Leaving Some Stones Unturned:
Dynamic Feature Prioritization for Activity Detection in Streaming Video

Yu-Chuan Su and Kristen Grauman
The University of Texas at Austin

1. Problem

Goal: detect activity in online video streams with per time-step computation budget (# of features extracted).

Current activity recognition strategies

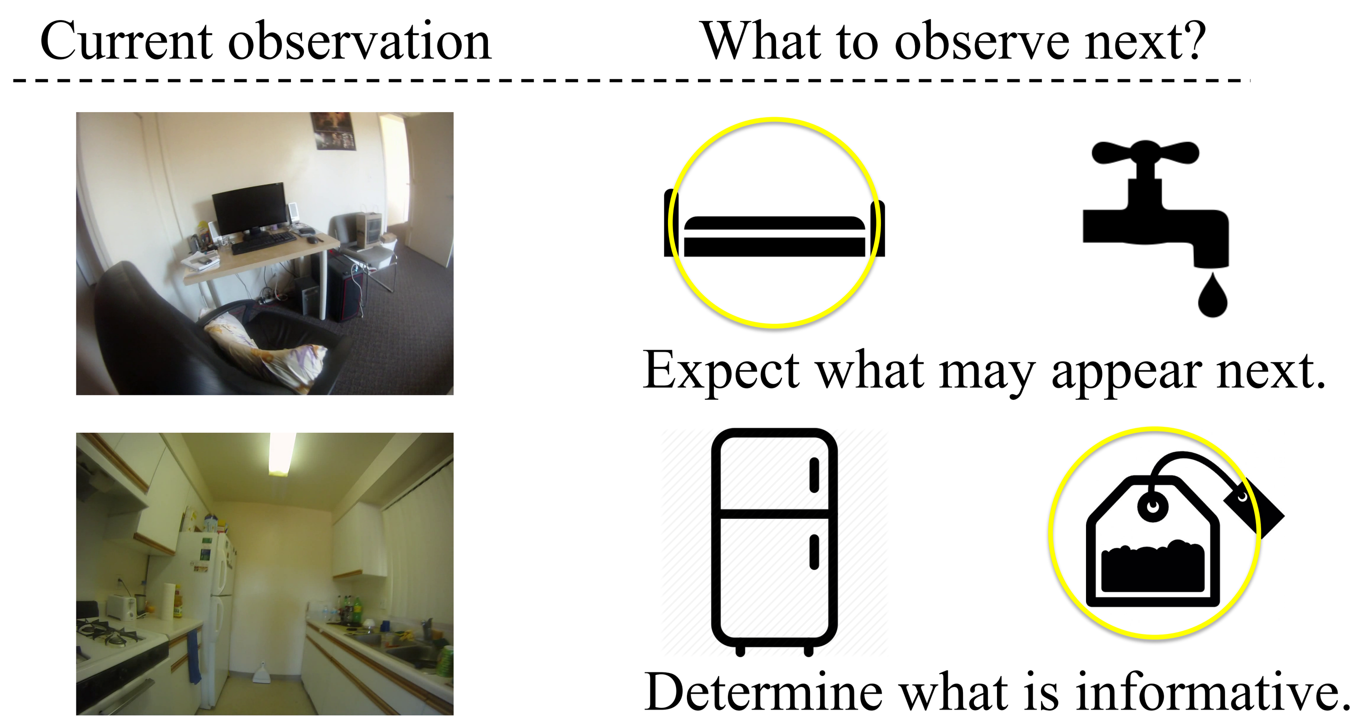
- Offline processing – assume full access to the entire video
- Unlimited resource – compute as many features as possible

Challenges

- Online video stream – can't perform random access
- Computation budget– can't enumerate all possible features

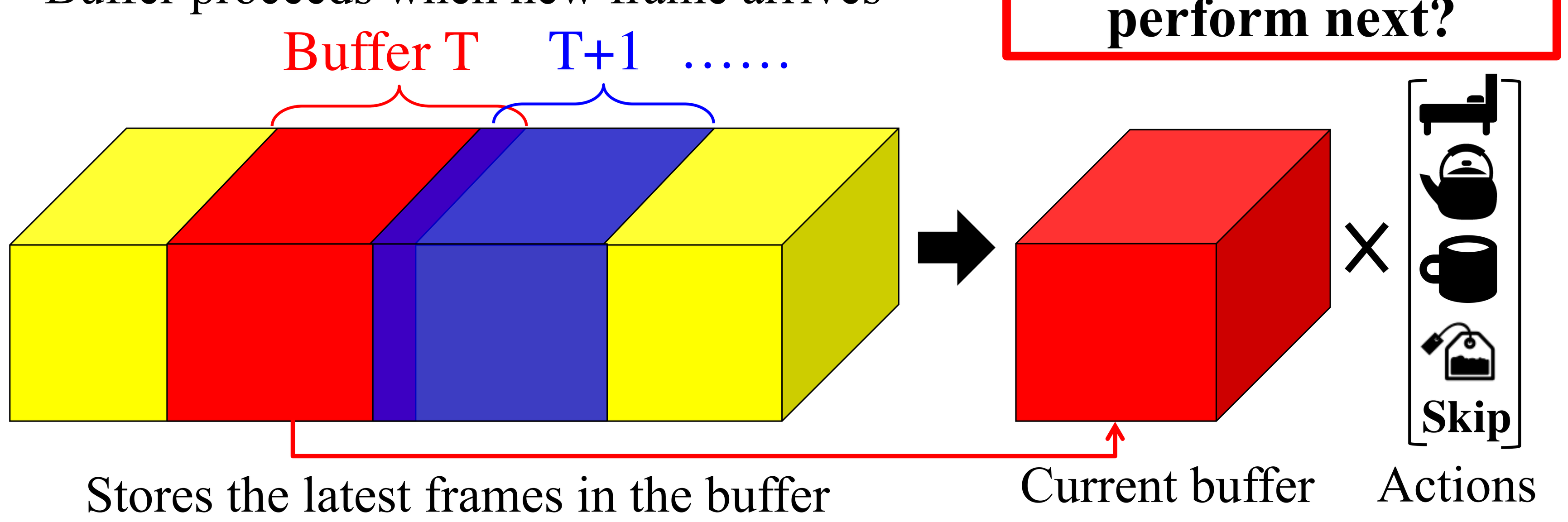
2. Proposed Solution

Select when and what to compute intelligently!



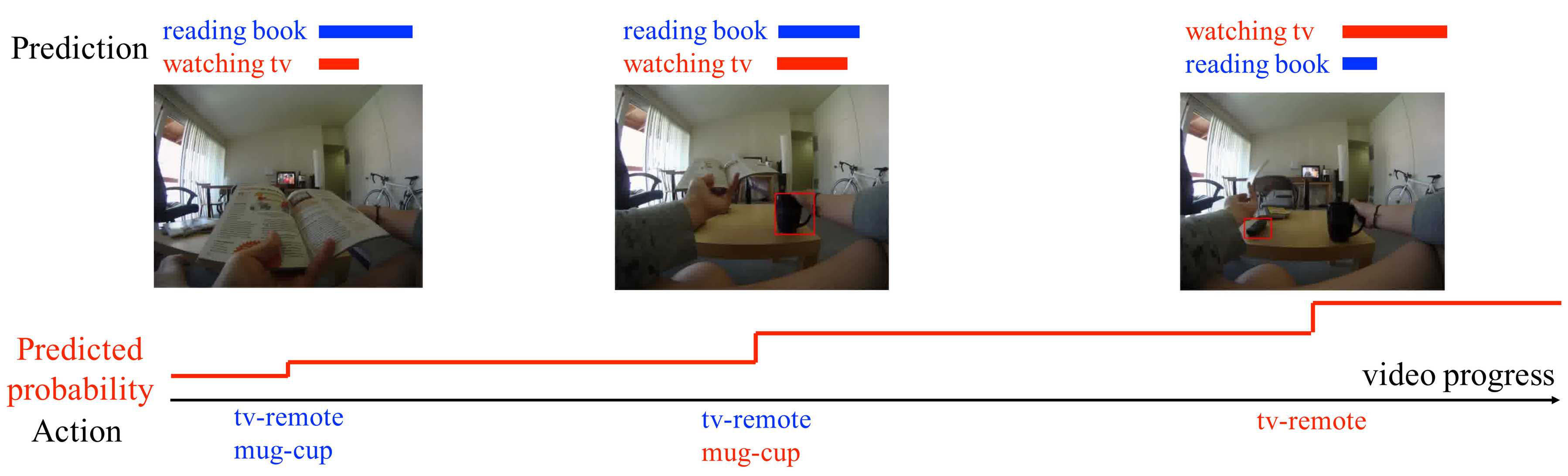
Running buffer model

Buffer proceeds when new frame arrives

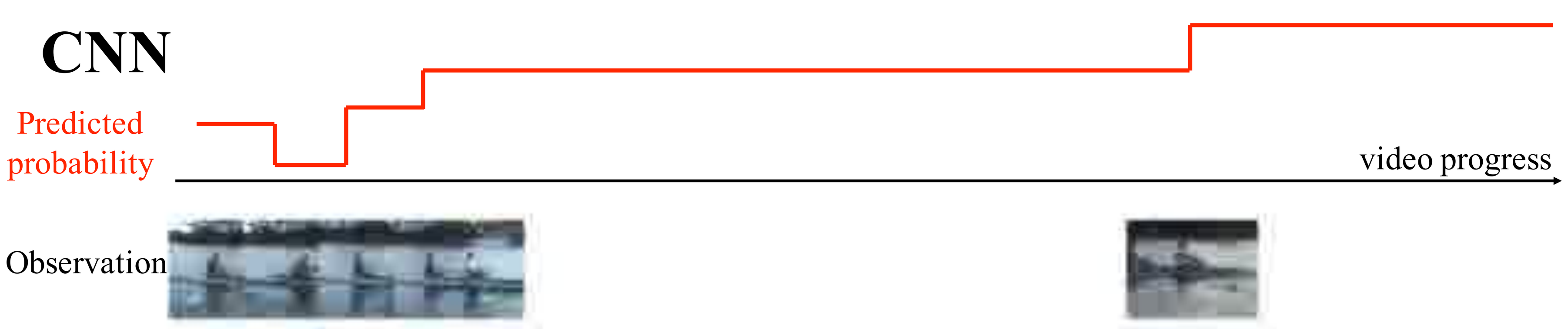


3. Policy in Action

Bag-of-Object

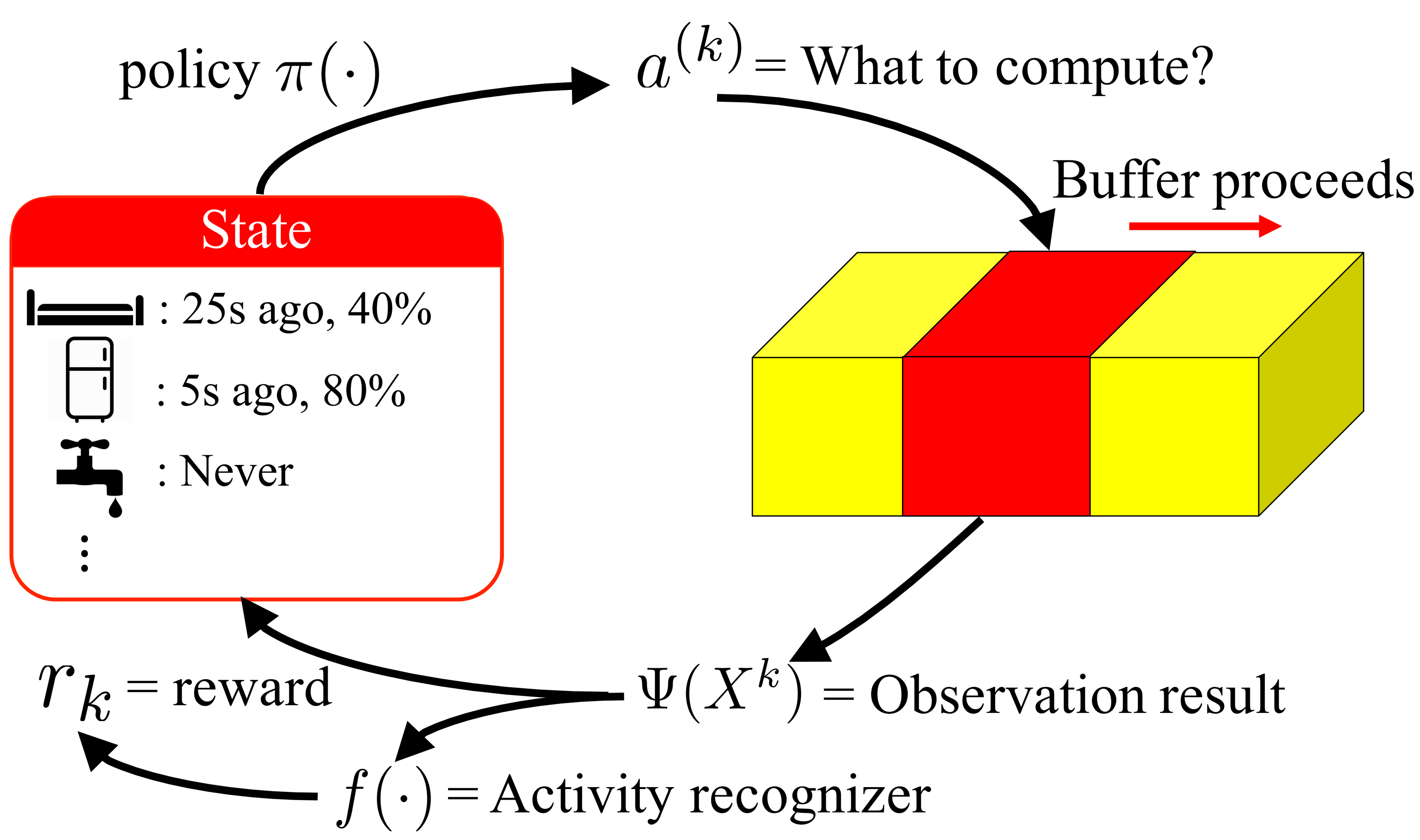


CNN



4. Approach

We formulate the problem as a Markov Decision Process (MDP).



Let X denotes video, $y \in \mathcal{Y}$ denotes activity label and given

- $\Psi(X^k)$ — video descriptor at step-k
- $f : \Psi \times \mathcal{Y} \to \mathbb{R}$ — activity classifier $f(\Psi(X), y) = P(y|X)$

We define the following components for MDP

- Actions $\mathcal{A} = \{a_m\}_{m=0}^{M}$ — {extract m-th feature} $\cup$ {**skip**}
- State-action feature $\phi(s_k, a) = [\Psi(X^k), \delta t^k]$
- Instant reward $r_k = f(\Psi(X^{k+1}), y) - f(\Psi(X^k), y)$

Learn policy by Q-learning with linear function approximation.

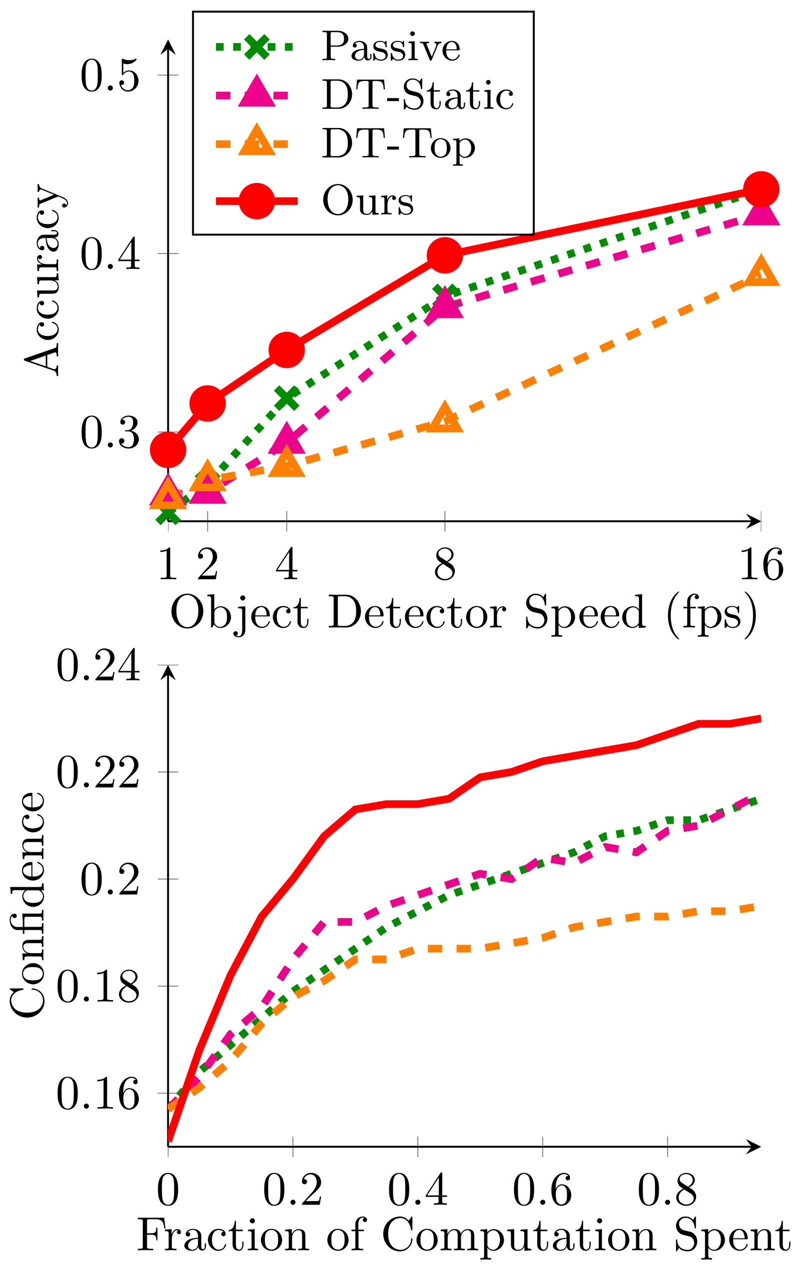
- $\pi(s_k) = \arg\max_a E[R|s_k, a, \pi]$
- $Q^\pi(s, a) = E[R|s, a, \pi] = \sum_k \gamma^k r_k = \theta^T \phi(s, a)$

5. Experiments

Baselines

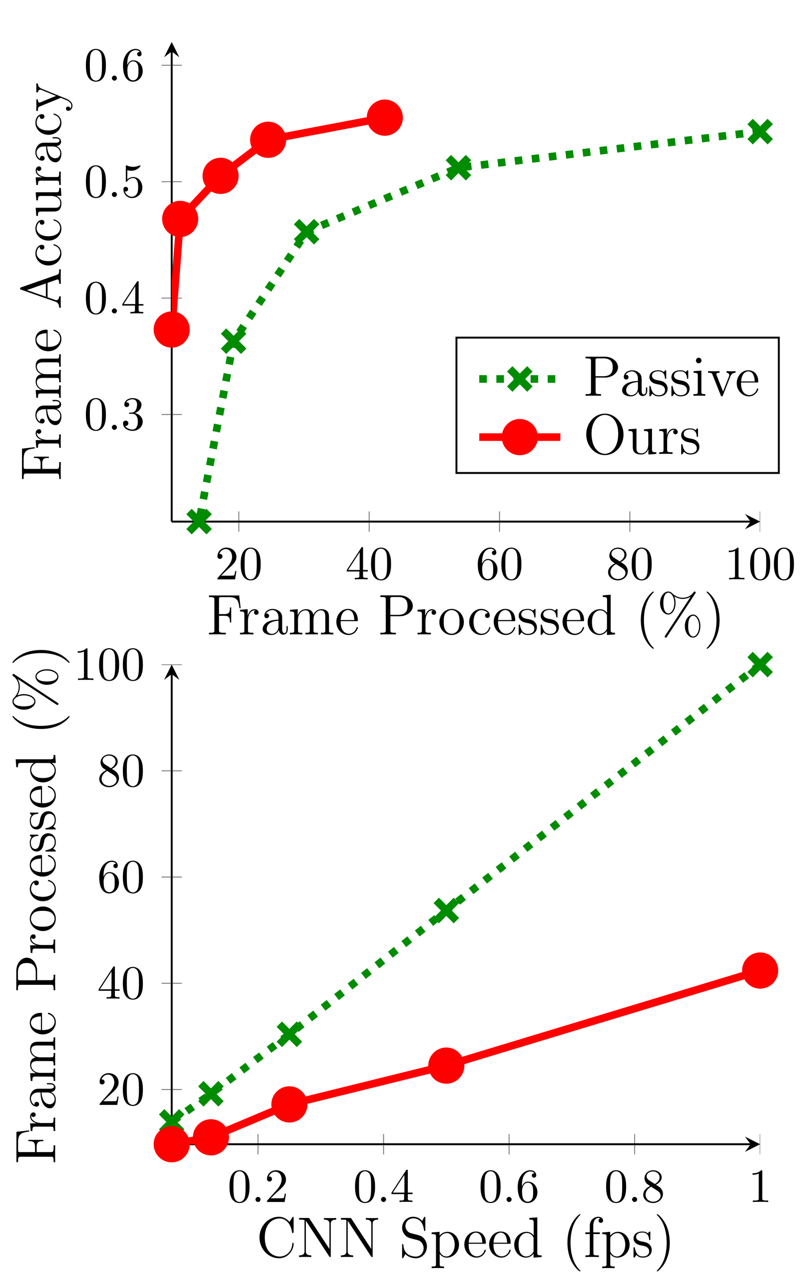
- Passive – no control on what action to take
- DT-Static – fixed order by Decision Tree importance
- DT-Top – only most important feature in Decision Tree

ADL + Bag-of-Object



- Skip computation intelligently
- Requires <40% computation

UCF-101 + CNN



- Performs the best under all budget
- Improve confidence more rapidly