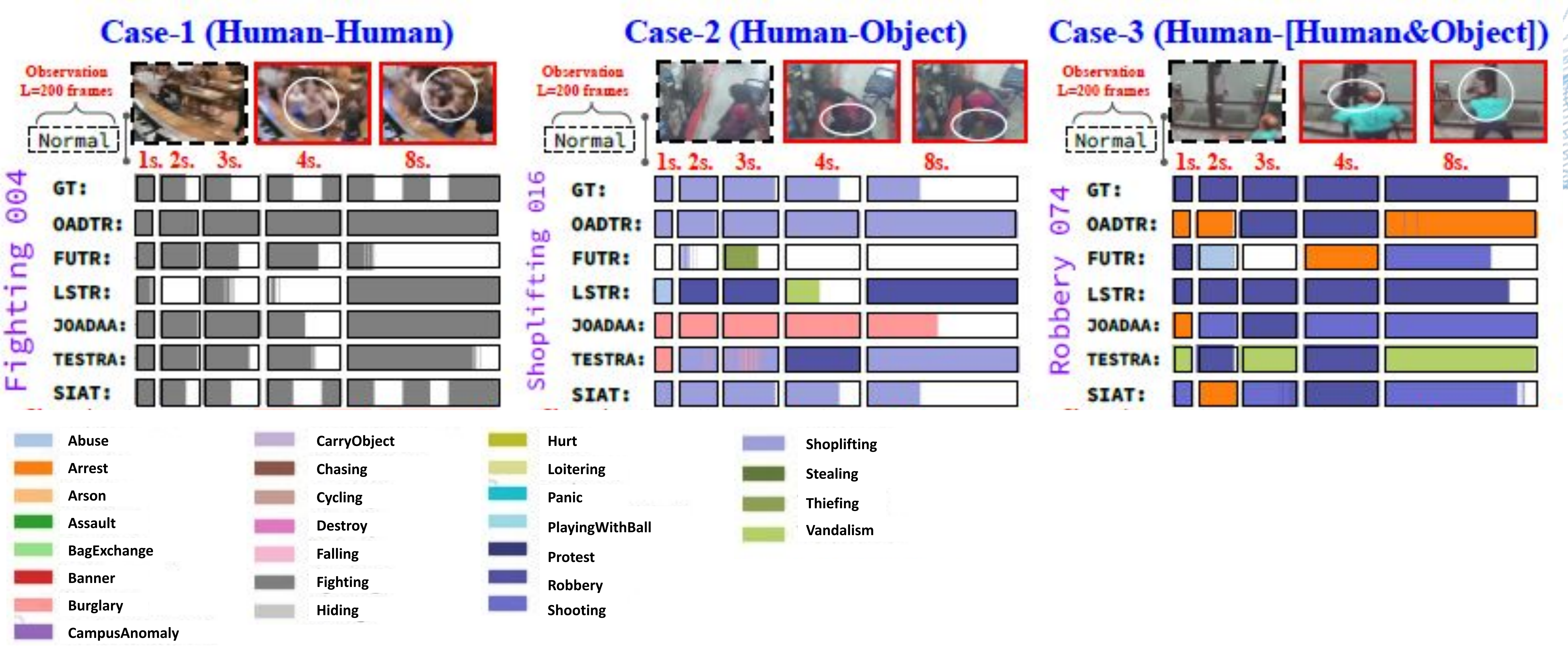


(a) Past Encoder: Space-time Interaction aware Transformer (SlaT)

(b) Future Decoder

(c) Operators of (a) & (b)

Qualitative Results



SIaT is effective in most scenarios but faces challenges in highly complex or ambiguous cases.

My Supervisors and Collaborators


Srijan Das


Rui Dai


Quan Kong


Lorenzo Garattoni


Gianpiero Francesca


François Brémond


Michal Balazia


Antitza Dantcheva


Mohammed Greumal


Giacomo D'Amicantonio

If you gain some interest in our works, do reach me out for further and detailed explanation.


THANK YOU