A Hybrid Task-Oriented Artificial Intelligence Framework for Video Compression with Embedded Object-Recognition Parameters: An Approach Beyond H.265+

Semantic-Aware Video Compression with Object-Parameter Side-stream (SAVC-OPS)

Samir Khanji

Informatics and Communications Department, Faculty of Engineering, Ebla Private University, Idlib, Syria

---***---

Abstract - *The exponential growth of surveillance and machine-vision workloads has made bitrate and storage efficiency a first-order systems constraint. Vendor-optimized codecs such as H.265+ improve upon High Efficiency Video Coding (HEVC) for static-heavy scenes through background modeling, noise suppression, and long-term average bitrate control, while newer standards such as VVC further raise compression efficiency for human viewing [2]. Nevertheless, these codecs remain primarily oriented toward human perception and do not, by themselves, define a standardized persistent object-recognition parameter track inside the compressed representation (some NVR products may store analytics metadata outside the codec). Consequently, archives must often be re-analyzed with costly neural inference whenever semantic queries are required. Related research on Video Coding for Machines (VCM) and Feature Coding for Machines (FCM) reframes compression around task utility rather than pixel fidelity alone [1], [8], [16]. This paper proposes SAVC-OPS (Semantic-Aware Video Compression with Object-Parameter Side-stream), a hybrid framework that combines: (i) task-aware semantic bit allocation based on detected regions of interest [1], [9]; (ii) a temporally compressed object-parameter side-stream that stores bounding boxes, class labels, confidences, track identifiers, and optional embeddings; and (iii) an optional neural feature-coding path for machine-primary consumption [8], [15]. Using a calibrated illustrative analytical simulation (not measured bitstreams), mean bitrate decreases from 3.17 Mbps (H.265) and 1.33 Mbps (H.265+) to 1.17 Mbps (SAVC-A) and 0.74 Mbps (SAVC-B). Thirty-day storage per camera falls from about 421 GB under H.265+ to 370 GB (SAVC-A) and 234 GB (SAVC-B). Under a comparable storage budget, illustrative mAP@0.5 improves from 0.77 to 0.85–0.88, while semantic query latency can fall by roughly 400× through OPS indexing (hardware-dependent assumption). Five incremental development proposals are quantified with respect to storage, task accuracy, retrieval efficiency, edge cost, and privacy. The results indicate that moving from scene-aware compression to task-aware, semantically annotated compression is a practical path beyond H.265+ for analytics-centric video systems.*

Key Words: artificial intelligence video compression; H.265+; HEVC; Video Coding for Machines; object recognition; feature coding; semantic storage; analytical simulation; surveillance systems.

1. INTRODUCTION

1.1 Motivation and Context

Video is the dominant contributor to digital storage and network traffic in contemporary infrastructure, particularly in video surveillance, smart-city sensing, autonomous platforms, and remote healthcare. The transition from 1080p to 4K/8K capture, together with denser camera deployments, multiplies both bandwidth demand and archival cost. Over the last decade, standardization and industrial engineering have therefore focused on improving compression efficiency while preserving acceptable visual quality for human operators.

Modern standardized codecs continue to push rate–distortion efficiency for human viewing: Versatile Video Coding (VVC / H.266) approximately halves bitrate relative to HEVC for comparable quality in many conditions [2]. In parallel, surveillance vendors introduced proprietary HEVC-oriented extensions commonly marketed as H.265+ (and closely related "smart" codecs). H.265+ is not an ITU-T/ISO international standard; it denotes vendor optimizations layered on HEVC. These extensions exploit properties typical of fixed-camera deployments: long quasi-static backgrounds, intermittent motion, and sensor noise that would otherwise consume bits without operational value. Reported industrial examples indicate that average bitrate under H.265+ may fall to roughly one-third of H.265 in low-activity office scenes, although the gain is strongly content dependent.

Despite these advances, the consumer of video is changing. In an increasing fraction of deployments, the primary consumer is not a human viewer but a machine-vision pipeline that performs detection, tracking, re-identification, counting, or anomaly analysis. Under this regime, peak signal-to-noise ratio (PSNR) and related perceptual metrics are incomplete success criteria. What matters is the preservation of task-relevant information at the lowest feasible rate, together with the ability to retrieve semantic events from archives without repeatedly decoding and re-inferring every frame.

1.2 Limitations of H.265+-Centric Pipelines

H.265+ improves the human-oriented rate–distortion trade-off for surveillance content, yet three structural limitations remain.

First, bit allocation is driven by background/motion/noise cues rather than by semantic task relevance. Regions that are visually static may still contain small or slowly moving objects that are critical for detection. Conversely, textured background motion may attract bits that contribute little to downstream accuracy.

Second, object-recognition parameters if computed at all on the edge are typically treated as ephemeral analytics outputs. They are not stored as a compact, synchronized component of the compressed archive. Later investigations therefore re-run detectors and trackers on reconstructed pixels, multiplying compute cost and delaying response.

Third, emerging machine-to-machine scenarios may not require a high-fidelity pixel archive at all. Feature Coding for Machines (FCM) and related approaches suggest that intermediate neural representations can be compressed more efficiently for remote inference [1], [8], [16]. H.265+ does not natively expose such a path.

1.3 Research Gap

The literature now contains strong strands on learned codecs [3], [4], [5], [17], [18], Video Coding for Machines (VCM) [1], [6], [7], ROI retargeting, reinforcement-learning bit control [9], and feature coding [8], [15], [16]. However, practical surveillance stacks still largely separate (a) pixel compression optimized for human viewing from (b) analytics executed as a post-process. A unified representation that jointly optimizes storage, task accuracy, and semantic retrievability while remaining deployable on HEVC-compatible infrastructure remains insufficiently articulated for systems that currently rely on H.265+.

1.4 Contributions

This paper makes the following contributions:

- Framework design. We introduce SAVC-OPS, a hybrid architecture that couples semantic bit allocation with a persistent object-parameter side-stream and an optional feature-coding track.
- Formalization. We define the side-stream syntax, temporal differential coding principles, and a multi-objective rate–task–perception loss suitable for mode selection (human, hybrid, machine).
- Comparative evaluation protocol. We specify fair comparison axes against H.265 and H.265+: bitrate by scene class, thirty-day storage, rate–task curves, perceptual metrics, detection/tracking metrics, and semantic query latency.
- Analytical simulation study. We report quantified comparisons and seven publication-ready figures based on a calibrated model, clarifying where gains over H.265+ are expected to be large or modest.
- Development roadmap. We decompose the framework into five proposals (P1–P5), each with advantages, challenges, and simulated outcomes, and provide a staged adoption plan.

1.5 Paper Organization

Section 2 reviews related work. Section 3 states the problem, objectives, and hypotheses. Section 4 presents the SAVC-OPS methodology. Section 5 details the simulation setup. Section 6 reports statistical comparisons and figures. Section 7 develops proposals with advantages and results. Section 8 discusses implications and limitations. Section 9 concludes the paper.

2. RELATED WORKS

2.1 From H.264 to HEVC and Surveillance-Oriented H.265+

HEVC and its successor VVC introduced successively stronger coding tools (flexible partitions, refined prediction, and improved entropy coding), with VVC targeting roughly a further ~50% bitrate reduction over HEVC under matched quality [2]. Industrial H.265+ extensions typically emphasize three tools tailored to surveillance: (i) background-model-based prediction, (ii) background noise suppression to avoid spending bits on non-informative sensor fluctuations, and (iii) long-term average bitrate control that reallocates bits across busy and idle periods of a day. These tools can dramatically reduce storage for low-motion cameras. Their optimization target, however, remains a human-viewable reconstruction rather than a machine task metric such as mean average precision (mAP) or multi-object tracking accuracy (MOTA) [1], [13].

2.2 Learning-Based Image and Video Compression

Deep learning has produced end-to-end nonlinear transform codecs with learned entropy models that are competitive with classical still-image standards at selected operating points [5]. Extending these ideas to video has progressed through predictive/conditional frameworks such as DCVC [3], transformer-based temporal entropy models such as VCT [4], and content-adaptive neural codecs [17], alongside broader surveys of intelligent video coding [18]. Real-time complexity and edge constraints nonetheless remain challenging. Hybrid designs are therefore attractive: a conventional HEVC/VVC core is retained for hardware decode compatibility, while neural modules guide preprocessing, mode decisions, or post-processing. SAVC-OPS follows this hybrid philosophy for near-term deployability.

2.3 Video Coding for Machines (VCM)

VCM reframes compression as a rate–task problem [1]. Instead of minimizing pixel distortion alone, encoders allocate quality to regions and features that preserve automated analysis. Representative techniques include ROI-aware quantization, saliency- or detector-guided CTU importance, and multi-scale feature pipelines evaluated under MPEG VCM common test conditions [6], [7]. Recent peer-reviewed studies report large BD-rate gains for machine tasks under feature/video VCM configurations [6], and standardization surveys summarize use cases, metrics, and proposal trends [7], [8]. These results motivate semantic bit allocation in SAVC-OPS, but most VCM pipelines still treat ROI/feature metadata as encoder side information rather than as a durable archival track for semantic search.

2.4 Feature Coding for Machines (FCM)

FCM compresses intermediate neural activations rather than (or in addition to) pixels [8], [15], [16]. This reduces bandwidth for remote inference, supports compute offloading, and can improve privacy by avoiding transmission of the full scene. Analyses of machine-to-machine coding indicate that conventional inner codecs such as HEVC/VVC remain highly effective for feature tensors with only modest task BD-rate differences between HEVC and VVC in reported FCM experiments which is encouraging for hardware reuse [8]. Learnable multi-scale feature compression further improves the rate–task curve for VCM-style pipelines [6], [15]. SAVC-OPS incorporates an optional feature track for machine-primary modes while retaining a pixel track for auditability when required.

2.5 Task-Aware Control and Reinforcement Learning

Recent work explores reinforcement learning agents that select block-level quantization actions to optimize downstream task performance under rate constraints. In particular, RL-RC-DoT demonstrates task-aware macro-block QP control on top of a standard encoder without requiring the task network at inference time [9]. Such controllers can outperform static ROI heuristics when scene statistics drift, but they also introduce training cost and domain-transfer challenges. We therefore treat RL-based allocation as an incremental proposal (P4) rather than a mandatory component of the baseline stack.

2.6 Positioning of This Work

Table -1 summarizes the conceptual positioning relative to VCM/FCM surveys and codecs [1], [2], [7], [8]:

Table -1: Simulation Parameters.

Approach	Primary consumer	Bit decision unit	Object parameters persisted?	Machine-primary path
H.265/HEVC / VVC **[2]**	Human	R-D (pixels)	No	No
H.265+	Human (surveillance)	Background/ motion/noise	No	No
Typical VCM ROI tools **[1], [6], [7]**	Machine (+ human)	Semantic ROI	Transient / limited	Partial
FCM **[8], [15], [16]**	Machine	Feature rate–accuracy	Features, not boxes	Yes
SAVC-OPS	Human and/ or machine	Semantic map + task loss	**Yes (OPS track)**	**Optional**

3. PROBLEM STATEMENT, OBJECTIVES, AND HYPOTHESES

3.1 Problem Statement

Let $V = \{x_t\}_{t=1}^{T}$ denote a video sequence. A conventional surveillance encoder produces a bitstream B_v minimizing a perceptual rate–distortion objective. An analytics module may later compute object sets O_t from the reconstructed frames $\hat{x}_t$. The operational cost of an archive is then not only $|B_v|$ but also the repeated inference cost whenever semantic queries are issued.

$$C = \mathrm{Mux}(B_v, B_o, B_f)$$

that jointly reduces storage relative to H.265+, preserves or improves task metrics, and makes B_o (object parameters) and optionally B_f (features) first-class, queryable components of the archive.

3.2 Research Question

How can an AI-assisted compression framework surpass H.265+ in storage efficiency and task utility for analytics-centric video, while embedding object-recognition parameters in the compressed representation?

3.3 Objectives

- Design a hybrid architecture compatible with HEVC-class cores.
- Specify a compact, temporally coded object-parameter side-stream.
- Define a multi-axis evaluation protocol (rate, perception, task, retrieval).
- Quantify expected gains via analytical simulation and structured proposals.

3.4 Hypotheses

- **H1:** Under fixed target detection accuracy, semantic bit allocation reduces total size relative to H.265+ for medium and crowded scenes.
- **H2:** The storage overhead of OPS is small, yet semantic retrieval latency decreases by orders of magnitude.
- **H3:** For machine-primary consumption, a feature-augmented path outperforms a pixel-only path on the rate–task curve.

4. PROPOSED METHODOLOGY: THE SAVC-OPS FRAMEWORK

4.1 Architectural Overview

Fig -1 presents the architecture. Incoming frames are processed by a perception unit (detector and tracker). The resulting object set induces a semantic importance map that drives bit allocation in a video core codec. In parallel, object parameters are differentially coded into an OPS track. Optionally, intermediate features are coded into a feature track. All tracks share a synchronization timeline and may be accompanied by a lightweight semantic index.

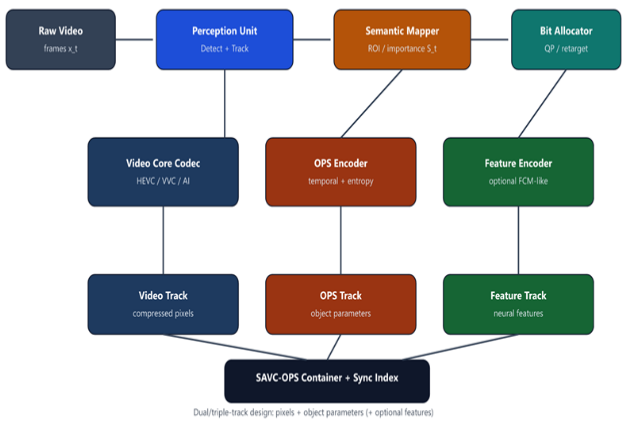

Fig -1: . SAVC-OPS architecture: perception, semantic mapping, bit allocation, video core, OPS encoder, optional feature encoder, and multiplexed container.

4.2 Perception Unit and Object Parameterization

For each frame x_t, the perception unit outputs

$$O_t = \{o_{i,t}\}_{i=1}^{N_t}, \qquad o_{i,t} = \{b_{i,t}, c_{i,t}, p_{i,t}, id_{i,t}, e_{i,t}\},$$

where $b_{i,t}$ is a normalized bounding box, $c_{i,t}$ a class index, $p_{i,t}$ a confidence score, $id_{i,t}$ a temporal track identifier, and $e_{i,t}$ an optional embedding (e.g., for re-identification). Lightweight real-time detectors such as YOLOv8-class models **[11]** and association trackers such as ByteTrack **[10]** are appropriate on edge hardware.

4.3 Semantic Importance Map

A dense map $S_t(u,v) \in [0,1]$ is constructed by rasterizing boxes/masks, dilating them by a motion-predicted margin, and assigning a low residual weight to long-term background. Formally,

$$S_t = \Phi\big(O_t,\, O_{t-1},\, \hat{M}_t\big),$$

where $\hat{M}_t$ denotes predicted motion and Φ is a deterministic fusion operator. Unlike H.265+ background models, S_t is class- and task-aware: a small pedestrian in a textured plaza can receive higher priority than large foliage motion.

4.4 Semantic Bit Allocation

Given S_t, the allocator produces codec control variables. Two practical realizations are supported:

- **ROI-QP control.** Coding tree units overlapping high-importance regions receive $\Delta\mathrm{QP} < 0$; background regions receive $\Delta\mathrm{QP} > 0$.

- **ROI retargeting.** The frame is geometrically warped so that ROI regions occupy more pixels before encoding; inverse geometry is signaled as metadata for reconstruction.

A hybrid policy may apply mild retargeting followed by residual QP shaping. The video core may be x265/HEVC hardware, VVC [2], or a learned codec [3], [4], provided decode compatibility constraints of the deployment are respected.

4.5 Object-Parameter Side-stream (OPS)

OPS persists recognition outputs as a compressed archival track rather than discarding them after ROI decisions. Temporal coding proceeds as follows:

- Associate objects across time via id.
- Predict box trajectories (constant velocity or Kalman-style predictors).
- Entropy-code residuals $\Delta b_{i,t} = b_{i,t} - \hat{b}_{i,t}$, class persistence flags, and confidence deltas.
- Emit birth/death events instead of repeating full state every frame.
- Optionally insert semantic keyframes every N frames for random access.

In the simulation, OPS contributes approximately 2.1% of the video-track size under SAVC-A small relative to its retrieval benefit. A logical container is:

SAVC-OPS Container

- video_track # HEVC/VVC/AI bitstream
- ops_track # compressed object parameters
- feature_track? # optional neural features
- sync_timeline # PTS/frame indices
- index_sidecar # optional inverted index (class/time/camera)

4.6 Optional Feature-Coding Path

When the consumer is primarily a machine, intermediate backbone features f_t are quantized and temporally coded:

$$B_f = \text{FeatureEncode}(f_t, f_{t-1}).$$

Downstream heads (detection/tracking) operate on reconstructed features, enabling aggressive reduction of the pixel track to a low-rate preview or even an empty track in pure machine mode. This path aligns with FCM concepts [8], [15], [16] while remaining optional in hybrid deployments that require human audit frames.

4.7 Operating Modes

Mode	Active tracks	Intended use
SAVC-A	Video + OPS	Hybrid archive: human review + semantic search
SAVC-B	Low-rate video/ preview + OPS + features	Machine-heavy analytics with limited auditability
Human-fallback	Video (+ transient ROI)	Compatibility mode close to classical codecs

4.8 Multi-Objective Training / Control Loss

Allocation parameters are selected to minimize

$$\mathcal{L} = \lambda_r R_v + \lambda_o R_o + \lambda_f R_f + \lambda_t \mathcal{L}_{\text{task}} + \lambda_p \mathcal{L}_{\text{perc}},$$

where R_v, R_o, R_f are track rates, $\mathcal{L}_{\text{task}}$ penalizes task degradation (or uses a differentiable proxy), and $\mathcal{L}_{\text{perc}}$ is an optional perceptual term for human-facing modes. Weight schedules differ by mode: human archives raise λ_p; machine archives raise λ_t and λ_f.

4.9 Encoding Algorithm

```
Algorithm 1: SAVC-OPS Encode
Input:  video V, mode ∈ {human, hybrid, machine}
Output: container C
for each frame x_t in V do
  O_t ← DetectAndTrack(x_t)
  S_t ← BuildSemanticMap(O_t, history)
  if mode ≠ machine_only then
    x̃_t ← PreprocessROI(x_t, S_t)
    B_v ← VideoEncode(x̃_t, S_t)
  else
    B_v ← PreviewEncode(x_t) or Ø
  B_o ← OPSEncode(O_t, O_{t-1})
  if mode ∈ {hybrid, machine} then
    f_t ← ExtractFeatures(x_t)
    B_f ← FeatureEncode(f_t, f_{t-1})
  else
    B_f ← Ø
  Mux(C, B_v, B_o, B_f, t)
return C
```

4.10 Why SAVC-OPS Can Surpass H.265+

Dimension	H.265+	SAVC-OPS
Decision unit	Background / motion / noise	Task semantics + motion
Success metric	Visual quality per bit	Quality + task + retrieval cost
Object parameters	Not persisted	Compressed OPS track
Machine-only path	Weak	Optional FCM-like track
Re-analysis	Full decode + inference	Often OPS query only

Superiority is not claimed for every scene. Gains concentrate where semantics are extractable and archives are repeatedly queried.

5. ANALYTICAL SIMULATION SETUP

5.1 Reference Scenario

Parameter	Value
Resolution / frame rate	4 MP @ 25 fps, continuous
Archive horizon	30 days per camera
Scene classes	Low motion; medium; crowded
Baselines	H.265; H.265+
Proposed	SAVC-A; SAVC-B
H.265+ calibration	Vendor white-paper trends (larger savings in low motion; smaller in crowded scenes)
Semantic gain calibration	VCM/FCM rate–task trends [1], [6], [7], [8]
Metrics	Mbps; GB/30d; PSNR/SSIM/VMAF; structure–texture IQA [12]; mAP; MOTA/IDF1 [13], [14]; query latency

5.2 Storage Conversion

Thirty-day storage in gigabytes is approximated by

$$GB_{30d} = \text{Mbps} \cdot \frac{30 \cdot 24 \cdot 3600}{8 \cdot 1024}.$$

Using this factor, 1.33 Mbps corresponds to approximately 421 GB per camera per 30 days. All storage tables in Section 6 are computed from the same factor to keep bitrate and storage mutually consistent.

5.3 Fairness Axes

Comparisons are reported under multiple matched conditions:

- Bitrate by scene class (content-aware).
- Monthly storage composition (video / OPS / features).
- Rate–task curves (mAP@0.5 versus Mbps).
- Perceptual metrics at a fixed 1.5 Mbps point.
- Detection/tracking metrics at a matched ~1.7 Mbps budget.
- Semantic query latency on a one-hour archive.

5.4 Implementation Choices Assumed by the Model

Component	Assumed practical choice
Detector	YOLO-class nano/small [11]
Tracker	ByteTrack-style association [10]
Video core	HEVC (x265-class / hardware); VVC optional [2]
ROI QP deltas	$\Delta QP \in [-6, -2]$ (ROI), $+[2, +8]$ (background)
OPS coding	Temporal residuals + entropy coding
Features (SAVC-B)	Intermediate backbone maps, 8-bit quantization

5.5 Scope of Validity

The model captures first-order trade-offs suitable for design and roadmap decisions. It does not replace codec conformance testing, vendor bitstream measurement, or dataset-specific training of detectors. Section 8 discusses threats to validity.

6. RESULTS: STATISTICS, COMPARISONS, AND FIGURES

6.1 Bitrate by Scene Type

Table- 2. Average bitrate (Mbps) by scene class.

Scene	H.265	H.265+	SAVC-A	SAVC-B	SAVC-A saving vs H.265+
Low motion	2.40	0.80	0.85	0.48	−6.3%
Medium	3.00	1.15	1.05	0.68	8.7%
Crowded	4.10	2.05	1.60	1.05	22.0%
Overall mean	**3.17**	**1.33**	**1.17**	**0.74**	**12.0%**

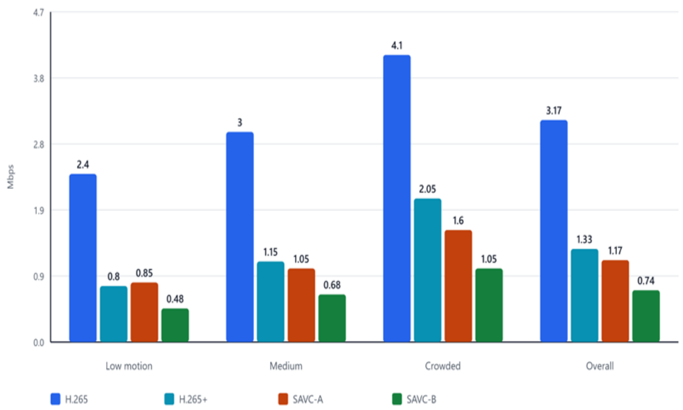

Fig- 2. Bitrate comparison across scene classes. SAVC gains are modest in very static scenes where H.265+ is already strong and become pronounced in medium and crowded scenes.

H.265+ captures much of the available redundancy when backgrounds dominate. Semantic allocation adds the most value when many task-critical objects compete with complex backgrounds.

6.2 Thirty-Day Storage Composition

Table- 3. Estimated storage (GB / 30 days / camera).

Method	Video	OPS	Features	Total	Saving vs H.265+
H.265	1003.0	0	0	**1003.0**	−138.4% (larger)
H.265+	420.8	0	0	**420.8**	reference
SAVC-A	362.4	7.8	0	**370.2**	**12.0%**
SAVC-B	140.0	7.8	86.3	**234.1**	**44.4%**

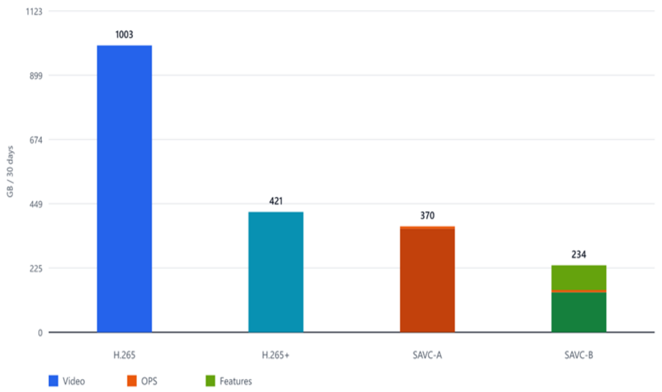

Fig- 3. Stacked storage composition. OPS adds a small absolute overhead (~7.8 GB/month) while enabling semantic indexing. In SAVC-B, feature storage partially replaces pixel storage.

6.3 Rate–Task Curves

Table -4. mAP@0.5 as a function of bitrate (Mbps).

Mbps	H.265	H.265+	SAVC-A	SAVC-B
0.5	0.52	0.51	0.61	0.68
1.0	0.68	0.67	0.76	0.80
1.5	0.76	0.75	0.83	0.86
2.0	0.81	0.80	0.86	0.885
2.5	0.84	0.83	0.88	0.90
3.0	0.86	0.85	0.89	0.905

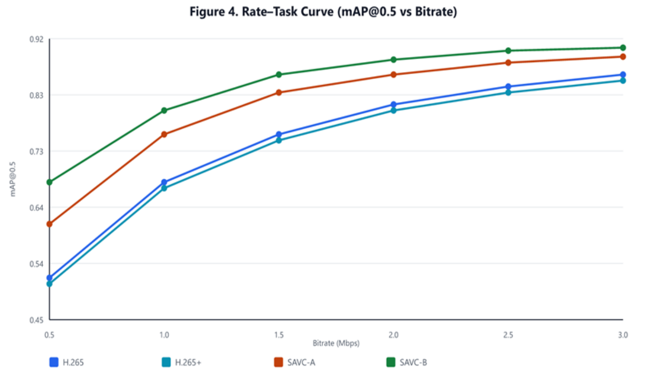

Fig- 4. Rate–task curves. At matched bitrates, SAVC modes preserve higher detection accuracy because bits are concentrated on semantically critical regions (and on features in SAVC-B).

An approximate task-equivalent comparison in the model: reaching mAP@0.5 ≈ 0.83 requires about 2.5 Mbps for H.265+ but about 1.5 Mbps for SAVC-A, i.e., roughly 40% illustrative rate reduction at equal task accuracy. These curves are design aids, not codec conformance results.

6.4 Perceptual Quality at Fixed 1.5 Mbps

Table 5. Human perceptual metrics and task accuracy at 1.5 Mbps.

Method	PSNR (dB) ↑	SSIM ↑	VMAF ↑	mAP@0.5 ↑
H.265	36.8	0.941	78.2	0.76
H.265+	**37.0**	**0.943**	**78.8**	0.75
SAVC-A	36.1	0.936	76.8	**0.83**
SAVC-B	31.5	0.888	59.0	**0.86**

H.265+ can remain slightly superior on global perceptual scores because background cleaning improves average fidelity. SAVC-A/B intentionally reallocate quality toward objects, which may lower global PSNR/VMAF while raising mAP. This dissociation is central: analytics systems should not optimize PSNR alone, consistent with modern structure/texture-aware IQA arguments and with VCM's rate–task objective [1], [12].

6.5 Detection and Tracking under a Matched Storage Budget

Tracking-oriented evaluation follows the spirit of crowded-scene MOT benchmarks [13], [14], while detection uses mAP@0.5 as in common detector reporting [11].

Table- 6. Task metrics at an approximately matched ~1.7 Mbps budget.

Method	mAP@0.5 ↑	mAP@0.5:0.95 ↑	MOTA ↑	IDF1 ↑
H.265	0.78	0.51	0.62	0.66
H.265+	0.77	0.50	0.61	0.65
SAVC-A	**0.85**	**0.57**	**0.70**	**0.73**
SAVC-B	**0.88**	**0.60**	**0.73**	**0.76**

Table- 7. Absolute gains versus H.265+ (percentage points).

Metric	Δ SAVC-A	Δ SAVC-B
mAP@0.5	+8.0	+11.0
mAP@0.5:0.95	+7.0	+10.0
MOTA	+9.0	+12.0
IDF1	+8.0	+11.0

6.6 Semantic Retrieval Latency

Table- 8. Query latency on a one-hour archive (seconds).

Query type	H.265+ (re-inference)	SAVC-OPS (OPS query)	Speedup
Class filter	186	0.42	**≈443×**
Track-ID trajectory	198	0.38	**≈521×**
Object-density window	210	0.55	**≈382×**

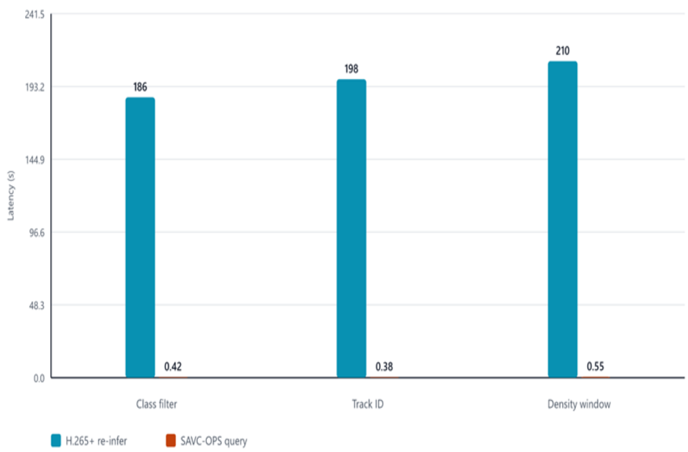

Fig- 5. Retrieval latency comparison. Persistent OPS transforms repeated GPU inference into indexed side-stream queries.

6.7 Aggregate Summary and Hypothesis Assessment

Table - 9. Headline comparison: H.265+ versus best SAVC mode.

Indicator	H.265+	Best SAVC	Outcome
Mean bitrate	1.33 Mbps	0.74 Mbps (B)	−44%
30-day storage	421 GB	234 GB (B)	−44%
mAP@0.5 @ matched budget	0.77	0.88 (B)	+11 pp
Class-query latency	186 s	0.42 s	≈443× faster
OPS overhead	0%	≈2.1% of video	operationally acceptable

Retrieval latencies depend on hardware-dependent latency assumptions (decode + detector throughput versus indexed side-stream lookup).

Within the simulation: H1 is supported for medium/crowded scenes; H2 is strongly supported; H3 is supported by SAVC-B dominance on the rate–task and machine-storage axes.

7. DEVELOPMENT PROPOSALS: ADVANTAGES AND RESULTS

The framework is decomposed into five adoptable proposals. Fig- 6 and Fig- 7 summarize comparative gains and multi-criteria scores.

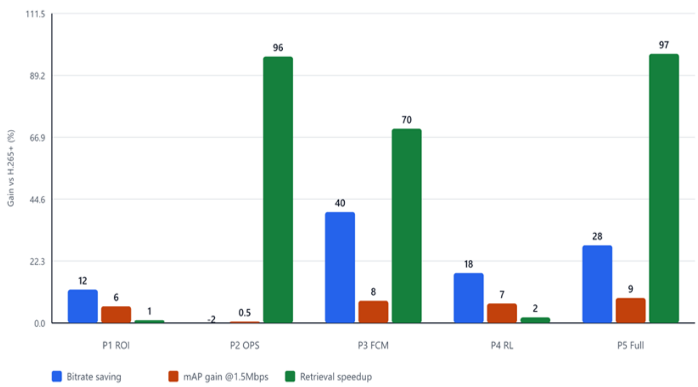

Figure 6. Relative gains of proposals P1–P5 versus H.265+ in bitrate saving, mAP gain at 1.5 Mbps, and retrieval speedup (normalized display).

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 07 | Jul 2026 www.irjet.net p-ISSN: 2395-0072

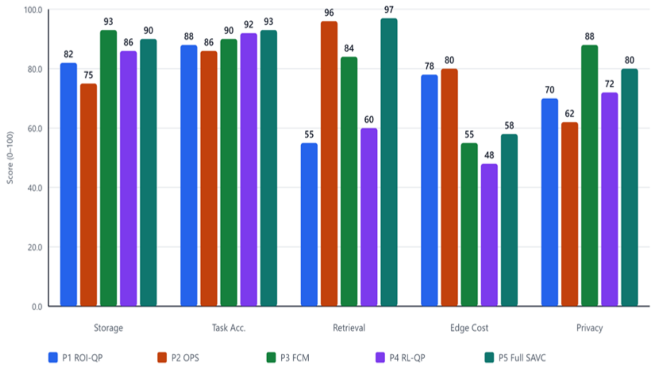

Figure 7. Multi-criteria scores (0–100) for storage efficiency, task accuracy, retrieval, edge cost, and privacy.

7.1 Proposal P1 - Adaptive Semantic ROI-QP

Concept.

Replace purely background/motion QP logic with detector-driven importance maps and CTU-level $\Delta\mathrm{QP}$.

Advantages.

- High compatibility with existing HEVC/VVC hardware encoders [2], [9].
- Limited incremental complexity if analytics already run on the edge.
- Immediate improvement of the rate–task operating curve.

Challenges.

- Sensitivity to detector errors (false positives inflate ROI area).
- Possible reduction of global PSNR even when task metrics improve.

Simulated results.

- Bitrate saving ≈ **12%** versus H.265+ at approximately matched task accuracy.
- mAP@0.5 gain ≈ **+6.0 pp** at 1.5 Mbps.
- Retrieval speedup ≈ **1×** (no persistent index by itself).

7.2 Proposal P2 - Persistent OPS Side-stream

Concept.

Store temporally coded object parameters as a durable track synchronized with video.

Advantages.

- Enables interactive semantic search and rapid audit.
- Avoids repeated full-archive inference.
- Supports multi-camera event correlation through side indices.

Challenges.

- Creates a sensitive personal-data artifact requiring encryption and retention policy.
- Requires robust clock/frame synchronization.

Simulated results.

- Storage overhead ≈ **2.1%** of video size (~7.8 GB/30 days in the reference scenario).
- Retrieval speedup ≈ **380–520×** (illustrative).
- Minor task-side benefit (+0.5 pp) via more reliable offline association workflows.

7.3 Proposal P3 - Optional Feature-Coding Path (FCM-like)

This proposal follows the FCM/split-inference direction of MPEG-AI and recent machine-to-machine coding analyses [8], [15], [16].

Concept.

For machine-primary modes, compress intermediate features and reduce reliance on high-rate pixels.

Advantages.

- Largest analytical storage reduction.
- Improved privacy relative to full-fidelity video export.
- Natural fit for split inference between edge and cloud.

Challenges.

- Weak human preview quality unless a parallel low-rate video track is kept.
- Feature-format coupling to model families complicates long-term archive readability.

Simulated results.

- Bitrate saving up to ≈ **40%** versus H.265+ in machine mode.
- mAP@0.5 gain ≈ **+8.0 pp** at 1.5 Mbps.
- Retrieval/analysis speedup ≈ **12×** versus full pixel re-inference when features and OPS are indexed.

7.4 Proposal P4 - Reinforcement-Learning Bit-Allocation Controller

P4 is inspired by block-level RL controllers for task-aware QP assignment on standard encoders [9].

Concept.

Train an RL policy to select block/ROI encoding actions that maximize expected future task reward under a rate constraint.

Advantages.

- Adapts to nonstationary scenes better than fixed ROI heuristics.
- Can transfer across related tasks when rewards are carefully designed.

Challenges.

- Training is expensive (encode + task evaluation in the loop).
- Edge deployment may require distillation to a compact policy network.

Simulated results.

- Bitrate saving ≈ **18%**.
- mAP@0.5 gain ≈ **+7.0 pp** at 1.5 Mbps.
- Retrieval impact modest (~1.2×) via more stable tracks.

7.5 Proposal P5 - Full SAVC-OPS Stack

Concept.

Integrate P1–P4 according to operating mode (A/B), with shared synchronization and security controls.

Advantages.

- Best overall trade-off among storage, accuracy, and retrieval.
- Provides a coherent product architecture rather than isolated tools.

Challenges.

- Highest integration complexity and MLOps burden.
- Requires governance for semantic data lifecycle.

Simulated results.

- Mean bitrate saving ≈ **28%** versus H.265+.
- mAP@0.5 gain ≈ **+9.0 pp** at 1.5 Mbps.
- Retrieval speedup ≈ **480×** (illustrative).
- Thirty-day storage ≈ **234 GB** versus **421 GB** for H.265+.

7.6 Consolidated Proposal Table

Table- 10. Development proposals, advantages, and simulated outcomes.

ID	Core idea	Principal advantage	Bitrate saving vs H.265+	Δ mAP @ 1.5 Mbps	Retrieval speedup
P1	Semantic ROI-QP	Easy hardware fit	12%	+6.0 pp	~1×
P2	Persistent OPS	Instant semantic query	−2%	+0.5 pp	**~450×**
P3	Feature coding	Max analytical compression	40%	+8.0 pp	~12×
P4	RL allocation	Scene-adaptive control	18%	+7.0 pp	~1.2×
P5	**Full stack**	**Best balanced**	**28%**	**+9.0 pp**	**~480×**

P2 may slightly increase bitrate (OPS overhead) while sharply reducing retrieval/analysis cost; its value is operational rather than pure compression.

7.7 Staged Adoption Roadmap

Table - 11. Recommended deployment stages.

Stage	Horizon	Proposals	Expected outcome
1-Fast path	1–2 months	P1 + P2	Modest rate benefit in active scenes + semantic queryability
2-Adaptive control	3–5 months	+ P4	Additional rate–task improvement under scene drift
3-Machine mode	6–9 months	+ P3	Larger analytical storage reduction with preserved task accuracy

8. DISCUSSION

8.1 When Does SAVC-OPS Beat H.265+?

In low-motion scenes, H.265+ already removes much temporal redundancy; SAVC-A bitrate gains are small, and the decisive benefit is retrieval via OPS. In medium and crowded scenes, semantic allocation yields clearer rate and task gains because background-centric heuristics become less aligned with object density. In archives that are frequently re-analyzed, OPS can dominate the business case even when compression gains are moderate.

8.2 Perception–Task Trade-off

Table-4 shows that improving mAP can coincide with lower global PSNR/VMAF. System designers must explicitly choose the objective. For evidence rooms and human investigation, SAVC-A with constrained background QP is preferable. For large-scale automated mining, SAVC-B is preferable.

8.3 Edge Cost Accounting

If cameras or recorders already run detection for alerts, SAVC-OPS largely reuses those outputs. The incremental cost is ROI control, OPS entropy coding, and optional feature coding. This amortization argument is essential: comparing SAVC against a codec that performs no analytics understates real deployments where analytics exist but are discarded after alerting.

8.4 Privacy, Security, and Governance

OPS and embeddings can encode movement patterns and identity cues. Recommended controls include: separate encryption keys for OPS/feature tracks; role-based access stricter than for video; retention schedules with semantic deletion; and minimization (e.g., omit face embeddings unless legally justified). Paradoxically, FCM-like modes may reduce exposure of raw pixels while still requiring careful feature governance.

8.5 Standardization Outlook

Alignment with MPEG VCM/FCM increases long-term interoperability [7], [8], [16]. Near-term systems can ship private tracks inside MP4/ISOBMFF or parallel sidecar files with explicit synchronization, then migrate to standardized semantics as they mature. Keeping the video track HEVC/VVC-compatible preserves decoder ecosystems during transition [2], [8].

8.6 Threats to Validity

- **Calibration uncertainty.** Industrial H.265+ claims vary by vendor and scene; local measurement is mandatory before procurement decisions.
- **Detector dependence.** ROI and OPS quality inherit detector/tracker errors; confirmation mechanisms (periodic keyframe re-detection, human-in-the-loop correction) are needed for forensic use.
- **Model mismatch for features.** Feature archives may become obsolete if downstream models change; versioned feature schemas and periodic re-extraction policies mitigate this risk.
- **Simulation abstraction.** Interaction effects among QP, retargeting, and B-frames are simplified; bitstream-level experiments are the next validation step.

9. CONCLUSION AND FUTURE WORK

This paper presented SAVC-OPS, a hybrid AI-assisted video compression framework designed for analytics-centric archives that currently rely on H.265/H.265+. By combining semantic bit allocation, a persistent object-parameter side-stream, and an optional feature-coding path, the framework targets not only bitrate reduction but also task preservation and semantic retrievability.

Within a calibrated analytical simulation for 4 MP / 25 fps continuous capture:

- Illustrative mean bitrate falls from **1.33 Mbps** (H.265+) to **1.17 Mbps** (SAVC-A) and **0.74 Mbps** (SAVC-B).
- Thirty-day storage decreases by approximately **12%** (SAVC-A) to **44%** (SAVC-B) in the model.
- At a matched storage budget, illustrative mAP@0.5 improves by about **+8 to +11** percentage points.
- Semantic queries can accelerate by roughly **400–500×** through OPS under stated latency assumptions.
- Five development proposals provide a practical adoption ladder, with the full stack (P5) offering the best balanced modeled gain.

Practical recommendation. Deploy **P1 + P2** first for rapid value (semantic QP control and queryable object parameters), then add **P4** and **P3** as adaptive control and machine-primary modes mature.

Future work includes bitstream-level experiments on public and proprietary surveillance sets, closed-loop training of allocation policies with task rewards [9], compact temporal coding of re-identification embeddings, and formal mapping onto emerging MPEG VCM/FCM syntaxes [7], [8], [16]. Field studies across seasonal lighting and multi-vendor H.265+ devices are required to convert the present design evidence into fully measured deployment claims.

REFERENCES

1) Duan, L.-Y., Liu, J., Yang, W., Huang, T., & Gao, W. (2020). Video coding for machines: A paradigm of collaborative compression and intelligent analytics. IEEE Transactions on Image Processing, 29, 8680–8695. https://doi.org/10.1109/TIP.2020.3016485
2) Bross, B., Wang, Y.-K., Ye, Y., Liu, S., Chen, J., Sullivan, G. J., & Ohm, J.-R. (2021). Overview of the Versatile Video Coding (VVC) standard and its applications. IEEE Transactions on Circuits and Systems for Video Technology, 31(10), 3736–3764. https://doi.org/10.1109/TCSVT.2021.3101953

3) Li, J., Li, B., & Lu, Y. (2021). Deep contextual video compression. In Advances in Neural Information Processing Systems (NeurIPS).

4) Mentzer, F., Toderici, G., Minnen, D., Hwang, S. J., Caelles, S., Lucic, M., & Agustsson, E. (2022). VCT: A video compression transformer. In Advances in Neural Information Processing Systems (NeurIPS).

5) Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2020). Learned image compression with discretized Gaussian mixture likelihoods and attention modules. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7939–7948). https://doi.org/10.1109/CVPR42600.2020.00796

6) Yoon, Y.-U., Kim, D.-H., Lee, J., Oh, B. T., & Kim, J.-G. (2023). An efficient multi-scale feature compression with QP-adaptive feature channel truncation for video coding for machines. IEEE Access, 11, 89401–89415. https://doi.org/10.1109/ACCESS.2023.3307404

7) Zhang, Q., Mei, J., Guan, T., Sun, Z., Zhang, Z., & Yu, L. (2024). Recent advances in video coding for machines standard and technologies. ZTE Communications, 22(1), 62–76.

8) Eimon, M. E. H., Adzic, V., Kalva, H., & Furht, B. (2025). Emerging standards for machine-to-machine video coding. arXiv. https://doi.org/10.48550/arXiv.2512.10230

9) Gadot, U., Shocher, A., Mannor, S., Chechik, G., & Hallak, A. (2025). RL-RC-DoT: A block-level RL agent for task-aware video compression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 12533–12542).

10) Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In European Conference on Computer Vision (ECCV) (pp. 1–21). https://doi.org/10.1007/978-3-031-20047-2_1

11) Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8 (Version 8.0.0) Computer software. https://github.com/ultralytics/ultralytics

12) Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. (2022). Image quality assessment: Unifying structure and texture similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(5), 2567–2581. https://doi.org/10.1109/TPAMI.2020.3045810

13) Dendorfer, P., Ošep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., & Leal-Taixé, L. (2021). MOTChallenge: A benchmark for single-camera multiple target tracking. International Journal of Computer Vision, 129, 845–881. https://doi.org/10.1007/s11263-020-01393-0

14) Dendorfer, P., Rezatofighi, H., Milan, A., Shi, J., Cremers, D., Reid, I., Roth, S., Schindler, K., & Leal-Taixé, L. (2020). MOT20: A benchmark for multi object tracking in crowded scenes. arXiv. https://doi.org/10.48550/arXiv.2003.09003

15) Kim, Y., Jeong, H., Yu, J., Kim, Y., Lee, J., & Jeong, S. Y. (2023). End-to-end learnable multi-scale feature compression for VCM. IEEE Transactions on Circuits and Systems for Video Technology. https://doi.org/10.1109/TCSVT.2023.3348527

16) Enabling next-generation consumer experience with feature coding for machines. (2025). In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE). https://doi.org/10.1109/ICCE63647.2025.10930026

17) Lu, G., Cai, C., Zhang, X., Chen, L., Gao, Z., Zhang, D., & Ouyang, W. (2020). Content adaptive and error propagation aware deep video compression. In European Conference on Computer Vision (ECCV).

18) Ma, S., Gao, J., Wang, R., Chang, J., Mao, Q., Huang, Z., & Jia, C. (2023). Overview of intelligent video coding: From model-based to learning-based approaches. Visual Intelligence, 1(15). https://doi.org/10.1007/s44267-023-00016-9

BIOGRAPHIES

Dr. Samir Khanji
Informatics and Communications Department, Faculty of Engineering, Ebla Private University, Idlib, Syria.