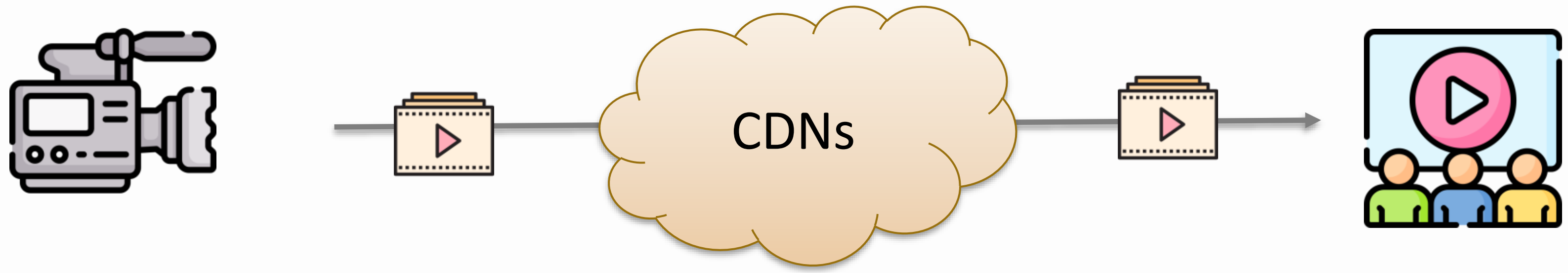
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

CASVA: Configuration-Adaptive Streaming for Live Video Analytics

Miao Zhang, Fangxin Wang, Jiangchuan Liu

BCKGROUND

Traditional video streaming



Human viewers
(entertainment)

Live video analytics

Pervasive camera deployment

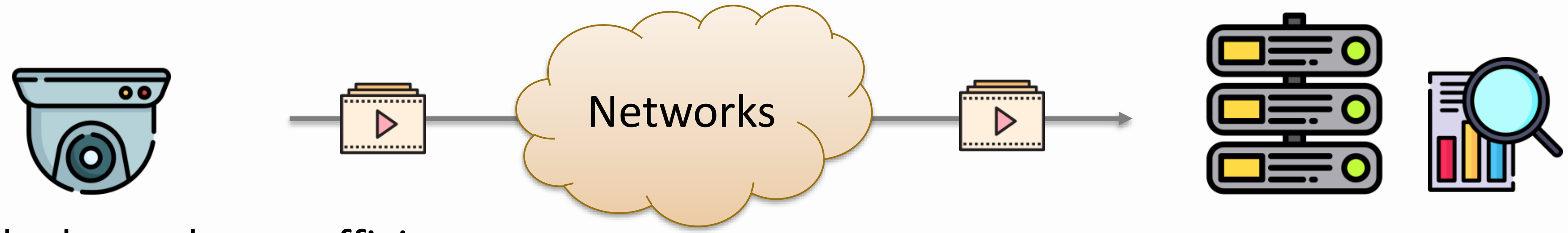
High-accuracy deep neural networks (DNNs)

Live Video Analytics
(automated analysis for real-time actionable insights)

Human beings are no longer the only consumers of videos!

BCKGROUND

Video analytics streaming



Cameras typically do not have sufficient resources for in-situ analytics.

Resource-rich backend server

Goals: Optimizing algorithm-perceived (DNN-perceived) QoE instead of human-perceived QoE.

How to adaptively and efficiently stream videos over the network

for live video analytics?

MOTIVATION

☐ Measurement Setup

Configuration knobs: Frame Rate (FR), Frame Resolution (RS), Quantization Parameter (QP)

Vision Tasks:

Object Detection (OD)
Bounding-box-based task

Semantic Segmentation (SS)
Pixel-based task

MOTIVATION

☐ Measurement Setup

Video dataset:

Video Name	Source	Type	Description
STA1	YouTube Live	stationary traffic camera	A video clip collected on a sunny day
STA2	YouTube Live	stationary traffic camera	A video clip collected on a rainy morning
STA3	YouTube Live	stationary traffic camera	A video clip collected on a sunny morning
DASH1	YouTube	Dashcam	Daytime drive in Chicago downtown
DASH2	YouTube	Dashcam	Night drive around London downtown

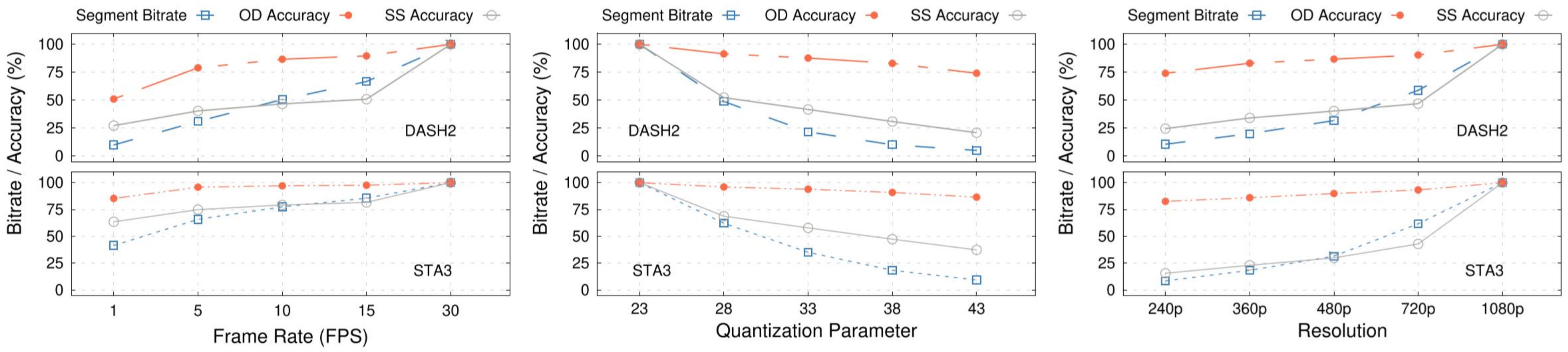
Metrics of Interest:

Bitrate: indicate the network resource requirement.

Accuracy: F1 for OD and mIoU for SS.

MOTIVATION

❑ Measurement Insights

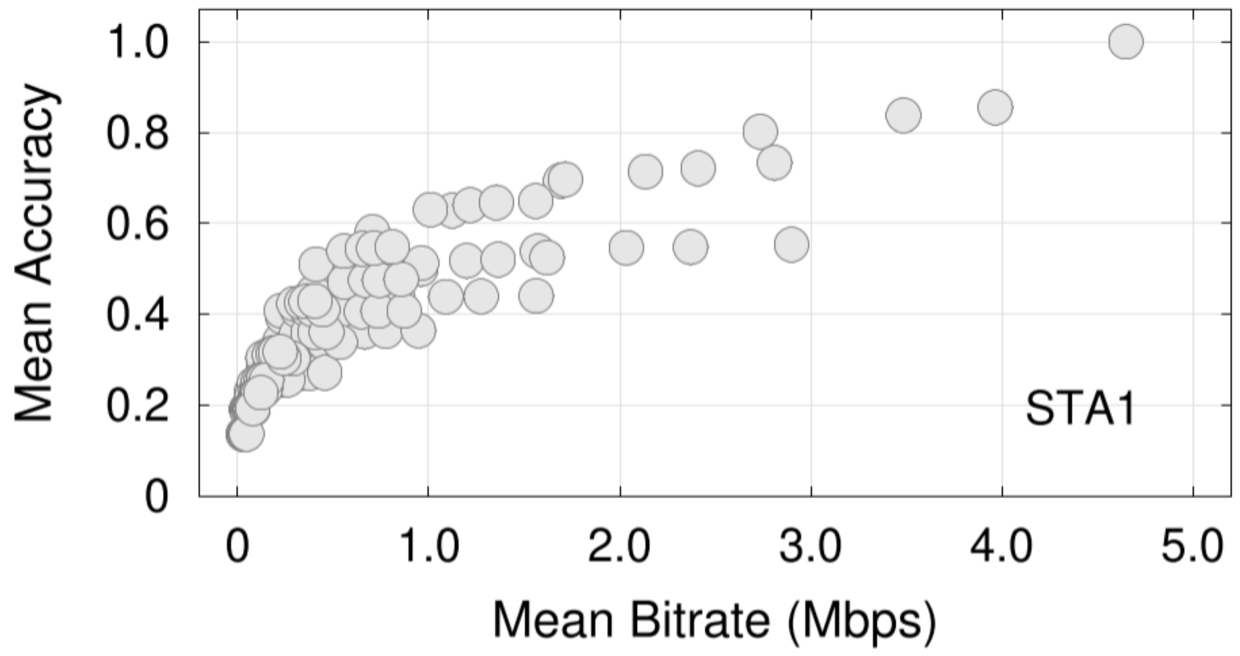


(a) Tune FR (RS: 1080p, QP: 23). (b) Tune QP (RS: 1080p, FR: 30). (c) Tune RS (FR: 30, QP: 23).

Different configuration knobs have different impacts on bitrate and accuracy, and such impacts are video-specific and task-specific.

MOTIVATION

Measurement Insights



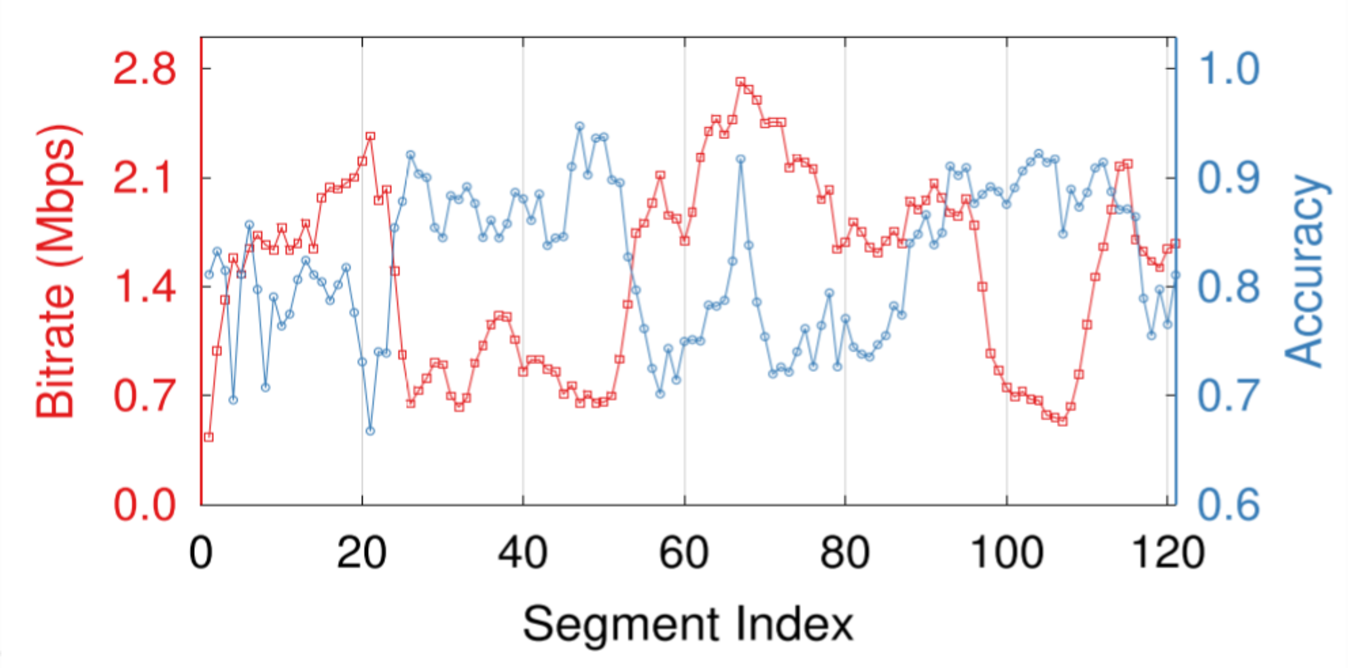
Mean bitrate and accuracy distribution of all configurations (Task: SS, video: STA1).

A higher bitrate does not necessarily lead to a higher accuracy, and configurations with similar bitrates can have very different accuracies.

Configuration tuning is necessary for bandwidth-efficient video analytics.

MOTIVATION

Measurement Insights



Segment bitrate and accuracy variations under a specific configuration (FR:10, QP: 28, RS: 720p; Task: OD, video: DASH1).

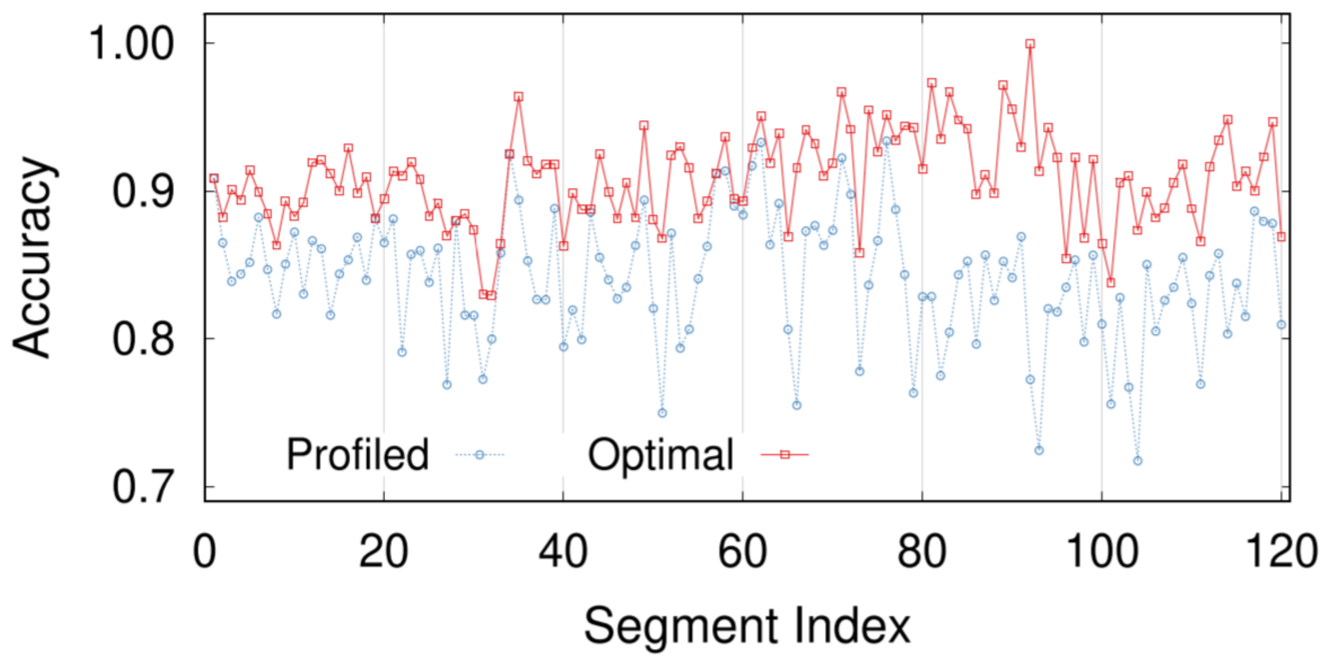
The relationship between configuration and bitrate (accuracy) is video content-dependent and highly variable.

⇩

Configuration-based streaming needs to be content-adaptive.

MOTIVATION

☐ Measurement Insights



Segment accuracy comparison of the profiled and optimal configuration (Task: OD, video: STA2, available bandwidth: 1.5 Mbps).

Profiling-based solutions fail to keep up with the intrinsic dynamics of bandwidth-accuracy trade-off.

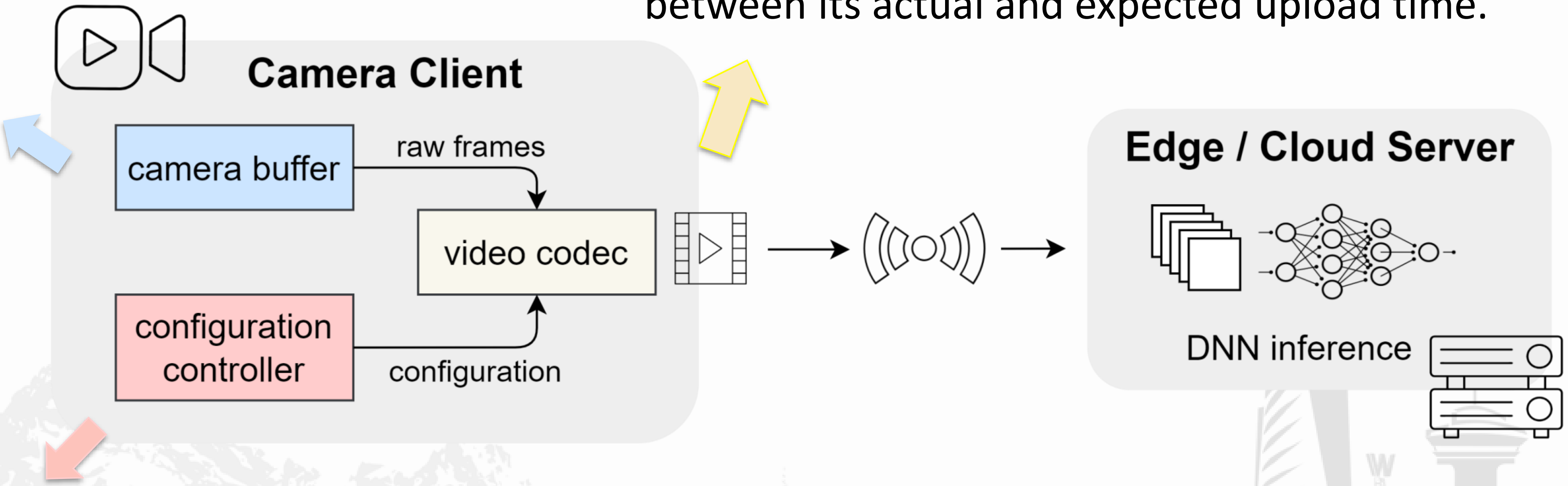
Continually fine-grained configuration adaptation is necessary.

CASVA

Configuration-Adaptive Streaming: Framework

Frames are encoded and delivered in segments. The upload lag of a segment is the time difference between its actual and expected upload time.

Frames captured by the camera are cached in a buffer.



Choosing the configuration for each segment to minimize upload lags while maximizing the server-side inference accuracy.

CASVA

Configuration-Adaptive Streaming: Challenges

- High accuracy and low latency are inherently conflicting goals.
- The server-side inference accuracy is affected by video content dynamics.
- The upload delays are influenced by dynamic segment bitrate and network conditions.
- In continuous live streaming scenarios, the upload lags can be accumulated.

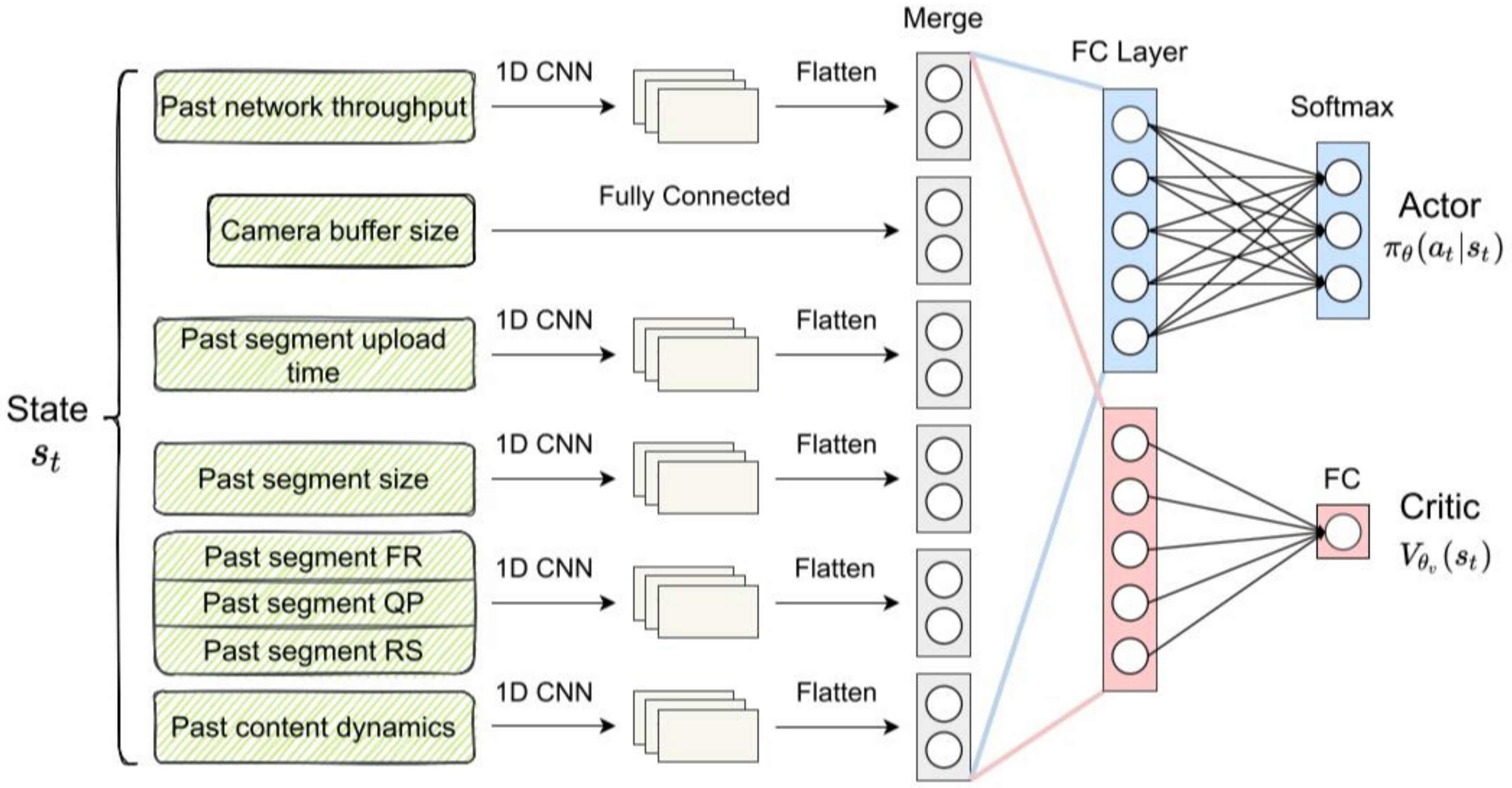
CASVA

Deep Reinforcement Learning Based Solution

State: past network conditions, buffer status, past configuration choices and video content characteristics.

Optimization goal: maximizing the long-term cumulative DNN-perceived QoE.

Policy gradient training: a dual-clipped Proximal Policy Optimization (PPO) method.



EVALUATION

Two streaming modes

Latency-first: $r_t = \alpha_1 Q_t - \alpha_2 \max(u_t - l, 0) / l - \alpha_3 M_t$

delivery-first: $r_t = \alpha_1 Q_t - \alpha_2 \max(u_t - l, 0) / l + \alpha_3 \mathbb{I}(b_{t+1} < b_t)(b_{t+1} - b_t)/l$

Network traces

An FCC fixed broadband dataset, a 4G/LTE bandwidth dataset

Evaluation metrics

Mean accuracy, mean lag, segment loss rate (latency-first mode only).

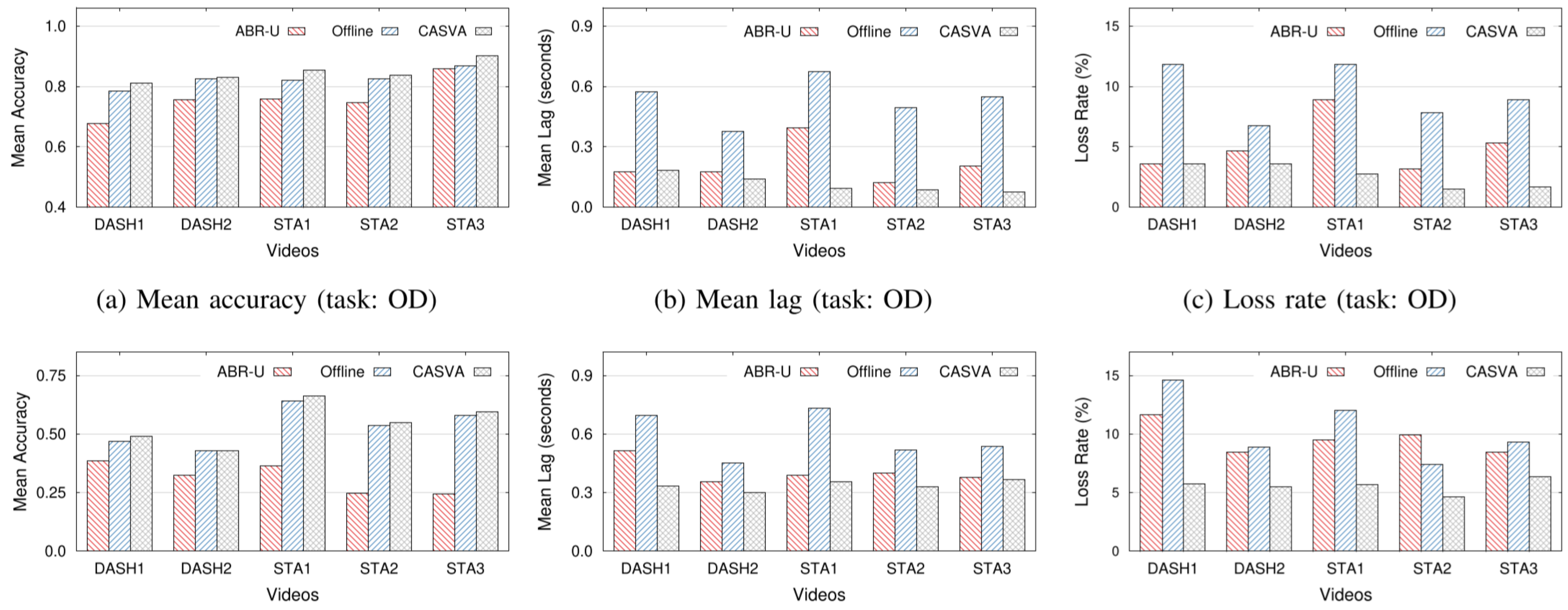
Baselines

ABR-U: a DRL-based ABR solution

Offline: a profiling-based solution

EVALUATION

Evaluation Results



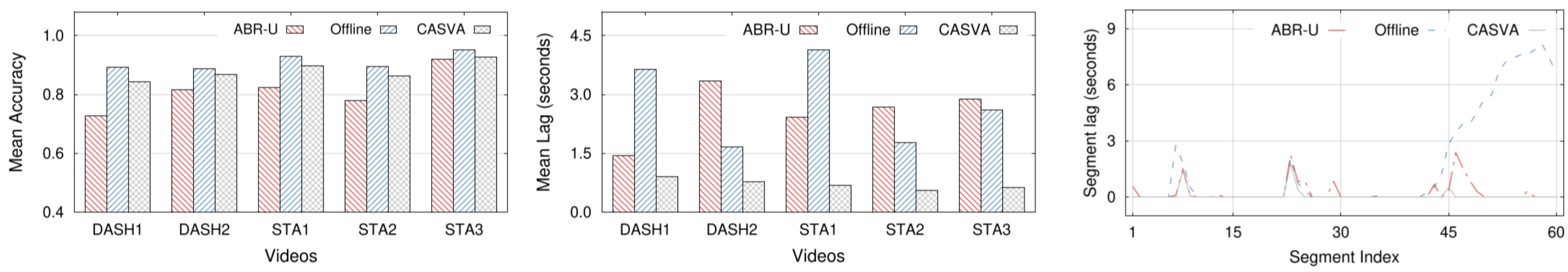
(a) Mean accuracy (task: OD) (b) Mean lag (task: OD) (c) Loss rate (task: OD)

(d) Mean accuracy (task: SS) (e) Mean lag (task: SS) (f) Loss rate (task: SS)

Performance of different methods in the latency-first mode (network traces: 4G/LTE)

EVALUATION

Evaluation Results



(a) Mean accuracy comparison (b) Mean lag comparison (c) Lag variations over time

Performance of different methods in the delivery-first mode (Task: OD; network traces: 4G/LTE)

SUMMARY

- Live video analytics creates new opportunities for video streaming, and it requires new designs of the streaming frameworks.
- Tuning video encoding configurations allows fine-grained adaptation to dynamic video content and network conditions.
- Deep reinforcement learning is well suited for addressing the challenges in configuration-adaptive streaming.

THANK YOU

Q & A