ARTIFACT EVALUATED usenix ASSOCIATION AVAILABLE

ARTIFACT EVALUATED usenix ASSOCIATION FUNCTIONAL

ARTIFACT EVALUATED usenix ASSOCIATION REPRODUCED

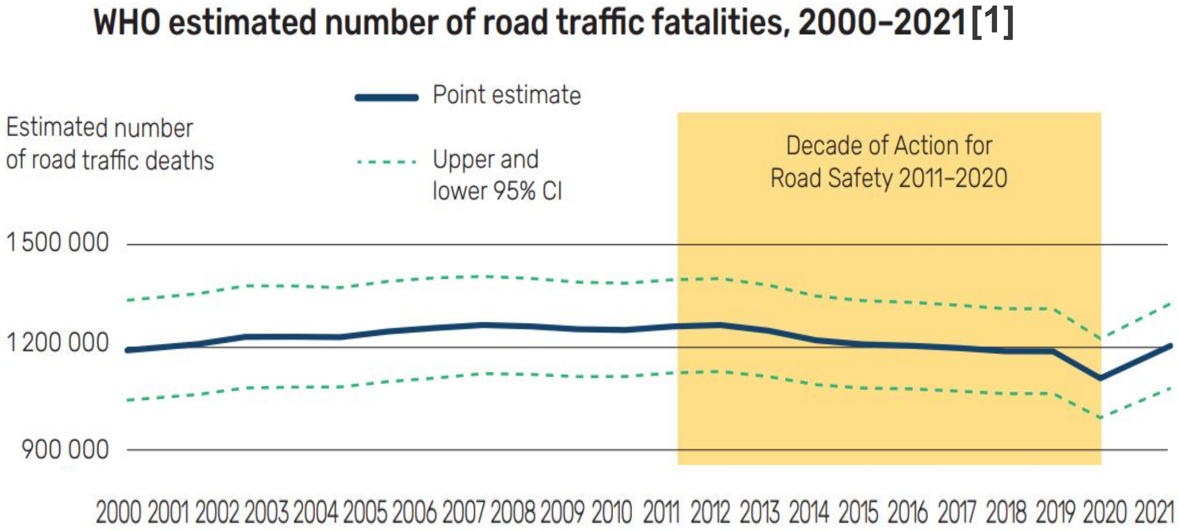
TileClipper: Lightweight Selection of Regions of Interest from Videos for Traffic Surveillance

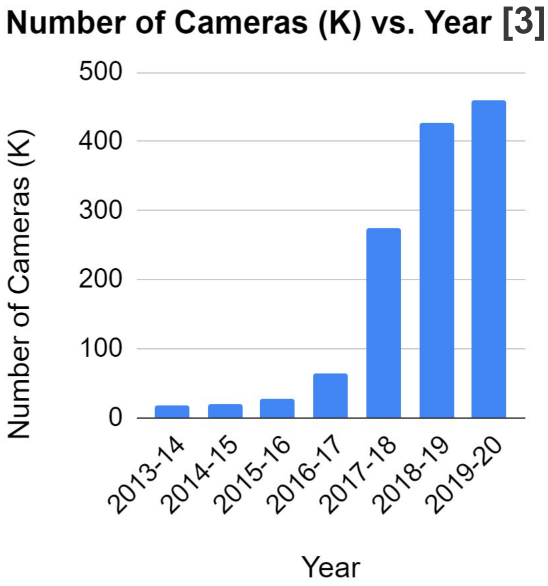
Shubham Chaudhary, Aryan Taneja, Anjali Singh, Purbasha Roy, Sohum Sikdar, Mukulika Maity, Arani Bhattacharya

Indraprastha Institute of Information Technology Delhi (IIITD), India

USENIX Annual Technical Conference 2024

Need of Automated Traffic Surveillance

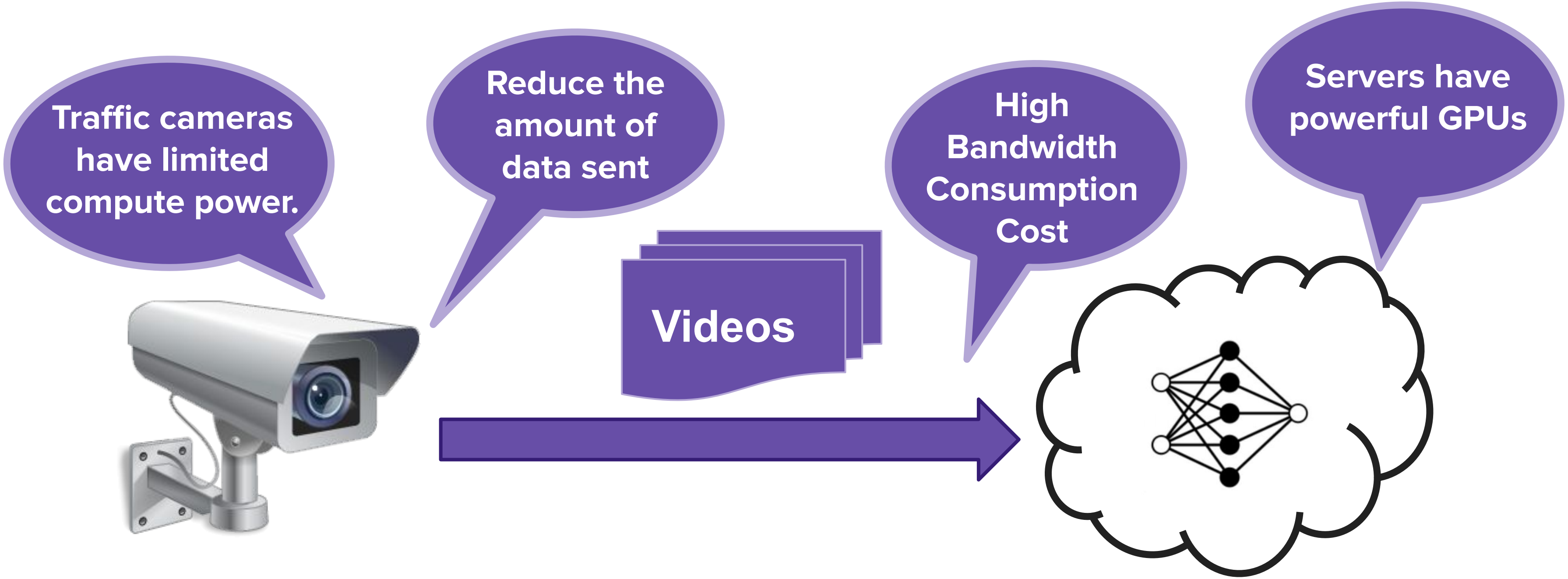

WHO estimated number of road traffic fatalities, 2000–2021[1]


Number of Cameras (K) vs. Year [3]

Cities are installing thousands of cameras for traffic monitoring [2].

[1] WHO Global status report on road safety 2023
[2] https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/
[3] India BPRD 2020 Report

Challenges of Traffic Surveillance



Traffic cameras have limited compute power.

Reduce the amount of data sent

Videos

High Bandwidth Consumption Cost

Servers have powerful GPUs

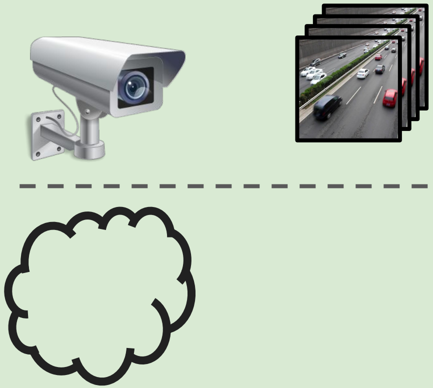
Sending data over a network has high bandwidth and latency cost

Image Sources:

1. https://samriddhi2958.medium.com/neural-networks-and-their-applications-44bc7062dd94
2. https://i.gifer.com/7VB.gif

Existing Pruning Techniques

Frame Pruning [1]

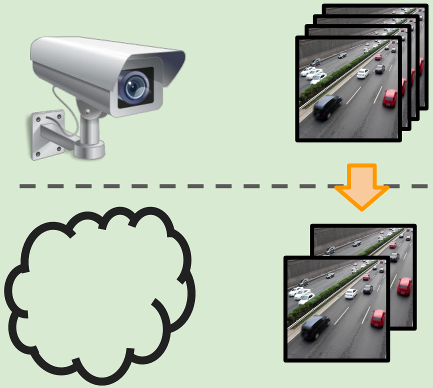


Needs access to raw frames

[1] Yuanqi Li et al. "Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics" SIGCOMM’20

Existing Pruning Techniques

Frame Pruning [1]

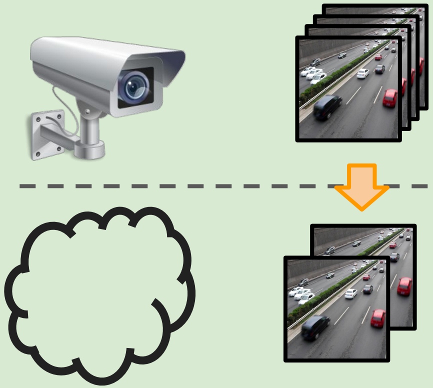


Needs access to raw frames

[1] Yuanqi Li et al. "Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics" SIGCOMM’20

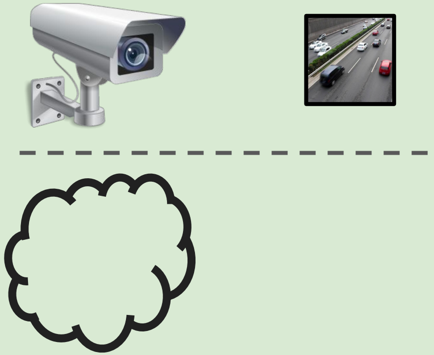
Existing Pruning Techniques

Frame Pruning [1]



Needs access to raw frames

Quality Pruning [2]

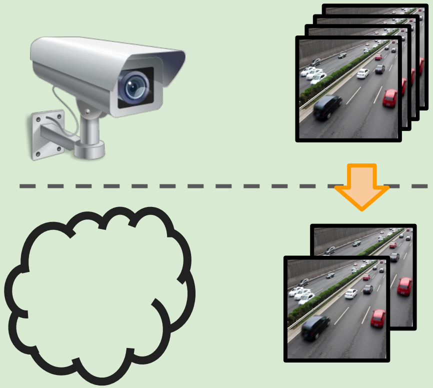


High server GPU usage cost

[1] Yuanqi Li et al. "Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics" SIGCOMM’20
[2] Yiding Wang et al. "Bridging the Edge-Cloud Barrier for Real-time Advanced Vision Analytics." HotCloud’19

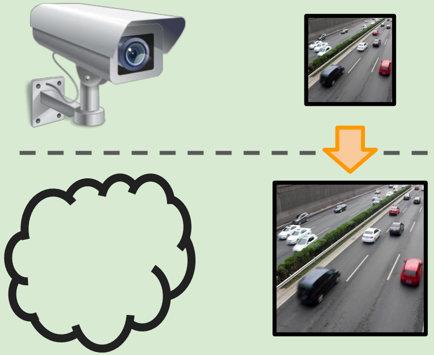
Existing Pruning Techniques

Frame Pruning [1]



Needs access to raw frames

Quality Pruning [2]

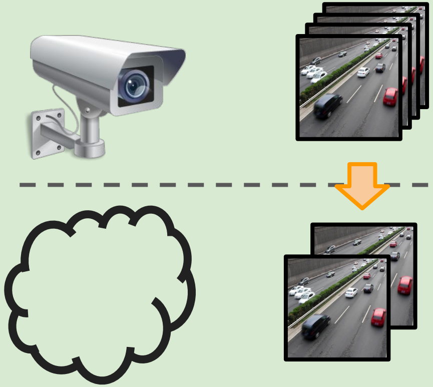


High server GPU usage cost

[1] Yuanqi Li et al. "Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics" SIGCOMM'20
[2] Yiding Wang et al. "Bridging the Edge-Cloud Barrier for Real-time Advanced Vision Analytics." HotCloud'19

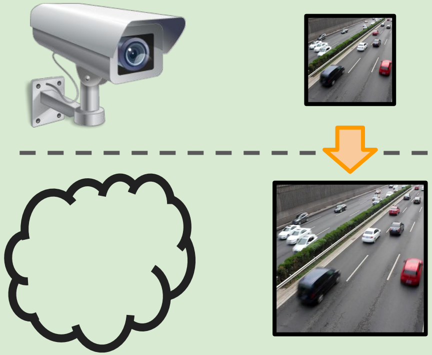
Existing Pruning Techniques

Frame Pruning [1]



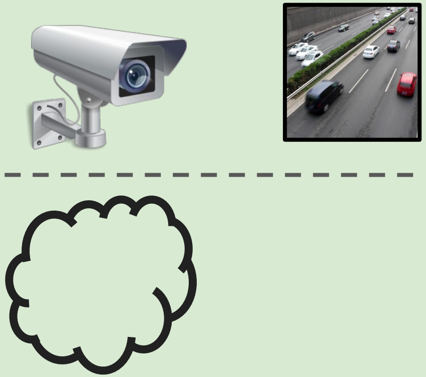
Needs access to raw frames

Quality Pruning [2]



High server GPU usage cost

Spatial Pruning [3]

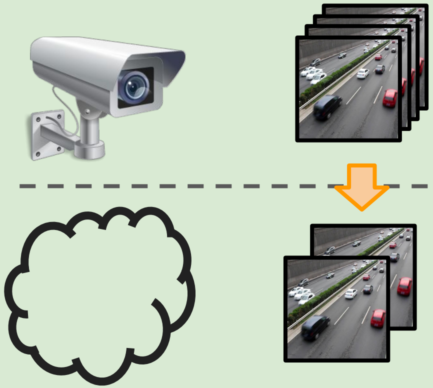


High camera-side overhead

[1] Yuanqi Li et al. "Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics" SIGCOMM’20
[2] Yiding Wang et al. "Bridging the Edge-Cloud Barrier for Real-time Advanced Vision Analytics." HotCloud’19
[3] Shengzhong Liu et al., “AdaMask: Enabling Machine-Centric Video Streaming with Adaptive Frame Masking for DNN Inference Offloading,” MM’22

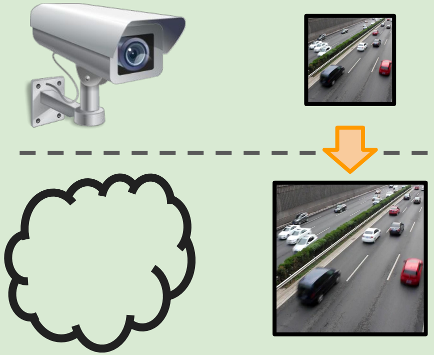
Existing Pruning Techniques

Frame Pruning [1]



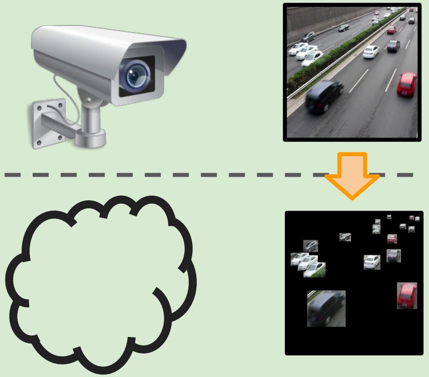
Needs access to raw frames

Quality Pruning [2]



High server GPU usage cost

Spatial Pruning [3]



High camera-side overhead

How to reduce the amount of redundant data sent to server without any additional compute overheads?

[1] Yuanqi Li et al. "Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics" SIGCOMM’20

[2] Yiding Wang et al. "Bridging the Edge-Cloud Barrier for Real-time Advanced Vision Analytics." HotCloud’19

[3] Shengzhong Liu et al., “AdaMask: Enabling Machine-Centric Video Streaming with Adaptive Frame Masking for DNN Inference Offloading,” MM’22

Outline

1. Background and Problem Statement

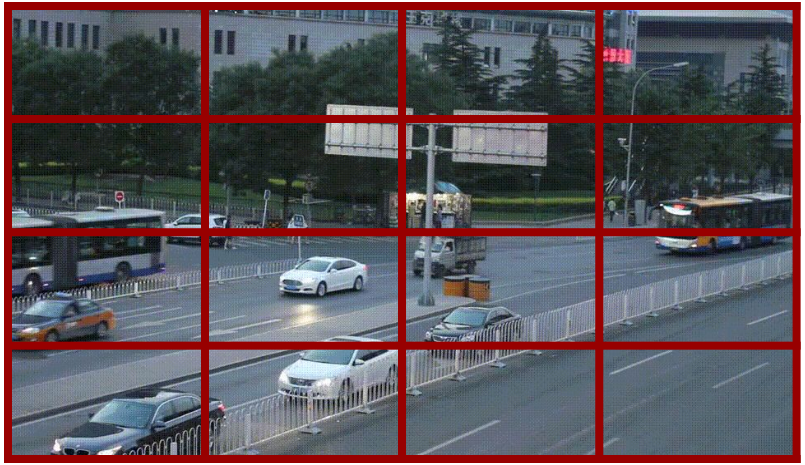
2. TileClipper: Approach and Design

3. Evaluation

4. Conclusion

Our Strategy: Leverage Tiles in Video Encoding

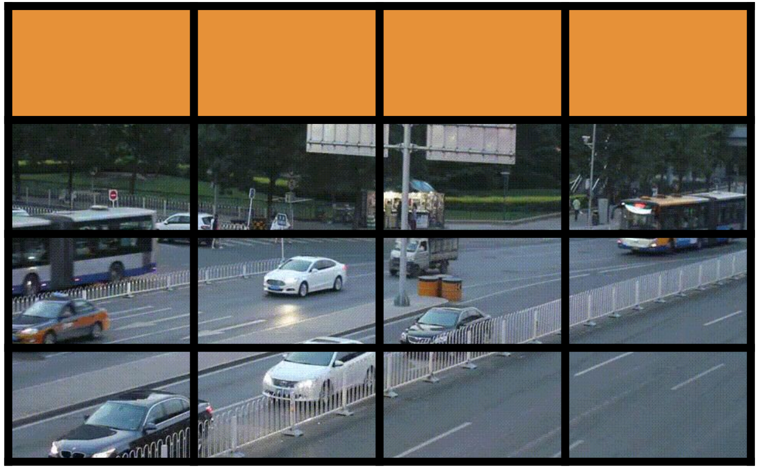
They act as independent video streams

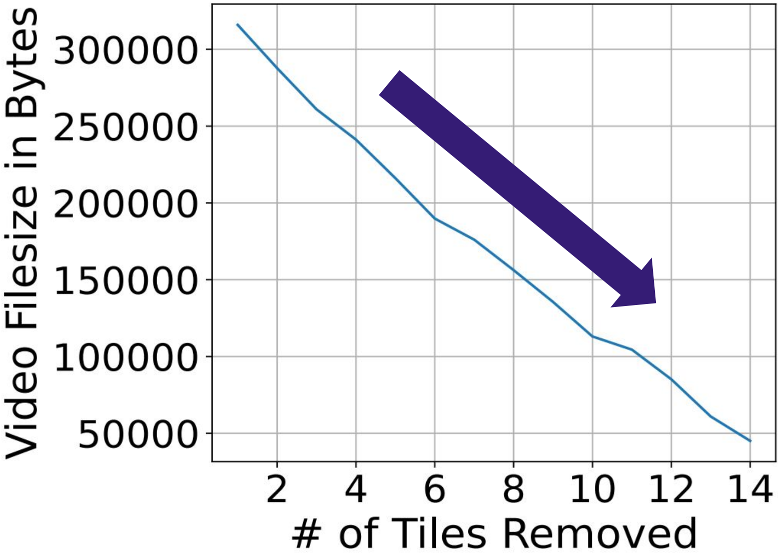


Tiles are spatial rectangular blocks

Tile manipulation in HEVC/H.265 codec does not require re-encoding

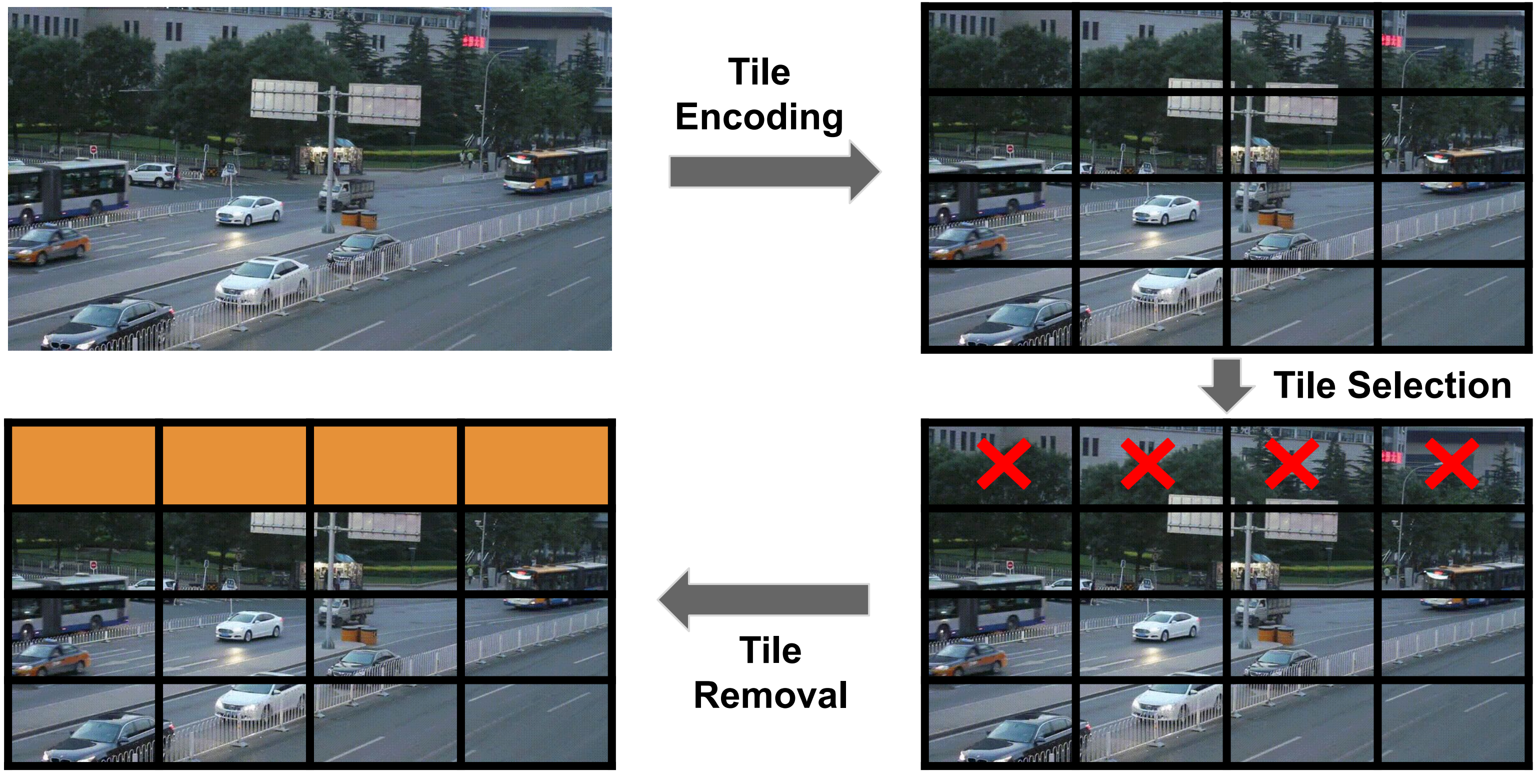
Can Tile Removal Reduce Bandwidth Consumption?





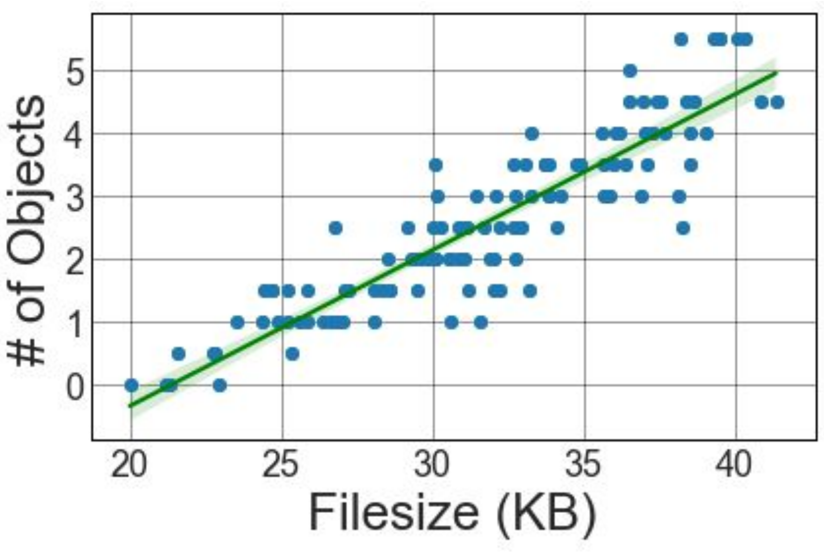
Removing tiles reduces filesize of a video

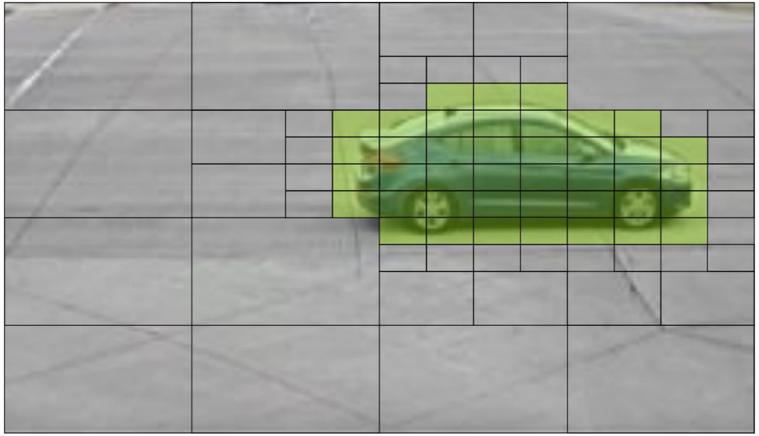
Using Tiles to Filter Irrelevant Spatial Portions



How to select tiles with objects at camera side without a neural network?

Correlation Between Tile Bitrate & Number of Objects



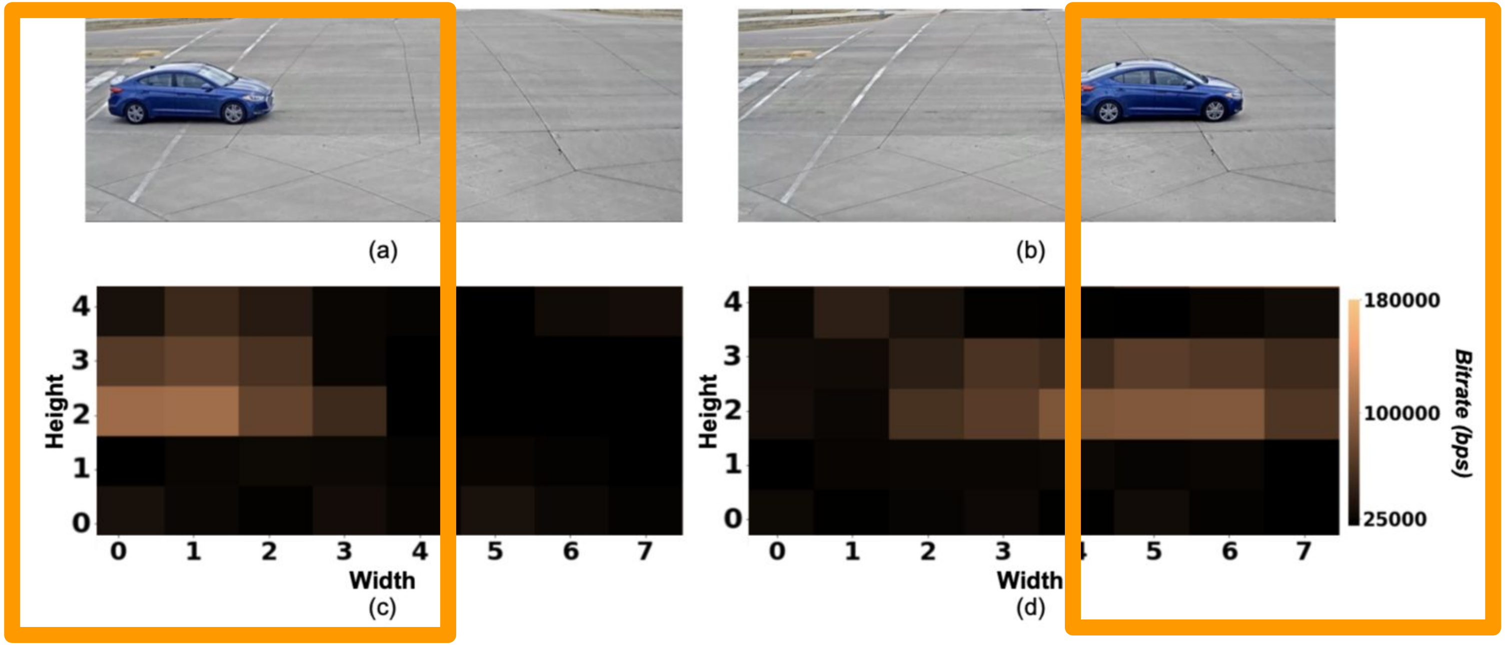


Complex scenes require more bits to encode

Spearman Correlation between 0.75 to 0.90

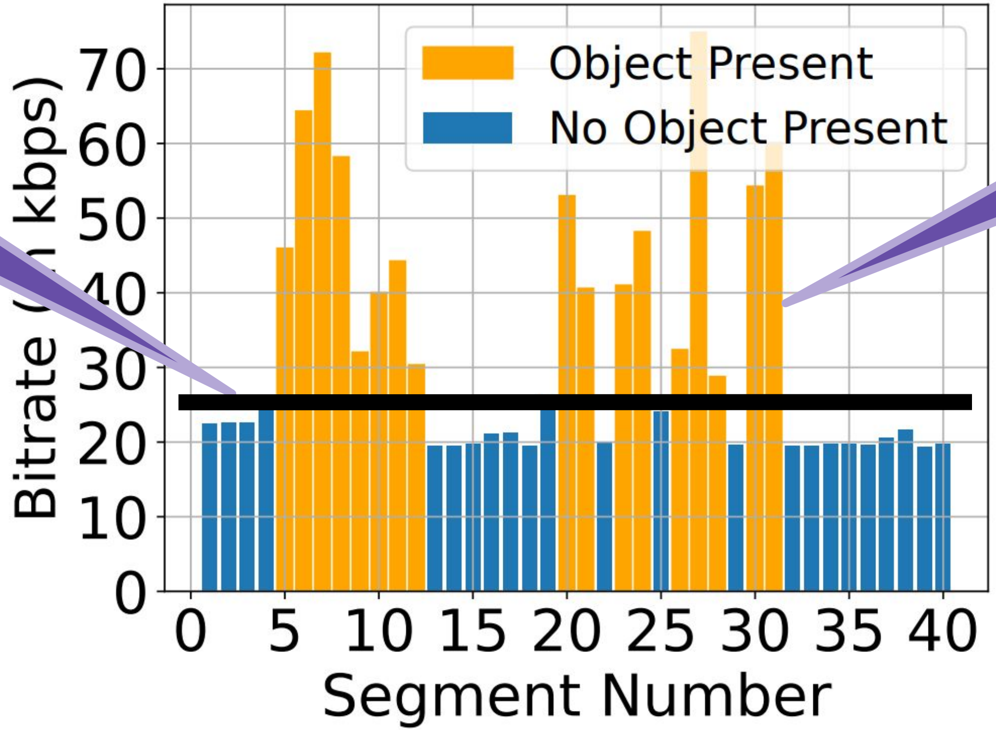
* Bitrate: Number of bits required to encode one second of a video

Can We Utilize Bitrate to Filter Irrelevant Tiles?



Higher bitrate is a signature of objects' presence.

The Bitrates are Noisy in Nature

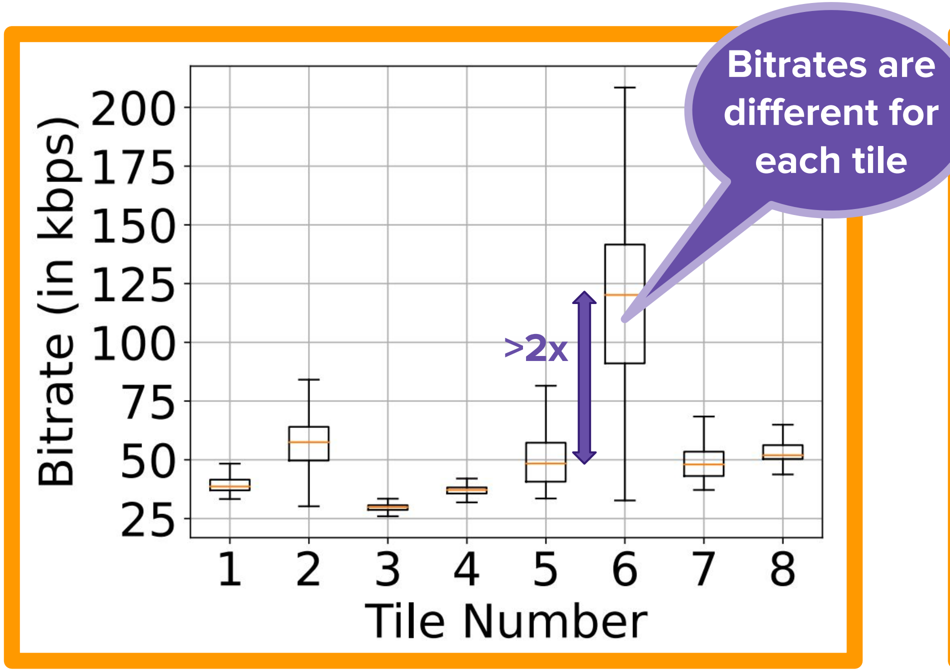


How to know detection threshold?

Spikes shows objects presence

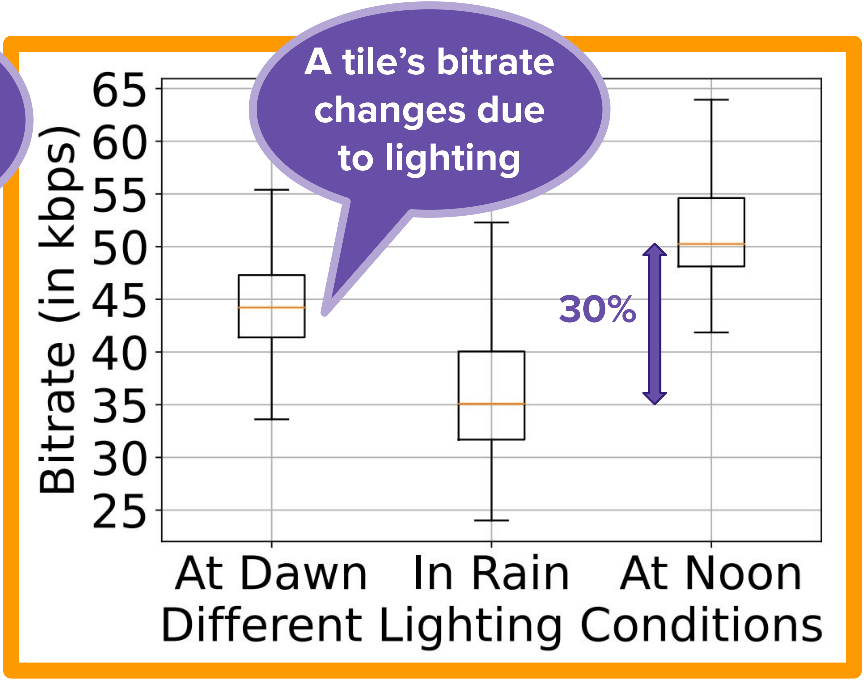
* **Segment: A bunch of encoded frames. A segment has 15 frames in our case.**

Lighting Conditions Affect The Threshold



Bitrates are different for each tile

Bitrate of tiles of the same video

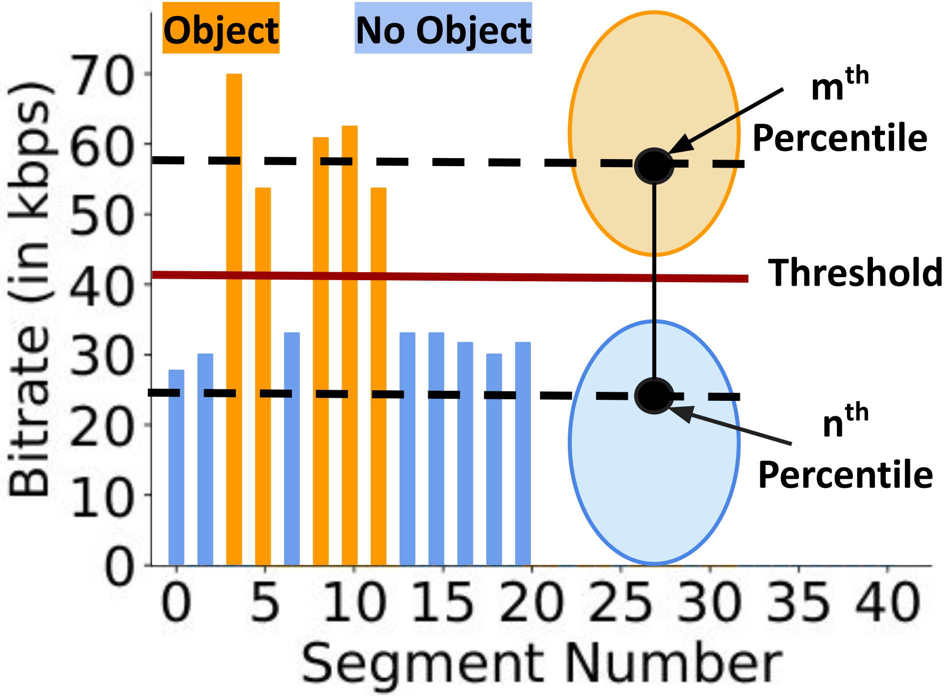


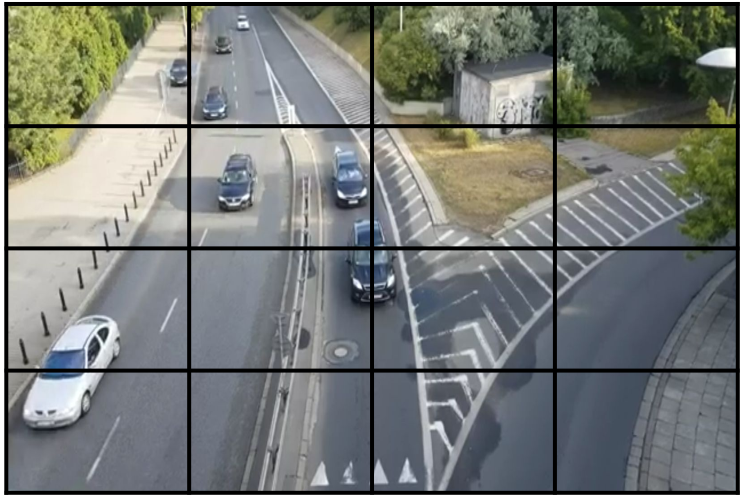
A tile's bitrate changes due to lighting

Same tile in different lighting condition

The threshold should be adaptive and different for each tile

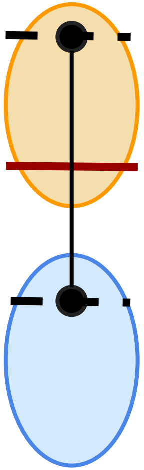
Clustering-Based Tile Selection Algorithm



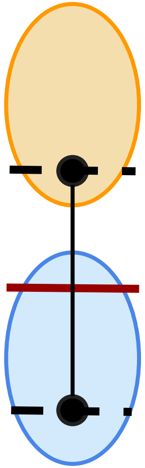


We run the algorithm for each tile independently because they have distinct bitrate distribution

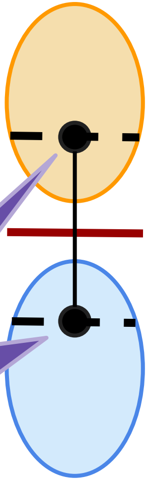
Imperative to Choose Right Percentile Values



High False Negatives

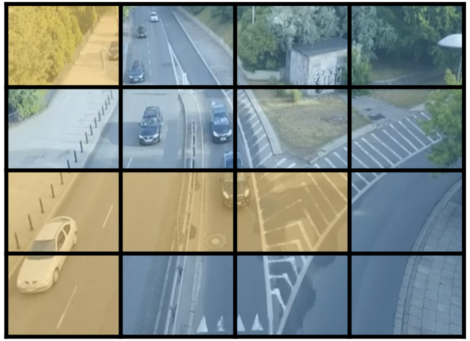


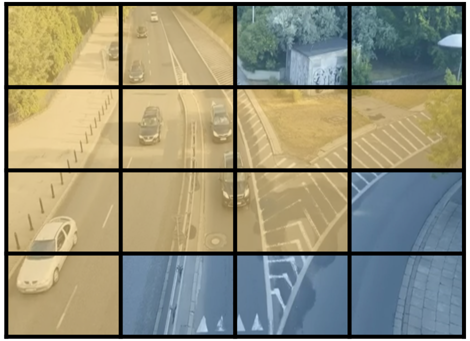
High False Positives

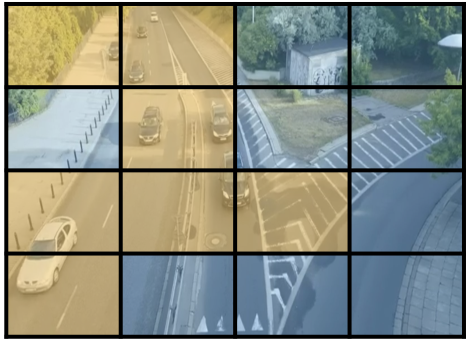


What should be the value of m and n?

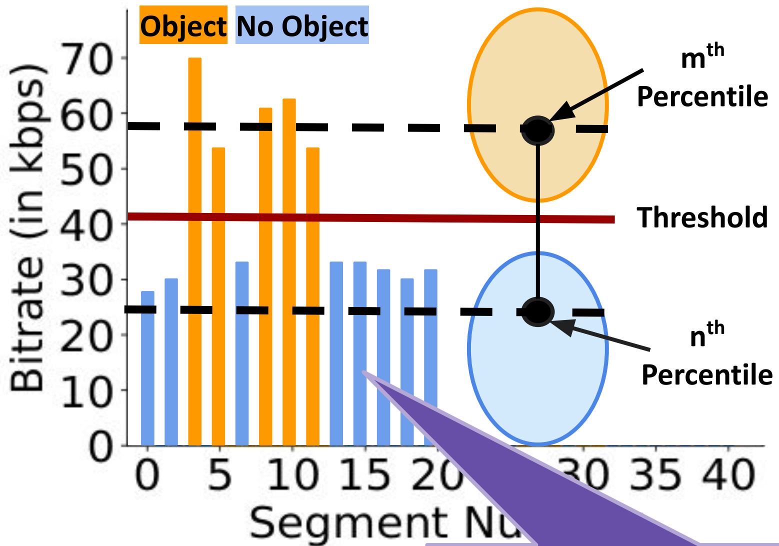
Balanced FP & FN





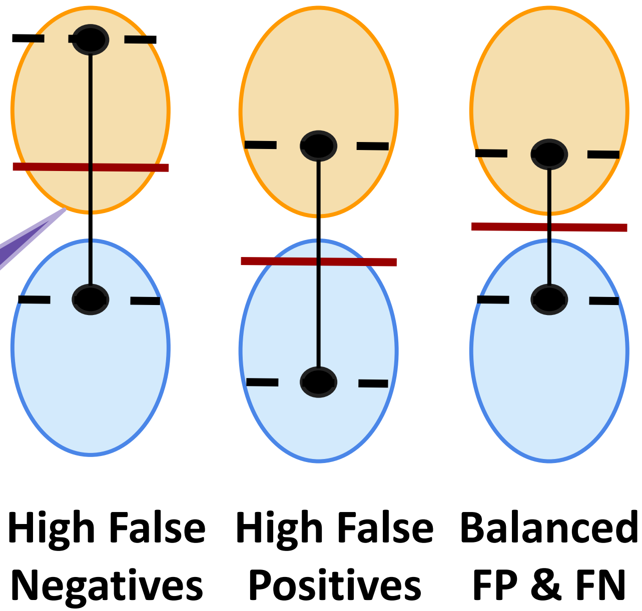


Need Calibration to Find Best Percentile



Do not have access to ground truth at runtime to estimate FPs and FNs.

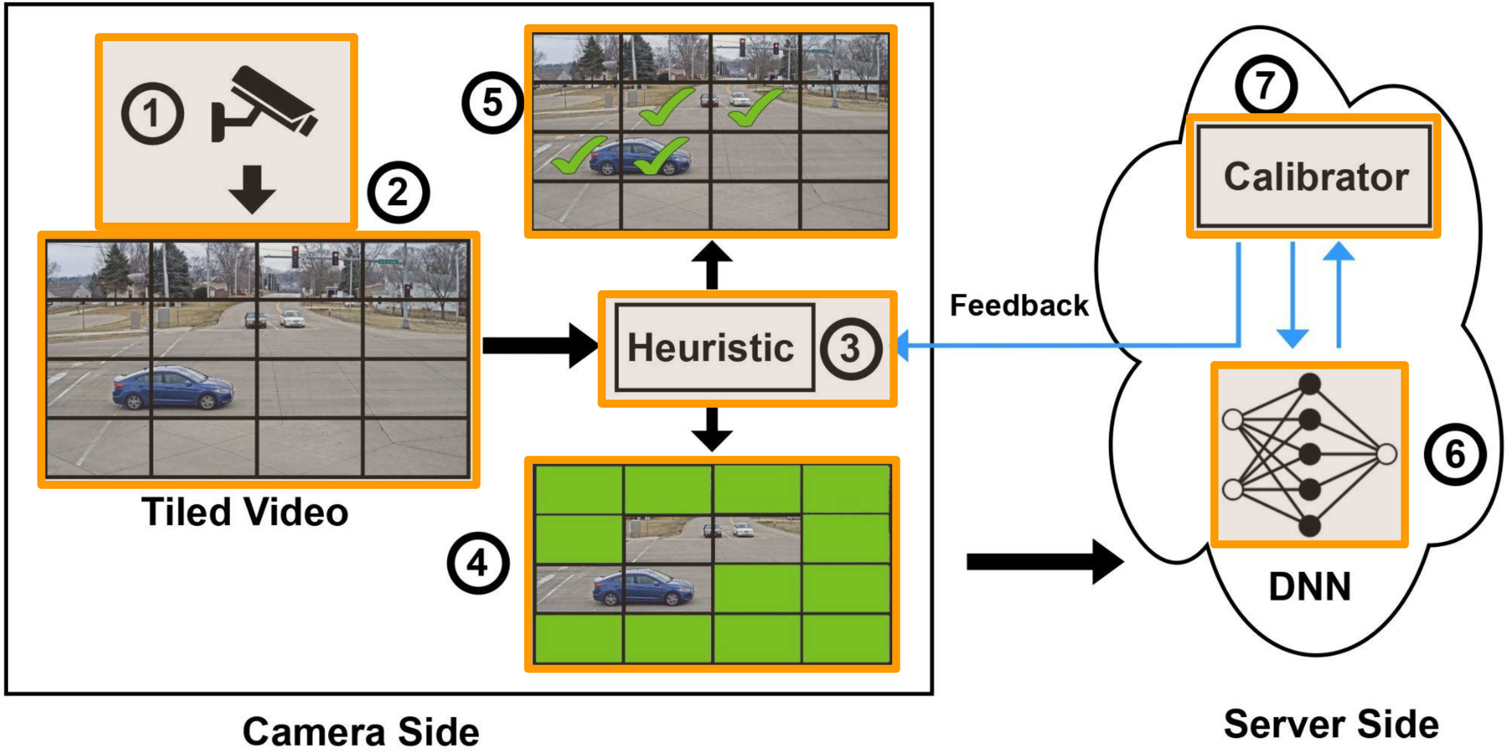
Misses (FNs) are more expensive [1]



We exhaustively search the best percentile <m, n> that maximizes F2 score

[1] Andrea Ceccarelli et al. Evaluating object (mis)detection from a safety and reliability perspective: Discussion and measures. IEEE Access

TileClipper's Architecture



Outline

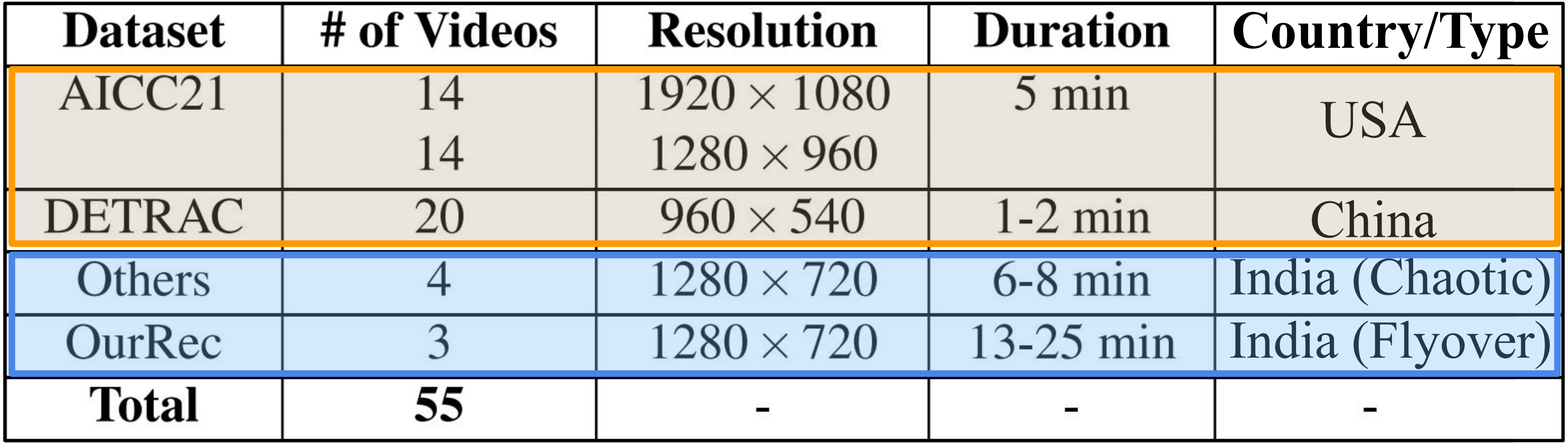
1. Background and Problem Statement

2. TileClipper: Approach and Design

3. **Evaluation**

4. Conclusion

Datasets Used For Evaluation



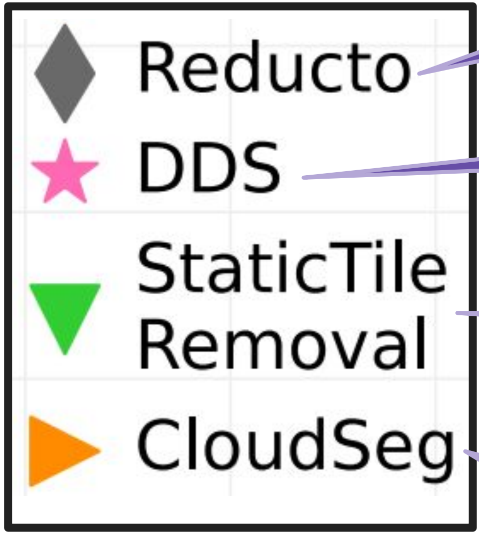
Dataset	# of Videos	Resolution	Duration	Country/Type
AICC21	14	1920 × 1080	5 min	USA
	14	1280 × 960		
DETRAC	20	960 × 540	1-2 min	China
Others	4	1280 × 720	6-8 min	India (Chaotic)
OurRec	3	1280 × 720	13-25 min	India (Flyover)
Total	**55**	-	-	-

We encode all videos into 4x4 tiles using Kvazaar encoder

Report object detection accuracy utilizing Yolov5 to get the ground truths

- **AICC: AI City Challenge 2021 Dataset**
- **OurRec: Our Recorded Videos**

Baselines



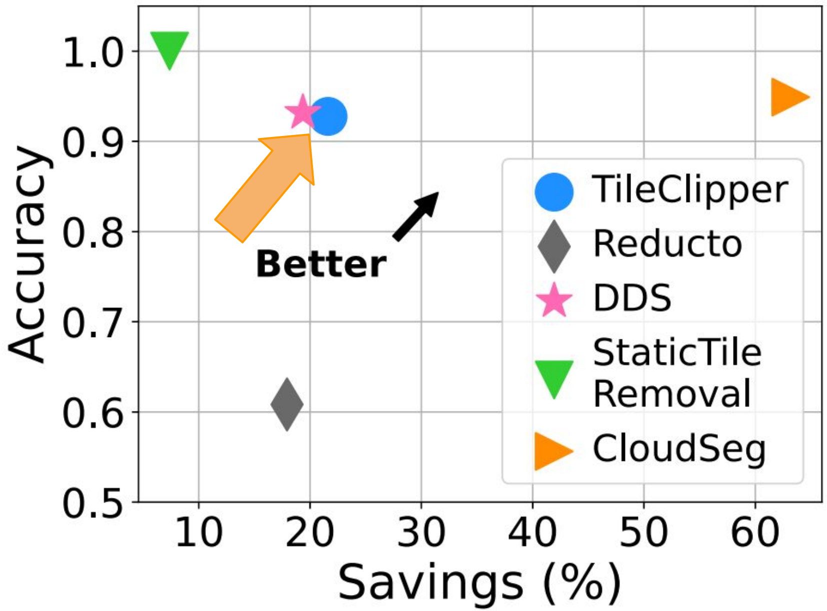
Uses frame filtering [SIGCOMM ‘20]

Sends high quality video if server fails to detect in low quality [SIGCOMM ‘20]

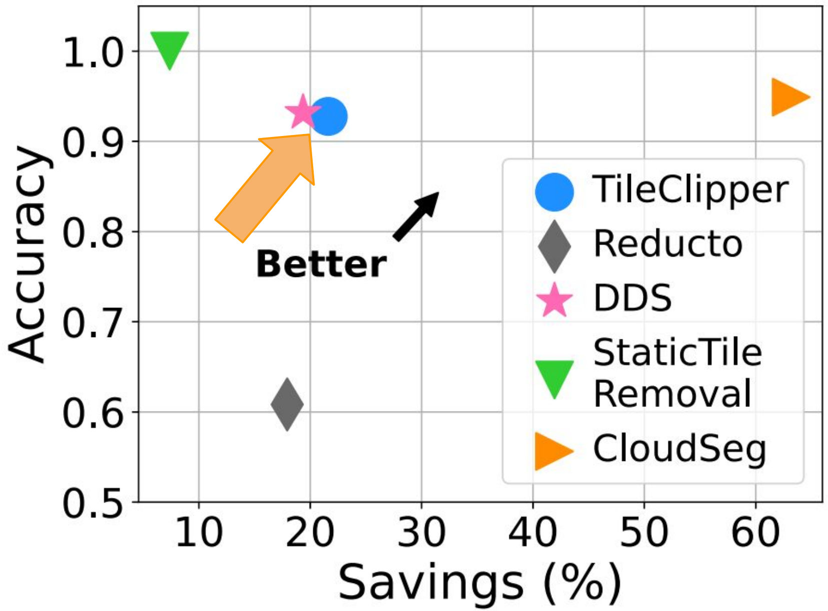
Filters only off side road tiles

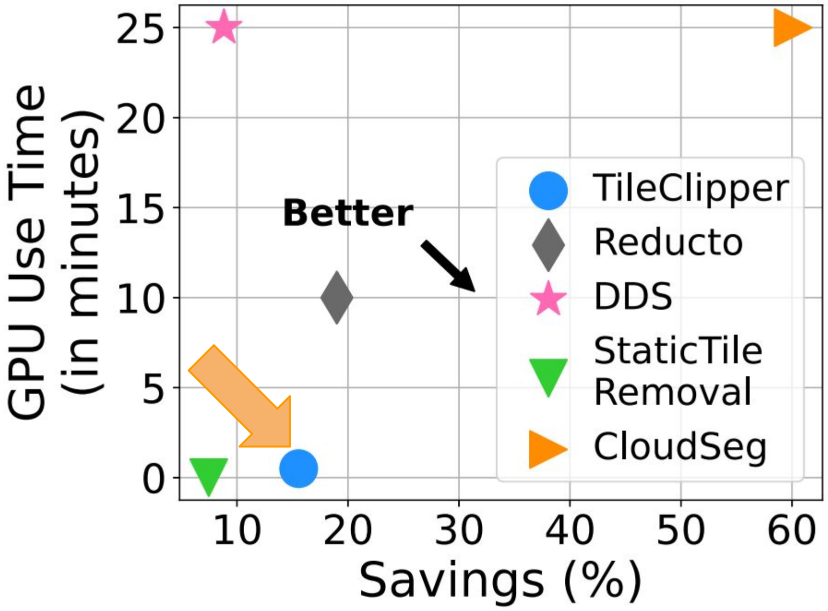
Uses super resolution for upsampling video at server [HotCloud ‘19]

Our Trade-offs: Accuracy vs Savings



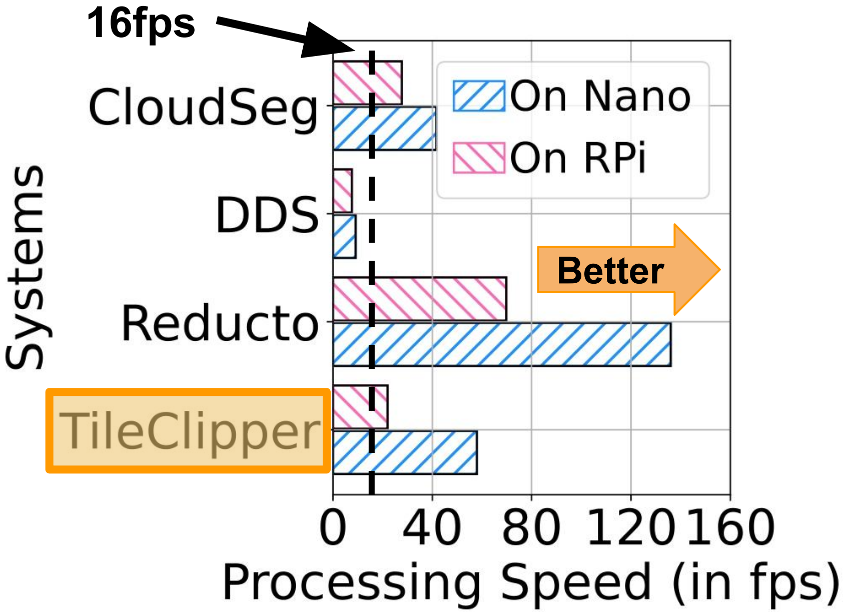
Our Trade-offs: Accuracy vs Savings vs GPU Use



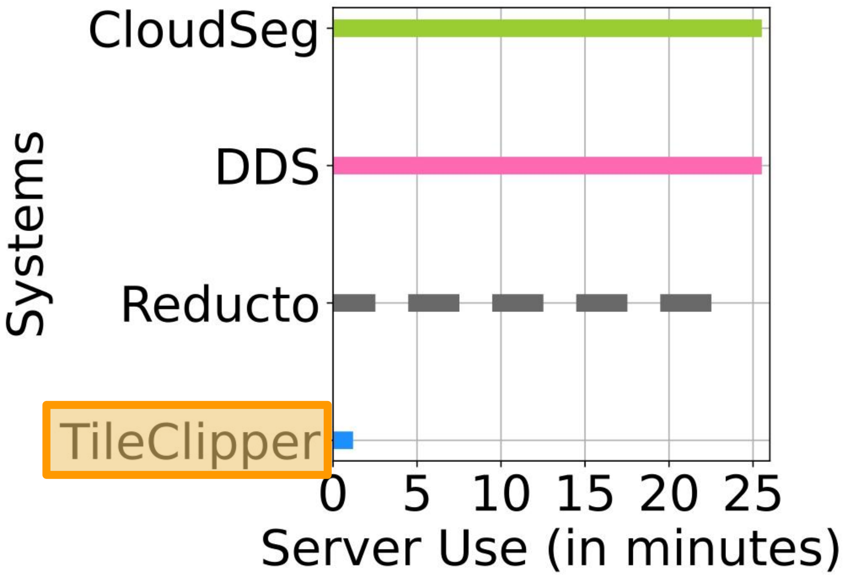


TileClipper gives best trade-off between accuracy, savings, and GPU usage

TileClipper Overheads



Camera-side



Server-side

TileClipper puts less overhead on both camera and server side

- **Most video analytics applications need minimum of 16fps**

Live Deployment Setup

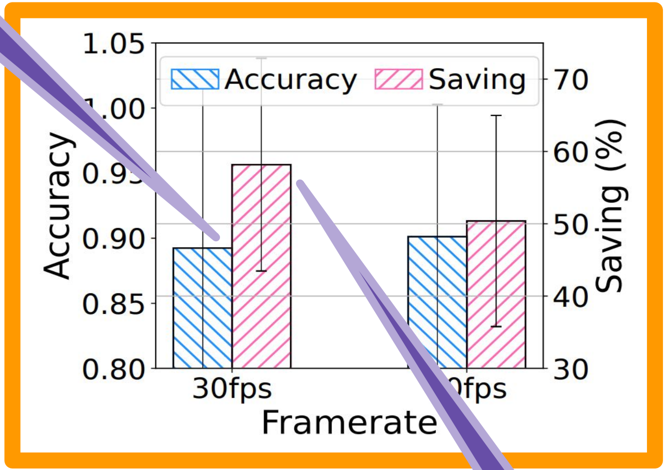
5 meters away from 2-way road

30 Kmph speed limit

Mix of vehicles and pedestrians

4G network to stream

>88% Accuracy



>50% Savings

Conclusion

Streaming videos to cloud incurs high bandwidth and latency cost

TileClipper utilizes tile filtering to amortize the streaming cost

Leverages the correlation between bitrate and # of object

Evaluated on diverse datasets under various lighting and weather conditions

Real life deployment validates its practical feasibility

ARTIFACT EVALUATED
usenix ASSOCIATION
AVAILABLE

ARTIFACT EVALUATED
usenix ASSOCIATION
FUNCTIONAL

ARTIFACT EVALUATED
usenix ASSOCIATION
REPRODUCED

Paper

shubhamch@iiitd.ac.in

Codes and Artifacts