Learning for Better Video Processing Systems

FastPath 2021

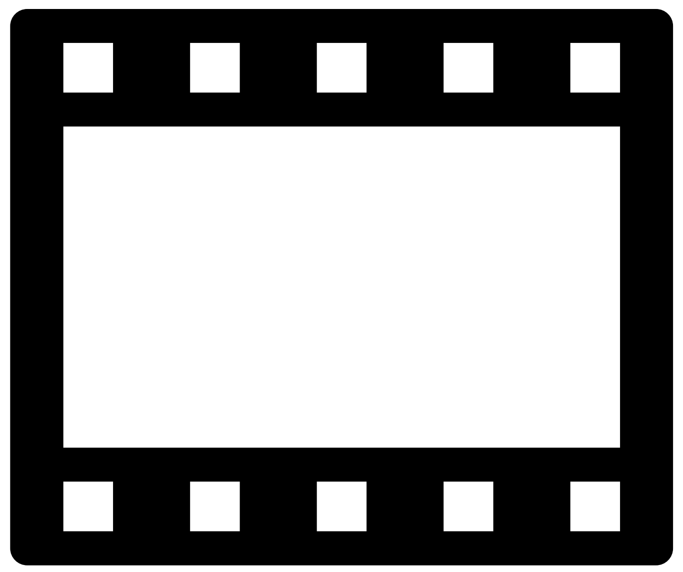
Amrita Mazumdar / Vignette AI & University of Washington

In collaboration with: *Maureen Daum, Brandon Haynes, Dong He, Magda Balazinska, Luis Ceze, Alvin Cheung, Mark Oskin*

Video is an increasingly popular source of data but presents challenges for streaming and ML processing pipelines.

Twitch streamed 75 million hours of video / month in 2020

Video communications consume 82% of internet traffic (Cisco 2019)

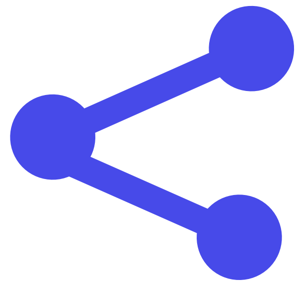


Cloud services use machine learning to process and understand video content.

Read video from storage and decompress

machine learning

features and parameters



saliency

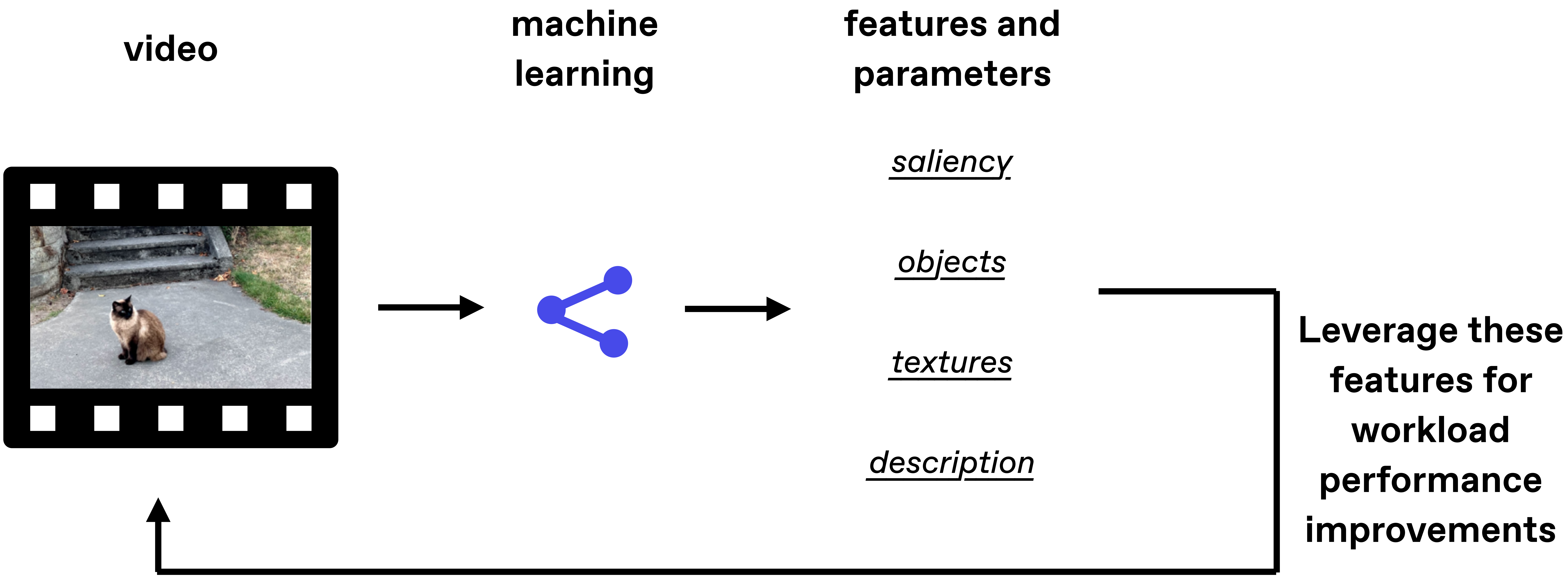
objects

textures

description

Decoding visual media takes 20x longer than accelerated DNN processing (Kang et al., VLDB 2021)

This talk: using learned features to improve performance



This talk: using learned features to improve performance

How can we use learned features to **reduce video streaming bandwidth** while maintaining quality?

Vignette (*Mazumdar et al., SoCC 2019*)

How can we use learned features to **reduce decode overhead for video analytics queries**?

TASM (*Daum et al., ICDE 2021*)

Video streaming systems trade off between visual quality and network bandwidth available.

fine details (noise, high frequencies)

color perception

fast motion

Baseline codec (HEVC) @ 20 Mbps
4 hours video playback

Source: Netflix Public Dataset

Saliency is a powerful perceptual cue for compressed video workloads.

4K 360° video
300 MB

AI-generated saliency map
only 15% of pixels are important

Source: Lo et al., MMSys 2017

Leveraging perceptual cues at scale presents design challenges.

Requires custom, outdated codecs

No integration with storage manager

No interface for applications

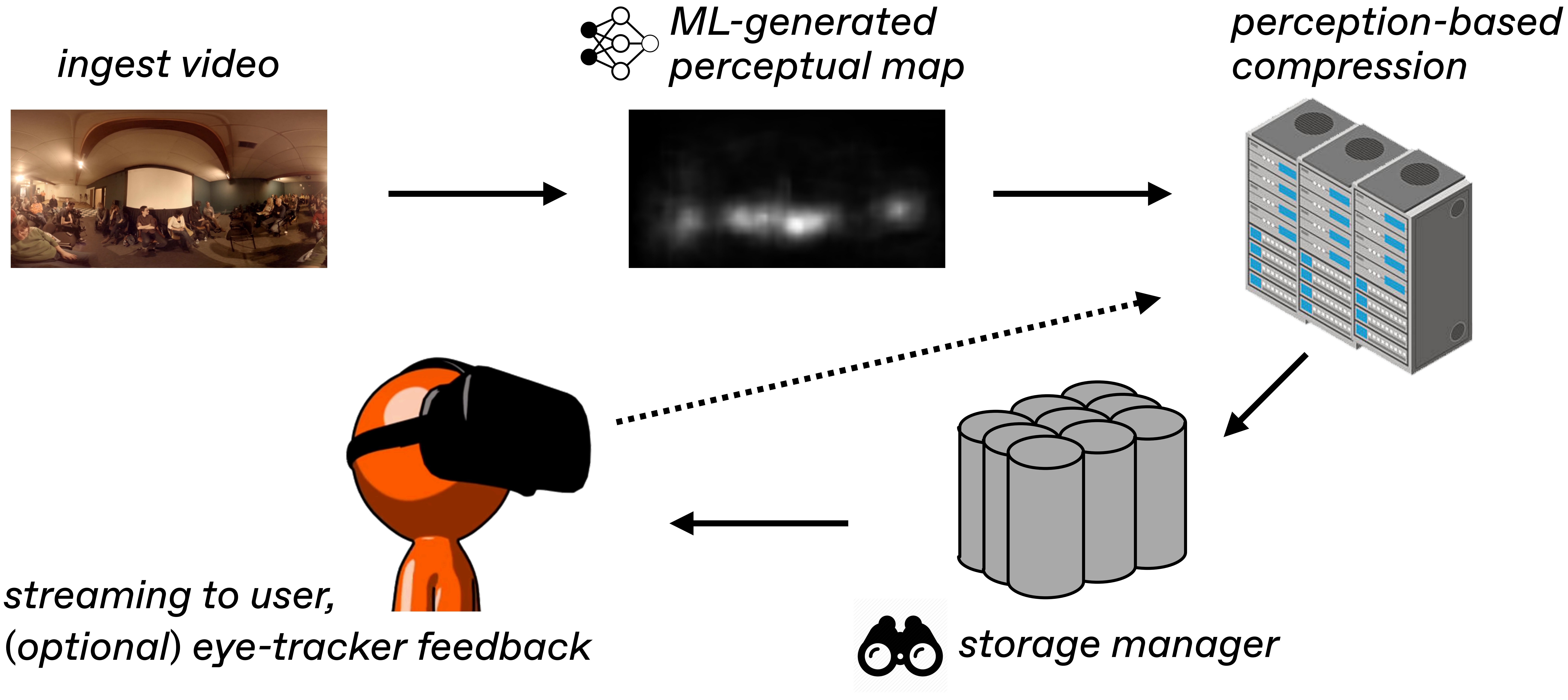
Goals:

✅ Modern codecs

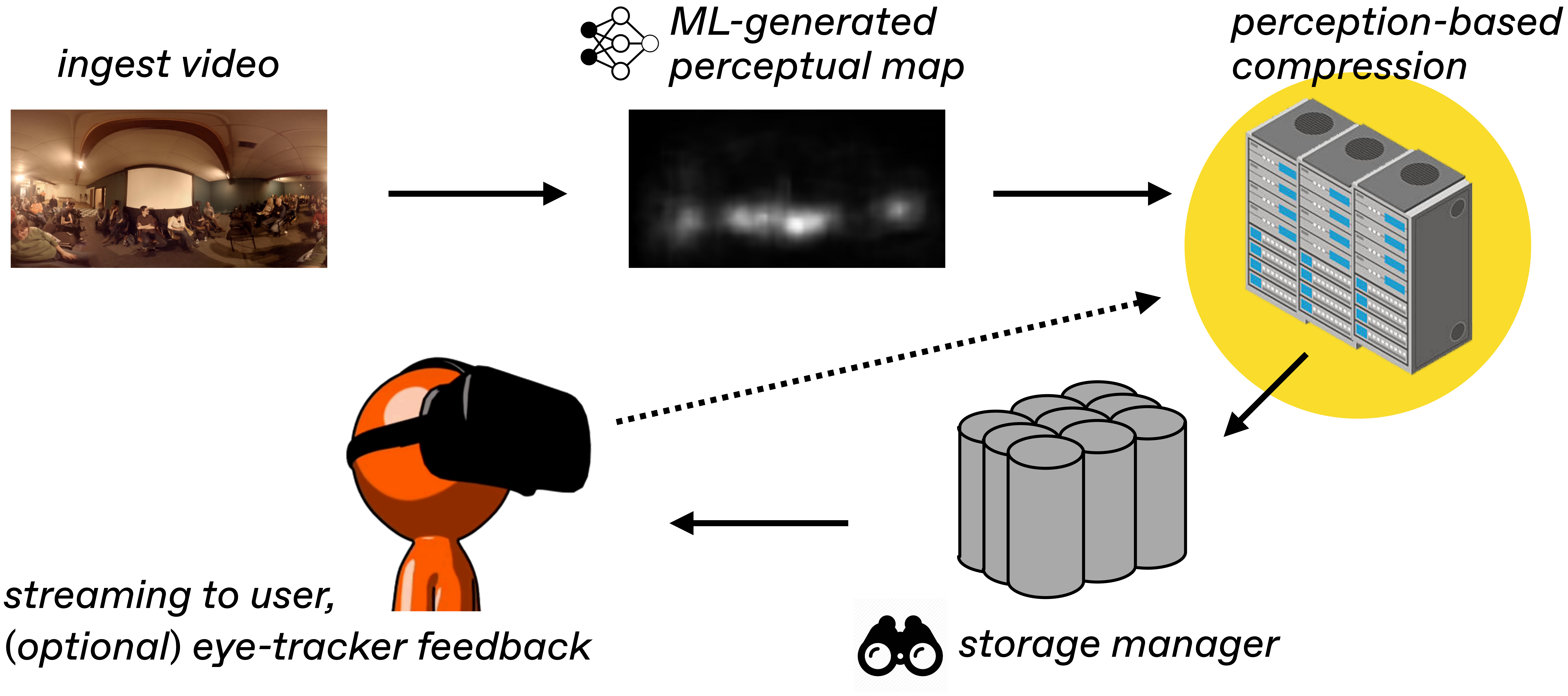
✅ API for storage

✅ Application portable

Vignette is a perception-aware video compression and storage system.



Vignette is a perception-aware video compression and storage system.



Vignette Compression uses tiles to convert saliency maps to video encoder parameters.

Automatically generate a saliency map

Split the video segment into tiles

Map saliency values to tiles

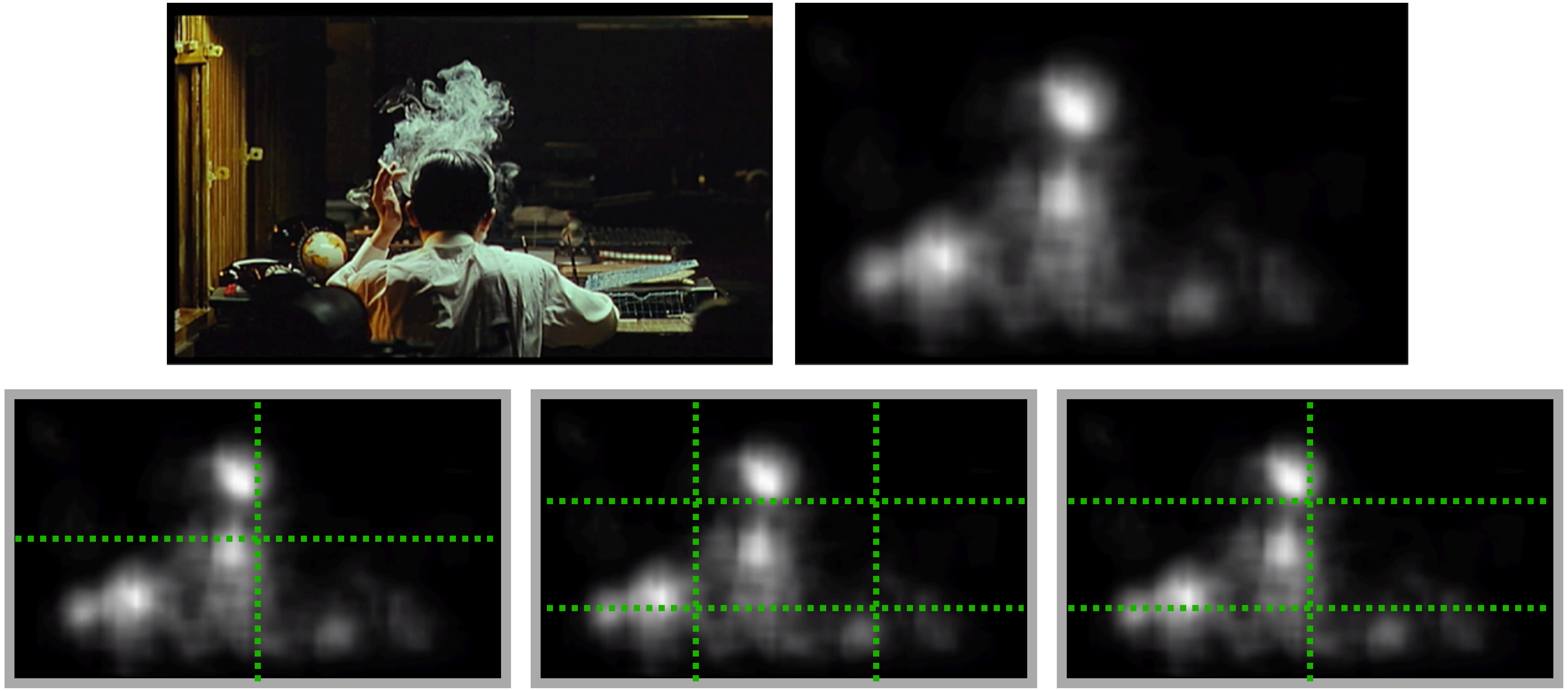
Source: Wong 2000

Vignette Compression uses tiles to convert saliency maps to video encoder parameters.

Automatically generate a saliency map

Split the video segment into tiles

Map saliency values to tiles



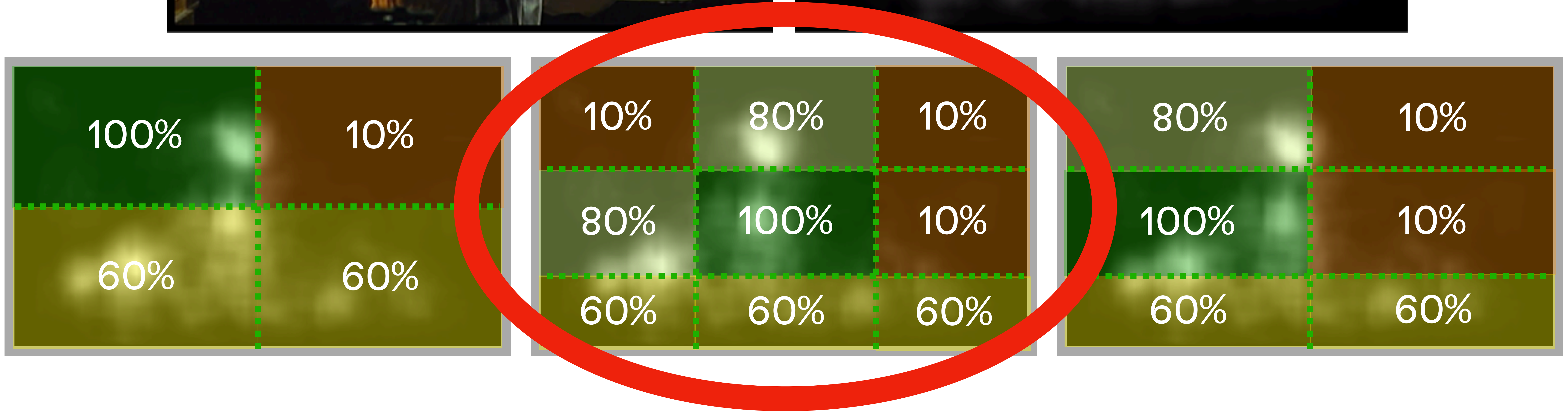
Source: Wong 2000

Vignette Compression uses tiles to convert saliency maps to video encoder parameters.

Automatically generate a saliency map

Split the video segment into tiles

Map saliency values to tiles



pick best quality, lowest overhead

Source: Wong 2000

Vignette Results

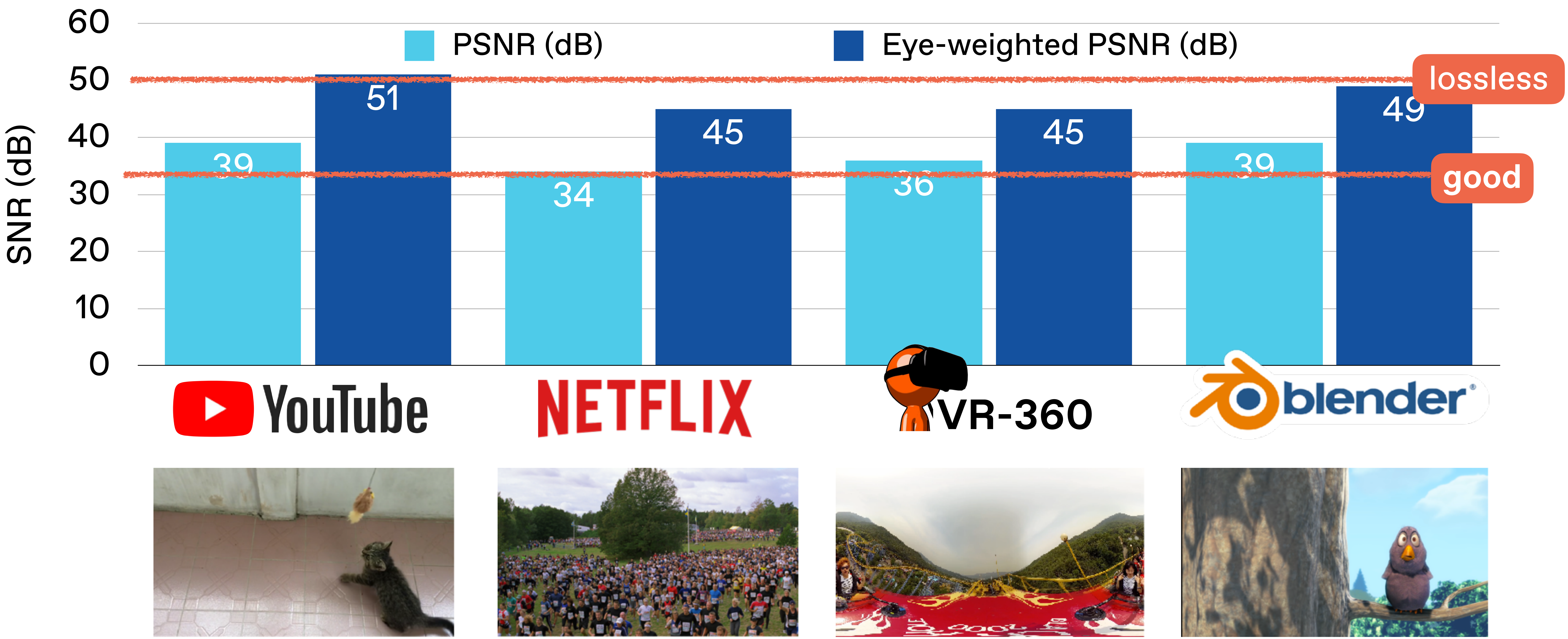
Participants either preferred Vignette or perceived no difference for 75% smaller videos.

Baseline HEVC @ 20 Mbps
4 hours video playback

Vignette @ 1 Mbps
6.5 hours video playback

Full Study Results: https://homes.cs.washington.edu/~amrita/vignette_socc19.html

Vignette videos reduce bitrate in non-salient regions, maintaining visual quality at lower storage sizes.



Vignette is a video processing system for perceptual compression and storage.

Vignette Compression
codec-agnostic perceptual video compression

Vignette Storage
storage manager for perceptually-compressed videos

Reduces storage by up to 75% with little quality loss

This talk: using learned features to improve performance

How can we use learned features to **reduce video streaming bandwidth** while maintaining quality?

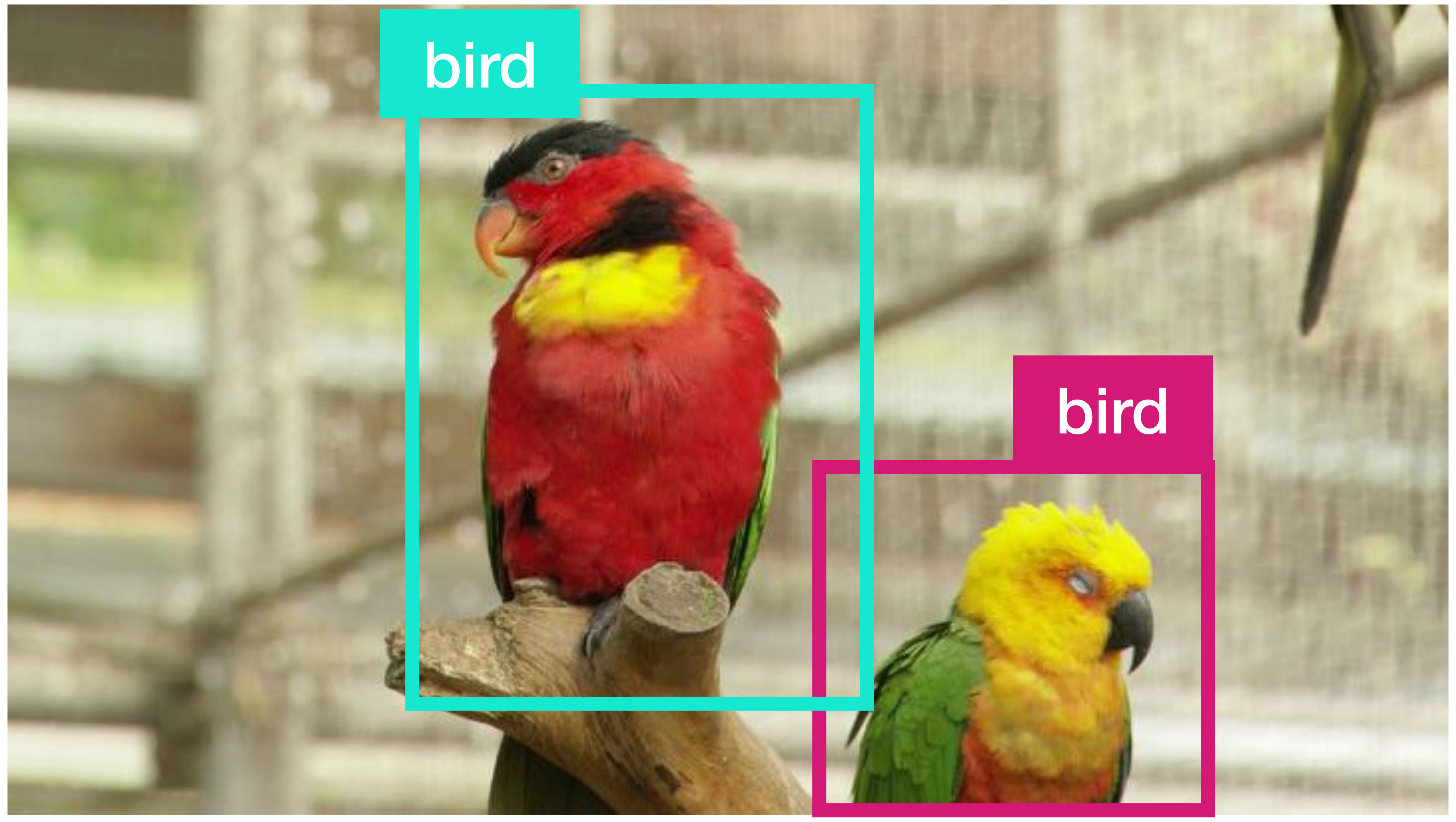
Vignette (*Mazumdar et al., SoCC 2019*)

How can we use learned features to **reduce decode overhead for video analytics queries**?

TASM (*Daum et al., ICDE 2021*)

Analytics queries extract a subset of pixels in video

```
Select bird FROM video;
```

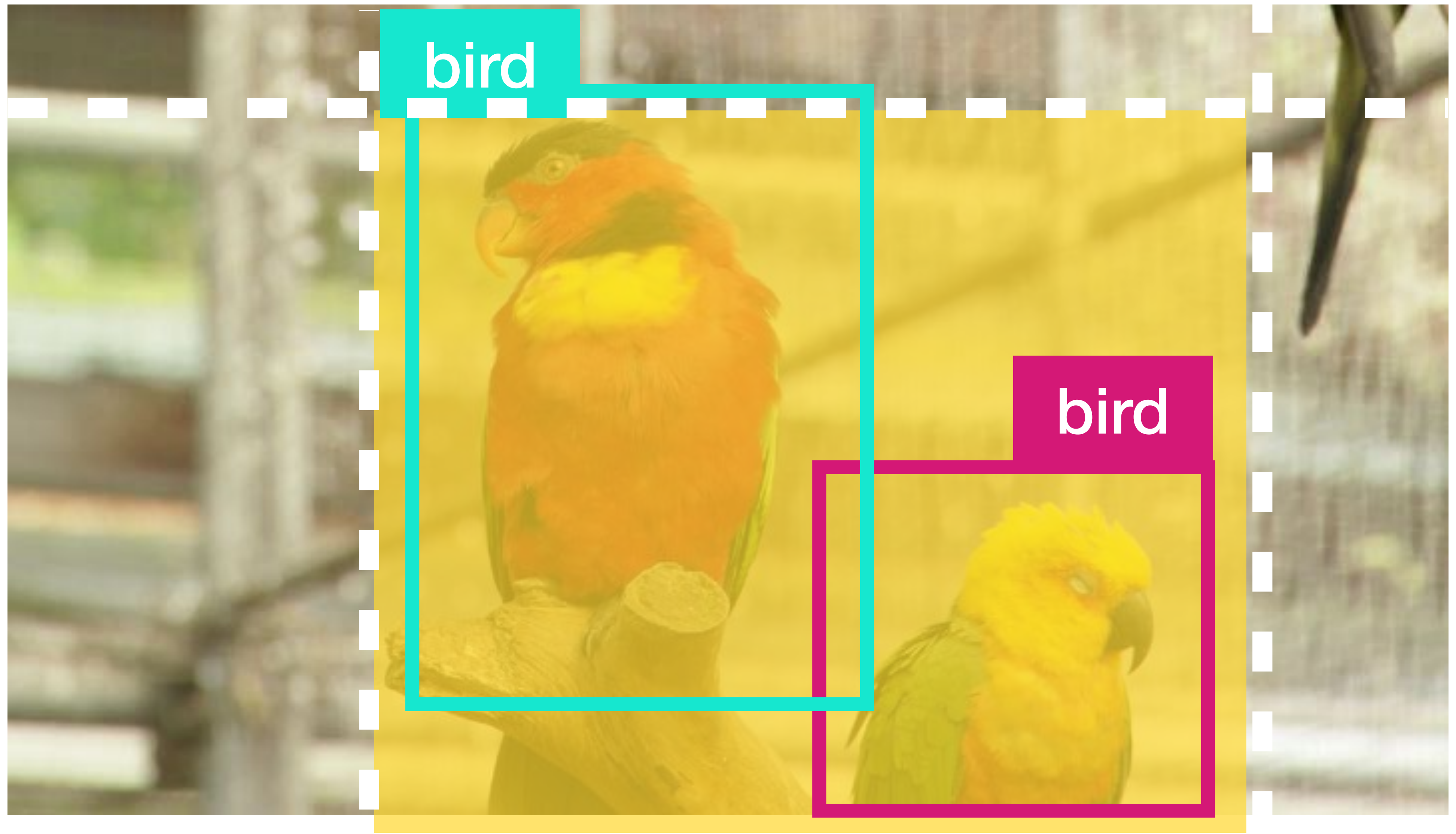


Typical workload:

- Identify objects in videos
- Extract pixels that correspond to objects of interest

Tiled video can speed up processing

```
Select bird FROM video;
```



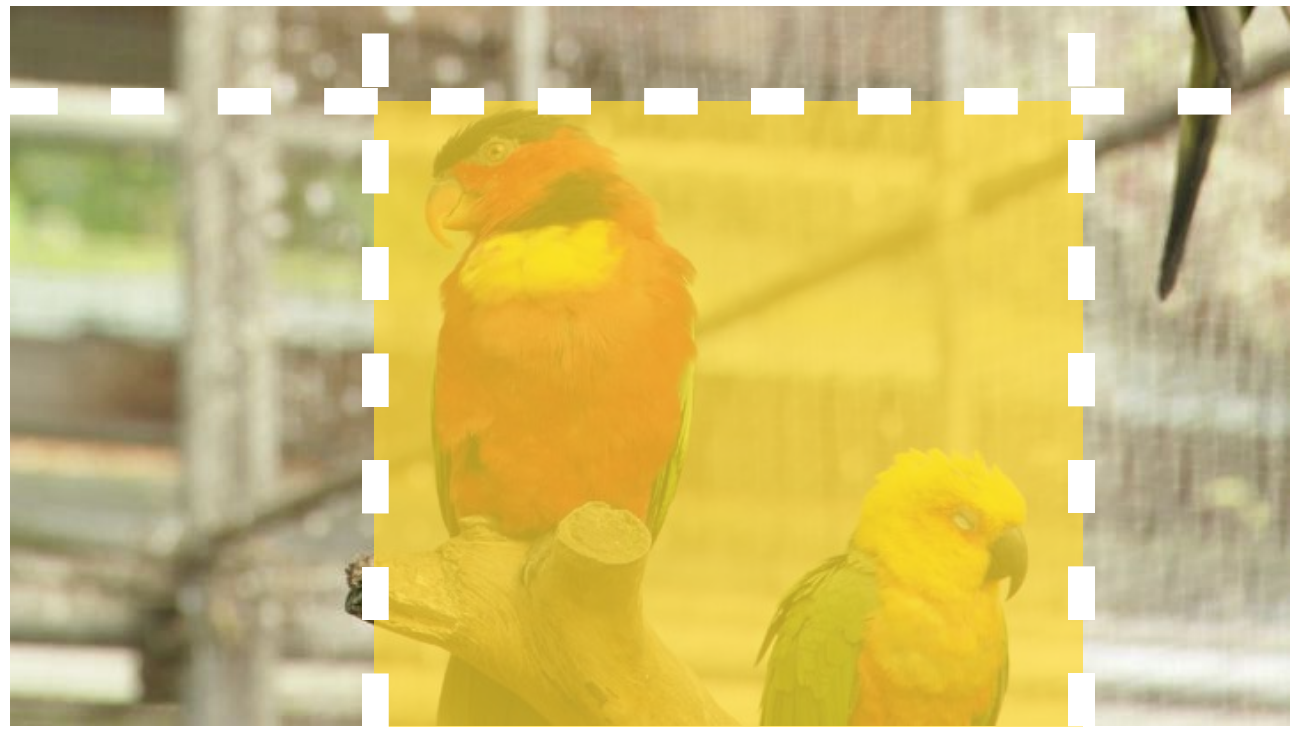
Tiles can enable *spatial random access* to video content

Knowing which tiles contain objects reduce video decode time and overall query processing time

Tiled video can speed up processing

but some tile layouts are better than others for analytics

```
Select bird FROM video;
```



Tile boundaries on objects can impact query accuracy

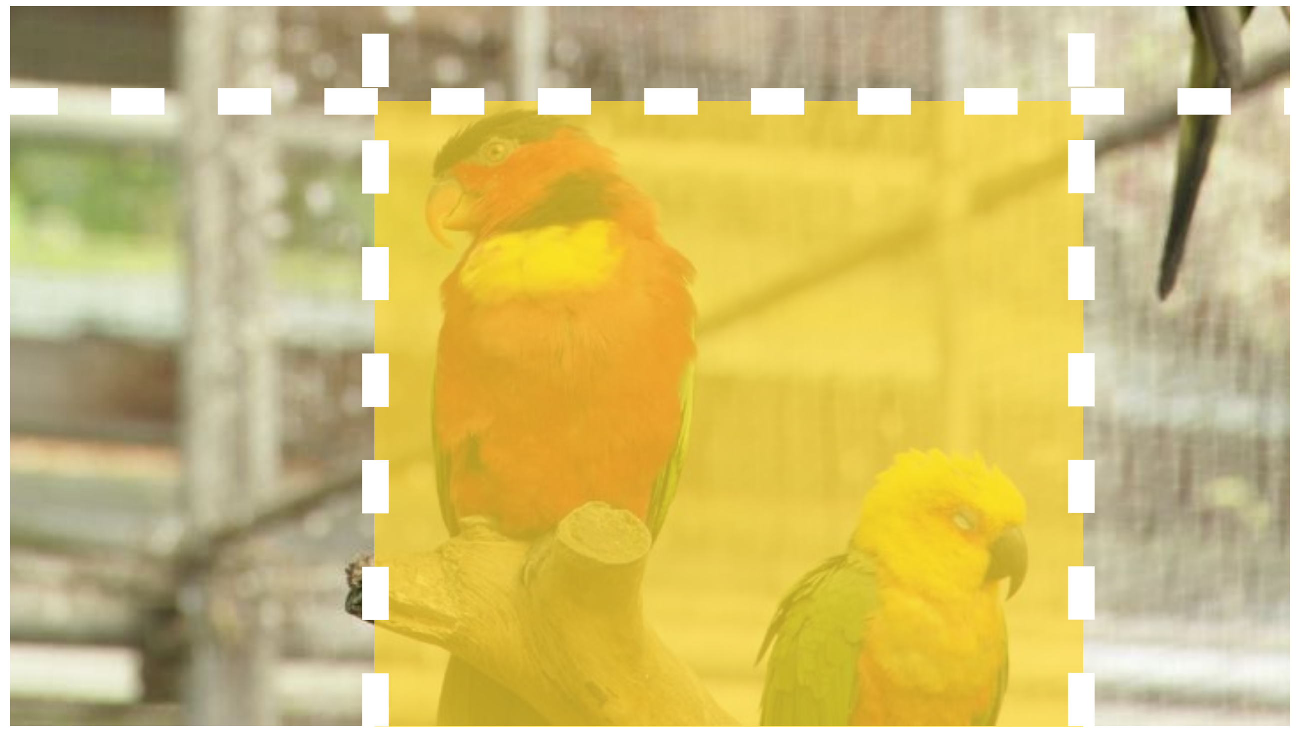
Videos can have many moving objects within the frame

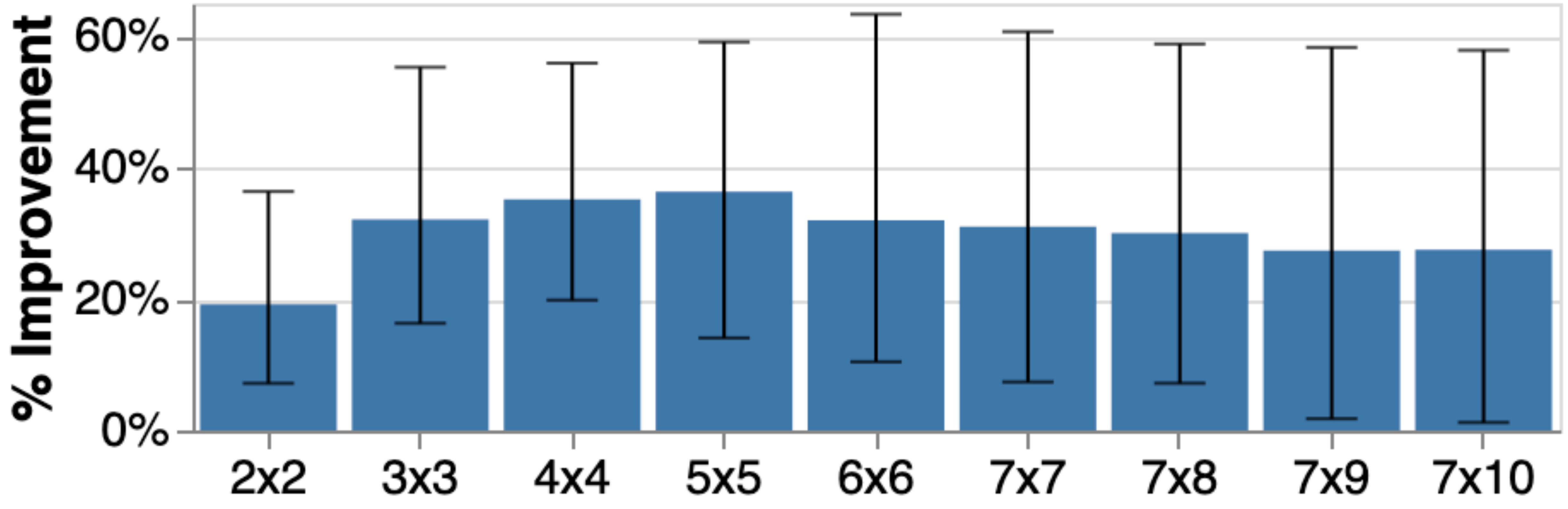
The optimal tile layout for a set of queries may not be known *a priori*

Tiled video can speed up processing

but some tile layouts are better than others for analytics

```
Select bird FROM video;
```

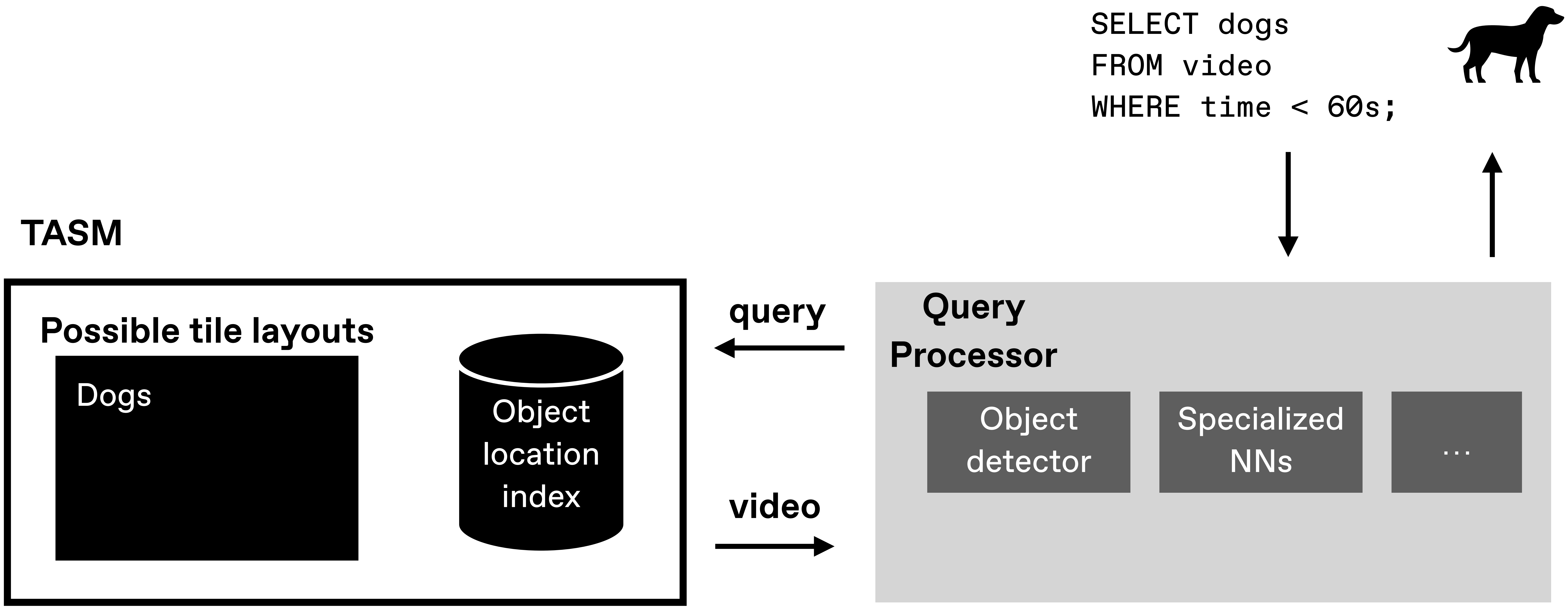




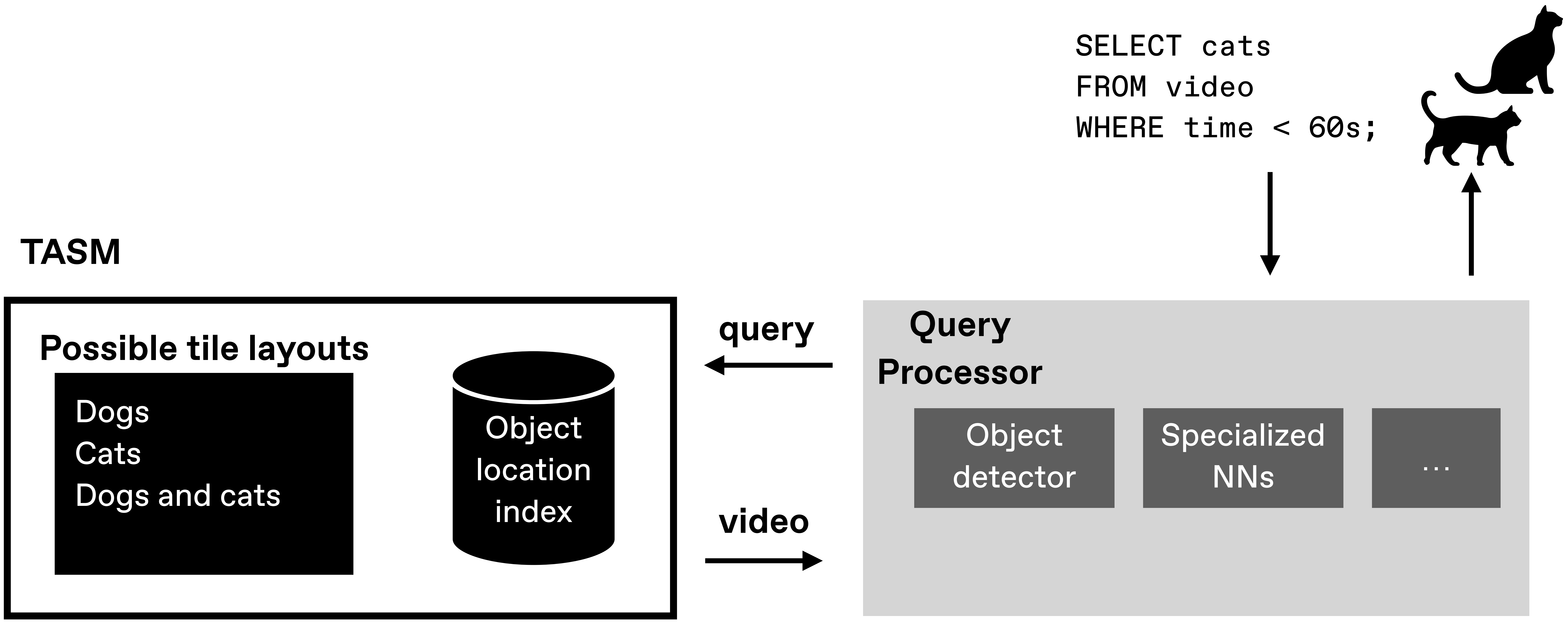
Even introducing uniform tiles can improve query performance

Overhead from too many tiles can outweigh benefits of subset selection

TASM is a storage system for video analytics queries that optimizes tile layouts for performance.

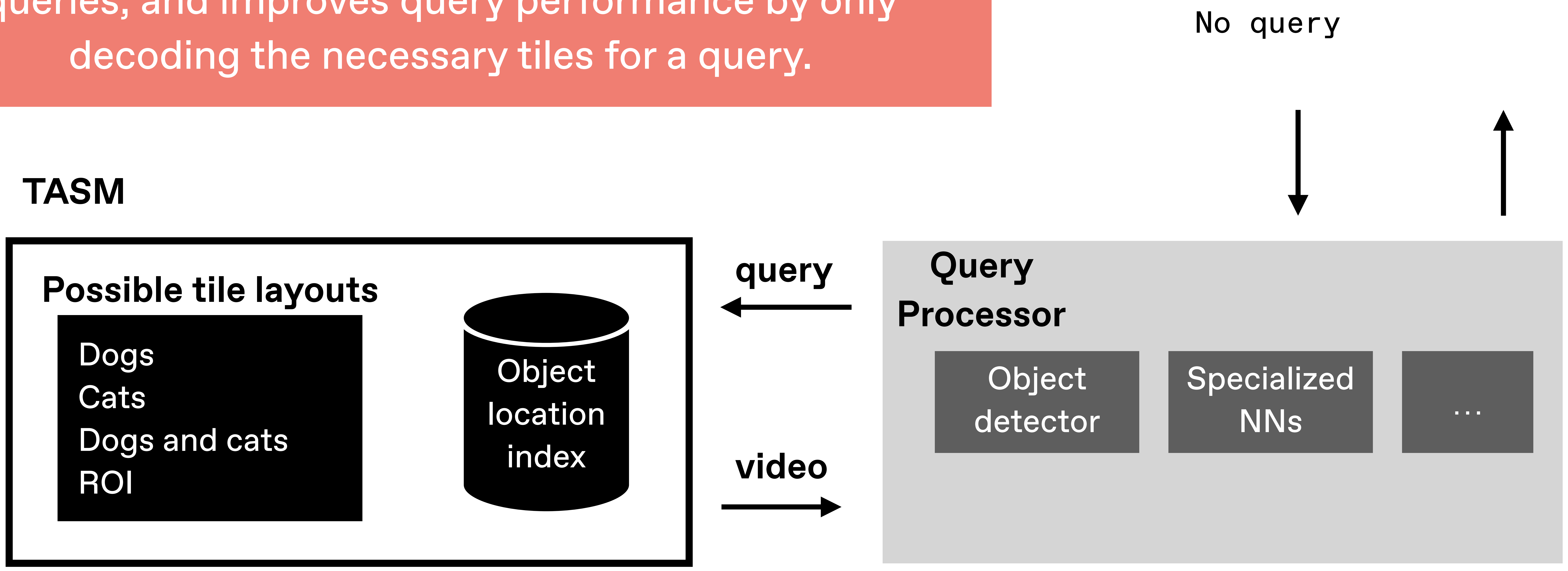


TASM is a storage system for video analytics queries that optimizes tile layouts for performance.



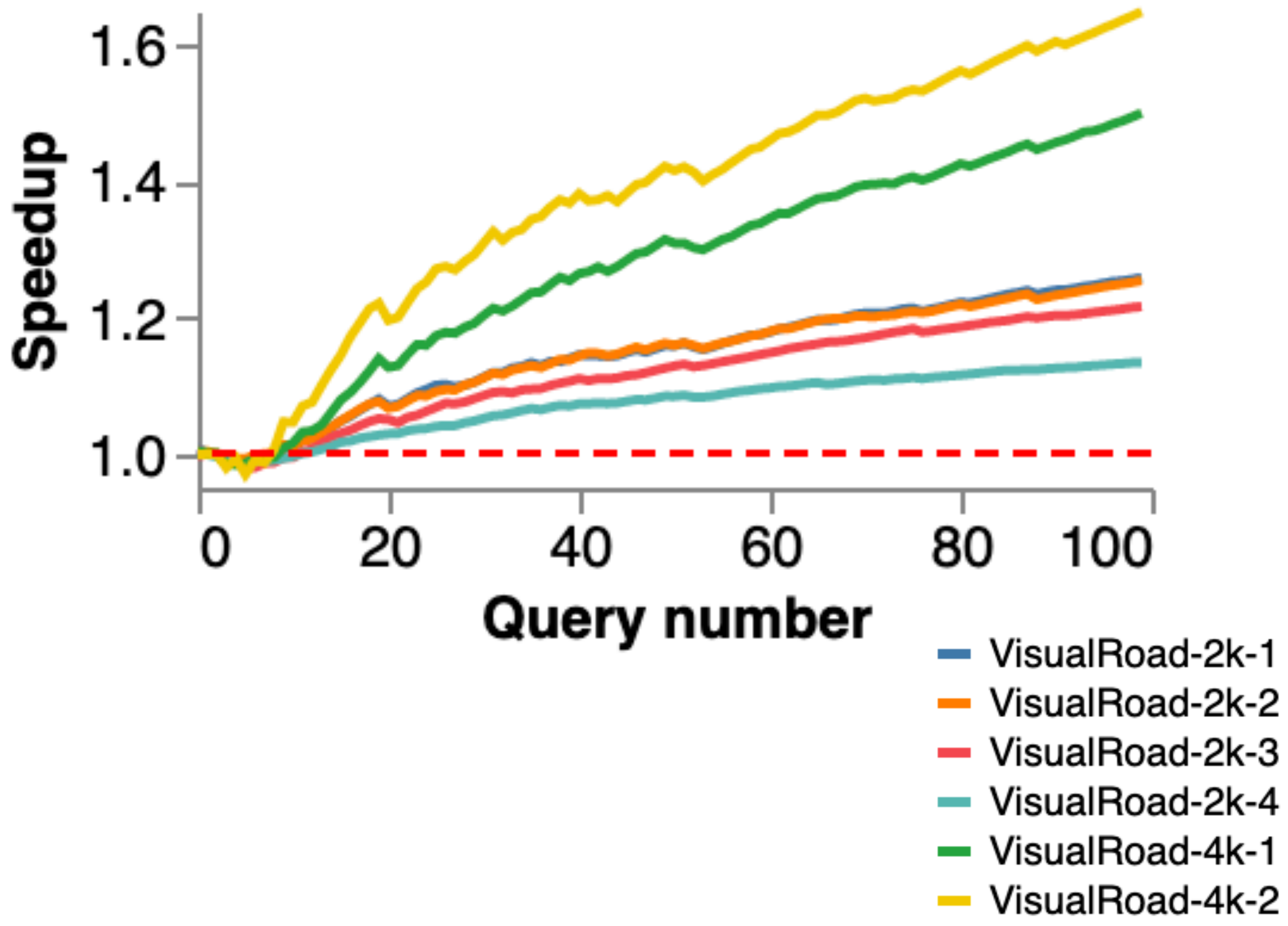
TASM is a storage system for video analytics queries that optimizes tile layouts for performance.

TASM incrementally re-tiles videos based on observed queries, and improves query performance by only decoding the necessary tiles for a query.



TASM reduces total workload runtime by processing only relevant regions of a frame.

- 60-second queries, randomly selected to be [cars, people]
- Comparing TASM with incremental regret-based tiling against untiled video
- TASM reduces total workload runtime by 12-39% across Visual Road benchmark



TASM is a storage system for video analytics queries that optimizes tile layouts for better performance.

TASM enables **spatial random access for video content**

TASM **incrementally tiles videos** as new queries are observed

Subframe selection queries show average of **50% speedup** (up to 94%)

This talk: using learned features to improve performance

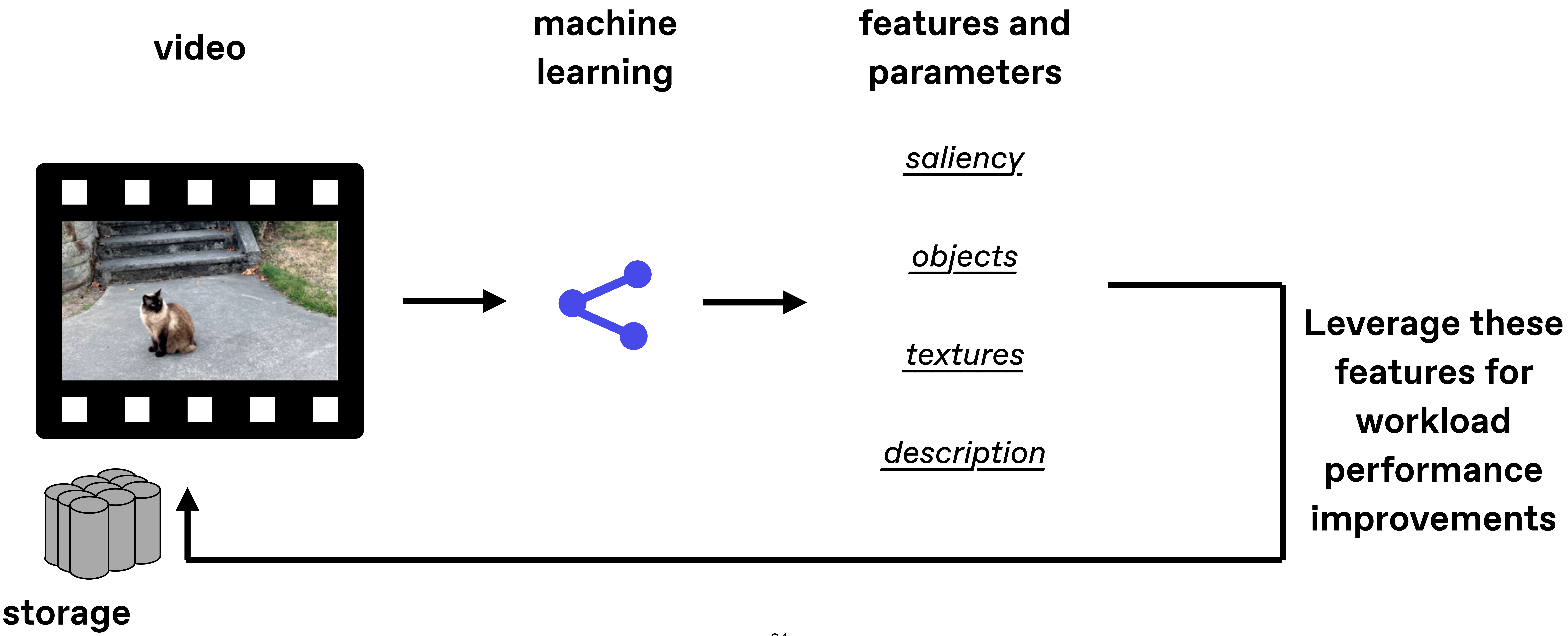
How can we use learned features to **reduce video streaming bandwidth** while maintaining quality?

Vignette (*Mazumdar et al., SoCC 2019*)

How can we use learned features to **reduce decode overhead for video analytics queries**?

TASM (*Daum et al., ICDE 2021*)

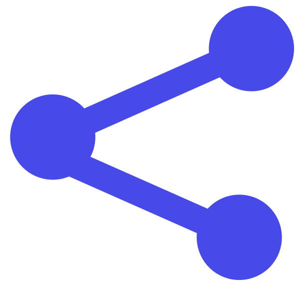
This talk: using learned features to improve performance



This talk: using learned features to improve performance

video

machine learning



features and parameters

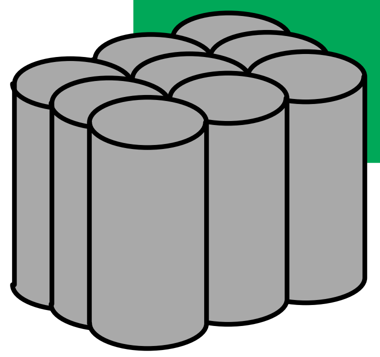
saliency

objects

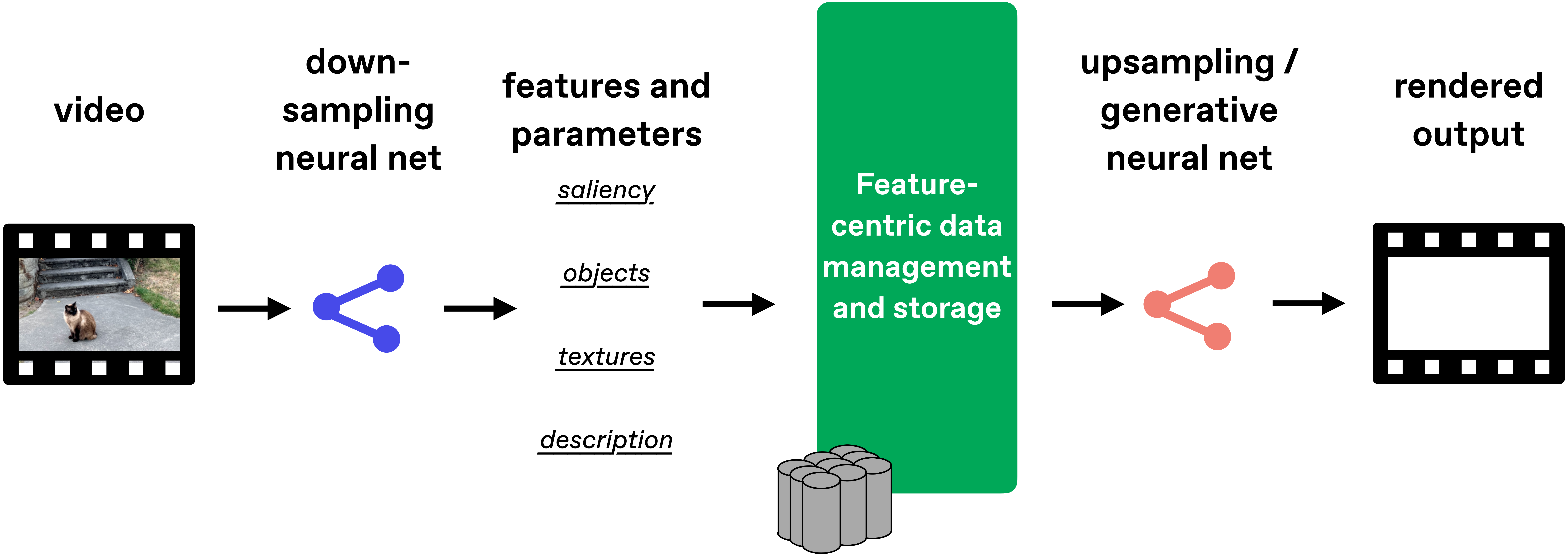
textures

description

Feature-centric data management and storage



Opportunity: depending on learned features to replace video content



Learning for Better Video Processing Systems

Thank you!

Amrita Mazumdar / Vignette AI & University of Washington

In collaboration with: *Maureen Daum, Brandon Haynes, Dong He,*
Magda Balazinska, Luis Ceze, Alvin Cheung, Mark Oskin