EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Spatial-Narrative-Aware Cooperative Edge Caching and Adaptive Layered Video Delivery for Industry 5.0

Huan Zhou[1], Xinlin Gao[2*]

[1]Nanchang Institute of Technology, Nanchang, China
[2]Jiangxi University of Water Resources and Electric Power, Nanchang, China

Abstract

INTRODUCTION: Industry 5.0 brings worker context, system resilience, and resource efficiency into the design of industrial digital services. Training video is a useful edge-delivery case because demand varies with where work is performed and with the type of instruction being requested, whereas conventional cache policies usually reduce demand to a single global popularity profile.
OBJECTIVES: We propose a spatial-narrative-aware delivery framework that combines zone-category-conditioned demand, layer-level cooperative caching, and adaptive Scalable Video Coding (SVC) transmission. Here, “spatial-narrative-aware” refers only to factory-zone and instructional-category metadata; it is not a claim of temporal shot segmentation or automatic narrative parsing.
METHODS: Evaluation uses 20 independent runs with fixed random seeds. Each run contains 1,500 requests, including a 300-request warm-up, across 20 edge nodes and 1,000 three-layer SVC items. All methods receive the same request sequence and per-request Rayleigh-fading realization. Experiments cover component ablation, a 10–100 MB cache-capacity sweep, θ2 sensitivity, controlled node failures, worker zone transitions, and an adaptation of a 2024 cooperative-SVC method under matched conditions. A separate public-data validation uses JSVM 9.19.15 on a 10-s construction-work clip and evaluates PPE detection on the 141-image Construction-PPE test split.
RESULTS: With 10 MB per node, Full ALT-PACC attains a 72.14% cache hit rate (95% CI: 71.60–72.67%), compared with 65.74% for ALT+LRU and 63.28% for CCo-Fog-adapted. Modeled energy is 80.62 mJ per served segment, 7.7% lower than ALT+LRU, and mean delay is 955.0 ms versus 1009.3 ms. In the auxiliary JSVM experiment, BL, BL+E1, and Full require 141.82, 319.16, and 606.34 kb/s and yield 29.57, 32.21, and 35.03 dB Y-PSNR. PPE mAP@0.5 is 0.391 on the original test images and 0.390 after base-layer coding; the paired-bootstrap retention interval includes 100%.
CONCLUSION: Layer-aware cooperation accounts for most of the cache-efficiency gain, while zone-category context and replica-diversity terms add a smaller increment under tight storage. The public-data checks show a measured SVC rate-quality progression and no detectable mAP@0.5 loss for PPE detection at the base layer in this external dataset. These results support network-side caching and safety-object semantic retention, not worker learning or task completion.

Keywords: Industry 5.0, edge caching, adaptive video streaming, spatial-narrative awareness, scalable video coding, cooperative caching Received on 04 August 2026, accepted on 04 September 2026, published on 24 September 2026

doi: 10.4108/eetsis.14316

*Corresponding author. Email: gaoxinlin850202@gmail.com

1. Introduction

Industry 5.0 extends manufacturing design beyond connectivity and automation to include worker needs, resilience, and sustainability [1,2]. Edge-enabled media access provides a relevant technical context for latency-sensitive and immersive services [3]. Evidence from construction safety training supports the use of computer-aided and video-based tools [4], while human–robot co-working and enabling-technology studies position such services within the broader Industry 5.0 agenda [5,6]. Industrial training video is therefore a practical edge application: safety procedures, maintenance demonstrations, and expert guidance often need to be retrieved near the point of work. The network problem is not simply one of maximizing average video quality. Demand changes across factory zones and instructional categories, while storage-limited edge nodes must decide which scalable layers to keep locally and when neighboring caches should be used.

Factory training libraries are typically smaller and more structured than open entertainment catalogues. Welding stations may request spark-protection material more often than assembly stations, whereas repair areas may favor fault-diagnosis sequences. We model this dependence through the requesting node's factory zone and the video's instructional category. In this paper, “spatial” means factory-floor zone and “narrative” means a reproducible instructional label such as safety, step demonstration, or scenario. Neither term refers to shot boundaries, temporal segmentation, or automatic narrative analysis.

Adaptive bitrate control and scalable video delivery are well established [7,8], as are distributed and proactive edge-caching principles [9,10]. The gap addressed here lies at their intersection: factory-zone and instructional-category context are used to condition demand, cache state is maintained per SVC layer, and delivery adapts to the instantaneous channel without solving a global placement problem online.

The study contributes a normalized node-specific demand model based on zone-category affinity, a layer-level PACC policy with explicit admission, eviction, insertion, replica management, and failure fallback, and an ALT rule that selects a feasible SVC layer from the current channel state. The network benchmark is complemented by an auxiliary public-data check: JSVM is used to measure layer rate-quality behavior, and an external PPE detector is evaluated on independently coded safety images. Worker learning, task completion, deterministic protocol failover, and hardware performance are not treated as measured outcomes.

2. Related Work

2.1. Industry 5.0 and Human-Centric Manufacturing

Industry 5.0 adds human-centeredness, resilience, and sustainability to the connectivity and automation goals associated with Industry 4.0 [1,2,5,6]. Energy-aware edge caching [11], 5G service requirements [12], and sustainability-oriented Industry 4.0 adoption [13] provide adjacent system-level context. In an industrial-training setting, human-centeredness is a design motivation for delivering context-relevant information; it cannot be inferred from cache hit rate, latency, or PSNR. Claims about learning, hazard recognition, or task completion require separate task- or user-level evidence.

2.2. Edge Caching and Content Delivery

Edge caching places frequently requested content near users to reduce backhaul demand and delivery delay [14–18], while cooperative caching has also been studied in mobile and vehicular edge settings [19]. Industrial digital-twin research links information services to factory assets and locations [20], whereas public video-workload studies such as the UMass YouTube trace characterize demand popularity rather than factory-zone semantics [21]. Video-specific edge work then moves toward adaptive bitrate caching and processing [22], multi-quality chunk caching [23], and cooperative scalable-video placement [24]. The latter two are close methodological references; our model differs by conditioning demand on factory zone and instructional category and by using a lightweight online placement policy.

2.3. Scalable Video Coding and Adaptive Transmission

SVC encodes a decodable base layer with optional enhancement layers [7], making it suitable for storage-constrained edge systems in which individual quality layers can be cached separately. Context-aware prefetching and task-oriented communication have also broadened the design space. Uriol et al. [25] used application-layer context for DASH prefetching in 5G MEC, while task-oriented video systems optimize downstream analytics rather than reconstruction quality alone [26,27]. We therefore model cache state at SVC-layer granularity and use PSNR only as a reconstruction-quality measure.

2.4. Context-Aware Industrial Video Delivery

Industrial information demand often follows the physical organization of production. Digital-twin research links information systems to factory assets and locations [20], while the UMass campus-video trace [21] is used only as an example of a non-industrial popularity marginal rather than evidence of factory semantics. That spatial organization motivates a context-conditioned cache model, not a video-segmentation task. Each training item carries an instructional-category label, and each edge node belongs to a factory zone; these metadata enter the request model. ROI or task-oriented encoders alter the visual representation itself [26,27], whereas the present method changes where SVC layers are cached. The research question is whether zone-category context improves resource use when added to layer-aware cooperative caching under a controlled industrial-training workload.

2.5. Positioning and Scope of Novelty

The novelty is limited to network-side integration rather than a new segmentation primitive. ROI and semantic-streaming methods operate on pixels, regions, or learned visual features [26,27]; layer/chunk-aware caching stores quality layers without the proposed factory context [22,23]; cooperative SVC fog caching [24] is the closest placement family but does not use the zone-category affinity term; and context-aware DASH prefetching [25] does not combine that context with cooperative SVC placement. ALT-PACC brings these three elements together—zone-category-conditioned demand, layer-level cooperative placement, and adaptive SVC delivery—and Section 5 tests their contributions with matched ablations and a recent cooperative-SVC comparator.

3. System Model

3.1. Network Architecture and Constraints

The system contains a cloud origin, a set N of edge nodes, and worker devices that request training video. Node n has cache capacity C_n and belongs to zone $z_n \in Z$. Video f has instructional category $c_f \in C$ and is represented by a base layer plus L enhancement layers. The caching logic is independent of a specific access technology. For video payloads, the intended setting is broadband industrial access such as private 5G or Wi-Fi, potentially integrated with TSN; LPWAN is not used as the video data plane. The principal system parameters and benchmark ranges are summarized in Table 1.

Table 1. Spatial-Narrative System Parameters

Parameter	**Value/Range**	**Design Rationale**
Cache Capacity (C_n)	10–100 MB	Storage-constrained benchmark range
Transmit Power (P_T)	14 dBm	Benchmark radio setting; not an LPWAN payload claim
Spatial Zone (z)	Welding/Assembly/Repair	Factory-zone metadata
Instructional Category (c)	Safety/Step/Scenario	Content-category metadata
SVC Layers (L)	3 total (base + 2 enhancement)	Layer-level caching and delivery

Table 1 lists modeling inputs rather than measured factory characteristics. Zone and instructional-category labels are metadata used by the request model, while $\gamma = 1.2$ in the primary benchmark describes the synthetic popularity generator. The radio and storage values should likewise be read as benchmark settings, not as evidence that a specific factory deployment has been profiled.

3.2. Channel and Transmission Model

The channel model follows a log-distance path-loss formulation. At carrier frequency f_c, the 1 m reference loss is $PL_0 = 32.4 + 20\log_{10}(f_c[\mathrm{GHz}])$ dB; with $f_c = 3.5$ GHz, $PL_0 = 43.3$ dB. For user–edge distance d_n and Rayleigh power gain h_n, $SNR_{n,dB} = P_{T,dBm} + G_t + G_r - PL_0 - 10\beta\log_{10}(d_n/1\ \mathrm{m}) + 10\log_{10}(h_n) - [N_0 + 10\log_{10}(B) + NF]$, where $N_0 = -174$ dBm/Hz. The achievable rate is $R_n = B \log_2(1 + 10^{(SNR_{n,dB}/10)})$. The benchmark samples h_n independently for every request.

Let $Q_f^{(0)}$ be the base-layer reconstruction quality assigned to video f, and let $\Delta Q_{f,k}$ denote the PSNR increment, in dB, associated with enhancement layer k. Expressing the increment directly in dB avoids the dimensional error in the earlier rate-distortion notation. These values are inputs to the network simulation rather than measurements from a particular encoder implementation:

$$Q_f(l) = Q_f^{(0)} + \sum_{k=1}^{l} \Delta Q_{f,k} \quad . \quad (1)$$

Here $l \in \{0,\dots,L\}$ is the highest delivered layer and $s_f(l)$ is the cumulative payload size through layer l. Equation (1) models reconstruction quality only; no PSNR threshold is interpreted as a guarantee of hazard recognition. Communication and processing energy for delivery from node n is modeled as:

$$E_{n,f,l} = (P_T/\eta_n) \cdot s_f(l)/R_n + P_c\, t_{proc} \quad . \quad (2)$$

where η_n is amplifier efficiency, P_c is codec-processing power, and t_{proc} is processing time per delivered layer. The benchmark uses $\eta_n = 0.35$, $P_c = 0.35$ W, $t_{proc} = 3$ ms per layer, cooperative-fetch energy of 0.012 J/MB, cloud-fetch energy of 0.035 J/MB, and cache-state synchronization energy of 0.00008 J per neighbor update. These are specified benchmark parameters, not device measurements.

3.3. Spatial-Narrative Caching Model

Let p_f denote normalized global popularity and $w_{z,c} \geq 0$ the affinity between factory zone z and instructional category c. The request probability at node n is $p_{n,f} = p_f\, w_{(z_n,c_f)} / \Sigma_j\, p_j\, w_{(z_n,c_j)}$, which avoids treating spatial and narrative terms as independent marginals. In the benchmark, $p_{n,f}$ is computed before each run and held fixed; no online EMA update is used. A deployment could re-estimate the table between cache-update epochs, but that adaptation is not evaluated here. Cache state is defined per SVC layer: $x_{n,f,l} = 1$ when layer l of video f is stored at node n.

$$\Sigma_f\, \Sigma_l\, x_{n,f,l}\, s_{f,l} \leq C_n, \quad x_{n,f,l} \leq x_{n,f,l-1} \text{ for } l \geq 1. \qquad (3)$$

The capacity constraint limits total cached layer size, while the dependency constraint prevents an enhancement layer from being retained without its lower layers. Designated safety items may receive a base-layer replica target $r_f \geq 2$, subject to capacity. PACC is an online heuristic; no global combinatorial optimum is claimed.

3.4. Optimization Target and Decomposition

The design target is represented by $J = \lambda_E\, \bar{E} + \lambda_D\, \bar{D} - \lambda_H\, H - \lambda_Q\, \bar{Q}$, where $\bar{E}$ and $\bar{D}$ are mean energy and delay, H is cache hit rate, and $\bar{Q}$ is delivered-layer quality utility. The non-negative weights express the intended trade-off. Feasible solutions must satisfy cache capacity, SVC dependency, transmission-rate constraints, and any configured base-layer replica target. No numerical λ weights enter the benchmark or the online policy decisions; J states the multi-criteria design target rather than a scalar objective solved in the experiments. The constituent metrics are therefore reported separately, and no weighted global optimum is claimed.

PACC and ALT operate on different timescales. Cache placement changes when content is admitted or evicted, whereas ALT chooses a layer for each transmission from the currently available set and instantaneous channel state. The sequential procedure is therefore an implementation-oriented decomposition: PACC determines the source and available layers, then ALT chooses the highest feasible layer. Global optimality is not assumed.

4. Proposed ALT-PACC Scheme

ALT-PACC is implemented as two sequential decisions. PACC maintains the layer-level cache state and identifies a local, neighboring, or cloud source for the requested content. ALT then selects the highest layer that can meet the current rate and delay constraints. This design keeps per-request processing simple and allows the contribution of caching and transmission control to be evaluated separately.

4.1. Adaptive Layered Transmission (ALT)

ALT selects the highest SVC layer whose cumulative bitrate and estimated delivery time fit the current channel and segment-delay budget. The benchmark uses 2 s segments with incremental bitrates of 0.8/1.2/2.0 Mbps (cumulative 0.8/2.0/4.0 Mbps) and layer sizes of 0.20/0.30/0.50 MB. With B = 1 MHz and $\theta_l = 2^{(b_f(l)/B)} - 1$, the corresponding SNR thresholds are −1.30 dB, 4.77 dB, and 11.76 dB. Section 5.4 varies a ±3 dB implementation margin around the full-layer threshold.

Algorithm 1: Adaptive Layered Transmission (ALT)

Input: SNR_n, available cached layers A_f, cumulative layer bitrates $b_f(l)$, bandwidth B, delay budget D_{max}

Output: selected highest layer l* or OUTAGE

1: compute $R_n = B \log_2(1 + SNR_n)$
2: for l from $\max(A_f)$ down to 0 do
3: estimate $D_{n,f}(l) = s_f(l)/R_n + D_{source}$
4: if $b_f(l) \leq R_n$ and $D_{n,f}(l) \leq D_{max}$ then
5: return l
6: end if
7: end for
8: return OUTAGE

The source-delay term D_{source} distinguishes local, cooperative, and cloud retrieval. For the three-layer representation used in the benchmark, the search is constant in size; with a general L-layer stream, ALT requires O(L) comparisons. If even the base layer cannot satisfy the rate and delay conditions, the request is recorded as an outage rather than being labeled semantically successful.

4.2. Popularity-Aware Cooperative Caching (PACC)

PACC uses node-specific demand and layer utility to decide admission and replacement. A local lookup is performed before any eviction; on a miss, eligible neighbors are checked for a dependency-complete layer chain, otherwise the request falls back to the cloud. Caches start empty at the beginning of each benchmark run (the CCo-Fog-adapted comparator is the only prefilled policy), and the demand table $p_{n,f}$ is fixed for that run. For reproducibility, equal eviction priorities are resolved by deterministic cache insertion order under the fixed seeded request sequence; neighbor-source ties are resolved by shorter distance, higher available layer, and then node index.

Admission is ranked by a bounded utility-per-byte score that combines conditional demand, layer utility, replica diversity, and any deficit relative to the replica target:

$$\varphi_{n,f,l} = (p_{n,f}\, u_l / s_{f,l})\, [1 + \mu\, \bar{d}_{n,f} + \nu\, \delta_{n,f}], \quad \mu,\nu \geq 0. \qquad (4)$$

Here u_l is layer utility, $\bar{d}_{n,f} \in [0,1]$ is the normalized distance to the nearest existing replica measured from the pre-admission cache snapshot, and $\delta_{n,f} \in [0,1]$ is the normalized replica deficit. Capacity is enforced by Equation (3), so the priority contains no singular capacity multiplier. A duplicate is admitted only when it improves the configured diversity or replica target.

Algorithm 2: Layer-Aware Popularity-Aware Cooperative Caching (PACC)

Input: request (n,f), requested layer l_{req}, local cache state, neighbor metadata, current popularity estimates

Output: delivery source, available layer set, updated cache state

1: read $p_{n,f}$ from the current demand table (fixed within a benchmark run)

2: if required layers $0\ldots l_{req}$ are in local cache then

3: source ← LOCAL; A_f ← local layers

4: else query K eligible neighbors for required layers

5: if a neighbor supplies the dependency-complete layer set then

6: source ← minimum-cost feasible neighbor; A_f ← returned layers

7: else source ← CLOUD; A_f ← origin layers

8: end if

9: end if

10: fetch/serve from source; apply ALT to choose delivered layer

11: for each delivered layer candidate in dependency order do

12: compute $\varphi_{n,f,l}$ from the pre-admission cache snapshot

13: while free capacity < $s_{f,l}$ and an evictable item exists do

14: evict the lowest-priority unprotected cached layer; preserve SVC dependency

15: end while

16: if sufficient capacity and admission criterion is met then insert layer l

17: end for

18: update local priority state; synchronize changed metadata with K neighbors

19: on node failure, invalidate its metadata and repeat neighbor/cloud lookup

20: return source, A_f, updated cache state

4.3. Complexity Analysis

Complexity is reported per event. ALT scans at most L candidate layers, $O(L)$, which is constant for the three-layer benchmark. Hash-based local lookup is $O(L)$, probing K neighbors is $O(KL)$, and admission/eviction with a min-priority structure is $O(L \log M_n)$ amortized for M_n evictable cached layers. Synchronizing the changed state of one object is $O(KL)$. An optional full refresh of all node-file-layer scores is $O(NFL)$, but that is not a per-request cost. Table 2 summarizes these per-event complexity bounds.

No ARM runtime, memory footprint, cache-synchronization traffic, SVC extraction timing, or device-power measurements are reported. Earlier numerical estimates for these quantities have therefore been removed. Section 5.6 reports only the wall-clock encoding and full-stream decoding time observed in the x86_64 reference-software run. The present paper claims simulation-level complexity and modeled energy results only; edge-hardware feasibility requires a separate hardware or emulation study.

Table 2. Per-Event Computational Complexity of ALT-PACC Components

Operation	Complexity	Scope
ALT layer selection	$O(L)$	Per transmission; $O(1)$ when $L=3$
Local layer lookup	$O(L)$	Per request with hash-indexed layer keys
Neighbor cooperative probe	$O(KL)$	Per local miss
Admission/eviction	$O(L \log M_n)$	Per admitted request with priority heap

4.4. Integration with Deterministic Networking

ALT-PACC can operate over industrial broadband networks that support traffic differentiation, including private 5G and TSN-integrated deployments [12,28]. A practical mapping could give the base layer higher transport priority than enhancement layers. This is a deployment option rather than part of the simulator; TSN scheduling and deterministic 5G behavior are not modeled here.

The paper therefore makes no fixed 10 ms control-latency or 15 ms failover claim. Demonstrating deterministic transport would require an explicit protocol stack, wired/wireless boundary, traffic classes, topology, scheduling policy, and recovery mechanism. TSN–5G studies show that bounded end-to-end behavior depends on the details of interworking and QoS translation [28].

Deterministic networking is treated only as a possible transport layer for a future deployment. Cooperative caching can provide alternate content sources, but end-to-end availability and recovery bounds must be measured under the chosen network stack rather than inferred from cache hit rate.

5. Simulation Results and Reproducible Evaluation

All results in this section come from the benchmark implementation of Sections 3–4. Each condition is repeated with the same set of 20 fixed random seeds; run-level observations are retained before aggregation, and figures report means with 95% confidence intervals.

5.1. Benchmark Configuration and Repeated-Run Protocol

The benchmark contains $N = 20$ edge nodes distributed across welding, assembly, and repair zones in a 500 × 500 m area and $F = 1{,}000$ training items. Node coordinates are regenerated deterministically from each seed; edges are formed within a 190 m cooperation radius, when the 190 m radius graph would otherwise give a node fewer than two neighbors, the three nearest other nodes are used as its fallback neighbor set. Each run contains 1,500 requests, of which 300 warm the caches and 1,200 enter the reported metrics. Request probabilities follow the normalized zone-category model in Section 3.3 with $\gamma = 1.2$. The primary ablation uses 10 MB per node, while the capacity sweep spans 10–100 MB.

Each request samples an independent Rayleigh power gain $h \sim \mathrm{Exp}(1)$. The radio settings are $B = 1$ MHz, $f_c = 3.5$ GHz, $P_T = 14$ dBm, 0 dBi antenna gains, $PL_0 = 43.3$ dB at 1 m, $N_0 = -174$ dBm/Hz, NF = 7 dB, $\beta = 3.2$, and a user–edge distance drawn uniformly from 5 to 100 m. The segment-delay budget is 2.5 s. Source delay is 2 ms for local access, 12 ms plus distance-dependent overhead for cooperative retrieval, and 45 ms for cloud retrieval. These settings represent a broadband industrial edge abstraction, not LPWAN video transport.

No explicit queueing process is modeled. Reported delay combines the configured source-access term and transmission time under the sampled channel realization; the latency results should therefore not be interpreted as queueing-calibrated or deterministic-network measurements.

The network benchmark uses a three-layer SVC abstraction with 2 s segments, incremental bitrates of 0.8/1.2/2.0 Mbps, layer sizes of 0.20/0.30/0.50 MB, a 31 dB base-quality input, and +3/+2 dB enhancement increments. These values remain benchmark inputs rather than measurements from a particular clip. Section 5.6 reports a separate JSVM 9.19.15 encoding experiment to check whether the assumed cumulative-rate shares and quality increments are plausible; those measurements are not used to retune the network results. The complete repeated-run benchmark settings are summarized in Table 3.

Table 3. Reproducible benchmark configuration (20 independent runs)

Parameter	**Value**	**Reproducibility note**
Nodes / catalog	20 / 1,000	3 factory zones; closed catalog
Requests	1,500/run	300 warm-up + 1,200 measured
Primary cache	10 MB/node	capacity sweep: 10,30,50,70,100 MB
SVC bitrates	0.8/1.2/2.0 Mbps	incremental; cumulative 0.8/2.0/4.0
Layer sizes	0.20/0.30/0.50 MB	2 s segment abstraction
Quality input	31 +3 +2 dB	base PSNR + enhancement increments
Bandwidth / carrier	1 MHz / 3.5 GHz	broadband edge abstraction
Radio / path loss	14 dBm; PL_0=43.3 dB; N_0=−174 dBm/Hz; NF=7 dB; β=3.2	3.5 GHz; 0 dBi gains; log-distance model
Fading	Rayleigh per request	h~Exp(1)
Delay budget	2.5 s/segment	local 2 ms; coop 12 ms+; cloud 45 ms
Energy inputs	η=0.35; Pc=0.35 W	coop 0.012 J/MB; cloud 0.035 J/MB
Neighbor graph	190 m radius; sparse fallback	seeded coordinates; if <2 radius neighbors, use the three nearest nodes
PACC utility / replicas	μ=0.20; ν=0.25; r_f=2	r_f=2 for top 5% safety items; otherwise r_f=1

5.2. Ablation and Recent Direct-Competitor Baseline

The ablation separates transmission control from cache policy. LRU+fixed-full and PACC+fixed-full require the full layer and serve as diagnostic controls for the effect of ALT; ALT+LRU gives the non-PACC cache the same adaptive delivery rule used by the proposed method. Global-PACC, spatial-only PACC, narrative-only PACC, and spatial+narrative PACC then isolate the cache-side context terms, while Full ALT-PACC adds the diversity and replica-deficit terms in Equation (4). Because the fixed-full controls intentionally reject requests that cannot sustain the full layer, they are reported as diagnostics rather than as preferred operating policies.

We also implement CCo-Fog-adapted from the 2024 cooperative-SVC family proposed by Sharifi et al. [24]. The published method assumes multiple content providers and fog tiers, so an exact drop-in implementation is not possible in this single-provider, single-edge-tier topology.

The adaptation keeps cluster-based cooperation, popularity-aware selection, and layer-aware placement while using the same catalog, cache budget, requests, channel realizations, and accounting model as ALT-PACC. Supplementary Package S1 documents each adaptation.

At 10 MB per node, Full ALT-PACC reaches a 72.14% hit rate (95% CI: 71.60–72.67%), compared with 65.74% for ALT+LRU and 63.28% for CCo-Fog-adapted. Relative to ALT+LRU, the difference is 6.40 percentage points; modeled energy falls from 87.36 to 80.62 mJ per served segment and mean delay from 1009.3 to 955.0 ms. Against CCo-Fog-adapted, the hit-rate difference is 8.86 points and modeled energy is 19.1% lower. Mean PSNR is slightly lower than both ALT+LRU (35.18 dB) and CCo-Fog-adapted (35.67 dB), so the comparison shows a resource-efficiency/quality trade-off rather than uniform dominance. The ablation and matched-comparator means are reported in Table 4.

The modeled energy difference can also be traced to its accounting components. Across the same 20 seeds, ALT+LRU averages 70.99 mJ radio, 2.76 mJ codec, 13.15 mJ cooperative/cloud fetch, and 0.46 mJ synchronization energy per served segment; Full ALT-PACC averages 67.36, 2.65, 10.15, and 0.46 mJ, respectively. The 6.74 mJ total reduction is therefore driven mainly by lower radio and remote-fetch terms, not by the synchronization term. These values are outputs of the benchmark accounting model rather than device power measurements. Table 5 gives the corresponding component-level energy means, and Figure 1 visualizes the adaptive-delivery hit rates with 95% confidence intervals.

Table 4. Ablation and direct-competitor results at 10 MB/node (means; 95% CIs shown in Figure 1)

Variant	**Hit %**	**Energy mJ**	**Delay ms**
LRU+fixed-full	56.00	95.7	1100
PACC+fixed-full	51.61	94.0	1098
ALT+LRU	65.74	87.4	1009
Global-PACC+ALT	71.66	80.9	957
Spatial-only PACC	71.40	81.0	957
Narrative-only PACC	71.65	80.9	957
Spatial+Narrative PACC	71.86	80.9	957
Full ALT-PACC	72.14	80.6	955
CCo-Fog-adapted	63.28	99.6	1129

Table 5. Modeled energy-component decomposition at 10 MB/node (means over 20 runs; mJ per served segment)

Component	**ALT+LRU**	**Full ALT-PACC**
Radio	70.99	67.36
Codec	2.76	2.65
Remote fetch	13.15	10.15
Sync	0.46	0.46
Total	**87.36**	**80.62**

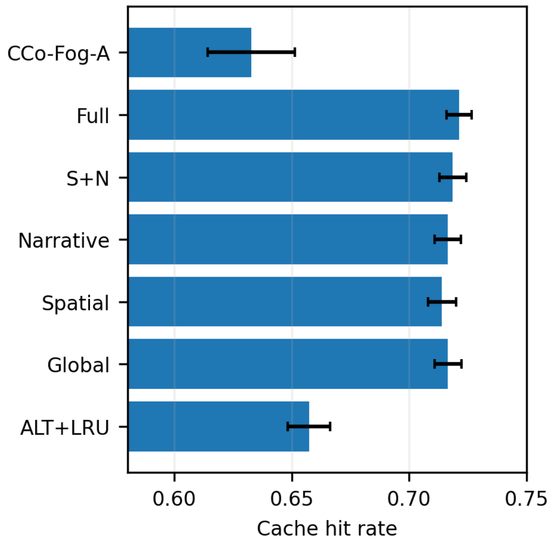

Figure 1. Hit rate for the adaptive-delivery ablation variants at 10 MB/node; error bars are 95% confidence intervals. Fixed-full diagnostic controls are reported in Table 4

5.3. Cache-Capacity Sensitivity

The capacity sweep reuses the same 20 seeds at 10, 30, 50, 70, and 100 MB per node. Policy differences are largest at 10 MB and narrow as storage grows; by 50 MB several online policies approach one another because a larger fraction of the active working set fits in cache. The gain is therefore concentrated in the storage-constrained regime for which PACC is intended. Figure 2 shows the cache-capacity sensitivity across the five tested capacities.

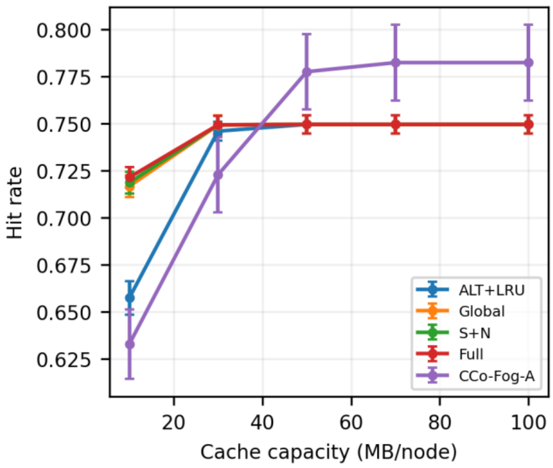

Figure 2. Cache-capacity sensitivity; means over 20 seeds with 95% confidence intervals

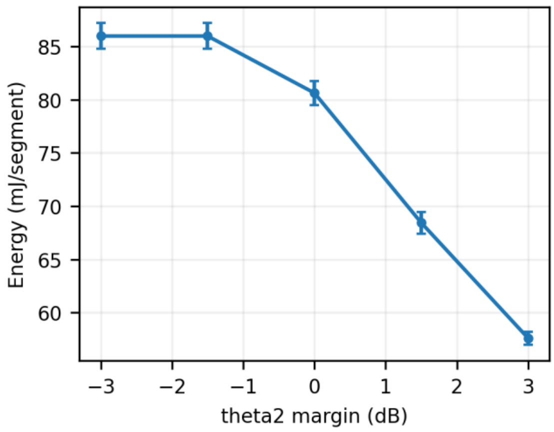

Figure 3. Modeled energy sensitivity to the full-layer threshold margin; 95% confidence intervals

5.4. SVC Threshold and θ2 Sensitivity

For B = 1 MHz and cumulative bitrates of 0.8, 2.0, and 4.0 Mbps, the Shannon-rate thresholds are −1.30, 4.77, and 11.76 dB. We vary the implementation margin around the full-layer threshold θ2 from −3 to +3 dB while holding requests, cache capacity, and channel seeds fixed.

A larger θ2 margin makes full-layer delivery more conservative. At the nominal threshold, mean PSNR is 34.94 dB, modeled energy is 80.62 mJ/segment, and delay is 955.0 ms. At +3 dB, PSNR falls to 34.03 dB, while modeled energy and delay fall to 57.60 mJ and 679.6 ms. These values describe a quality-resource trade-off; no threshold is labeled globally optimal. Table 6 reports the numerical θ2-margin results, and Figure 3 shows the associated modeled-energy trend.

Table 6. θ2 margin sensitivity

Margin dB	PSNR dB	Energy mJ	Delay ms
-3.0	35.10	86.0	1024
-1.5	35.10	86.0	1024
+0.0	34.94	80.6	955
+1.5	34.49	68.4	804
+3.0	34.03	57.6	680

5.5. Controlled Node Failures and Worker Zone Transitions

Failures are injected after the warm-up period. The scenarios include one random failed node, 10% and 20% random failures, a 10% high-connectivity-node failure, and two adjacent failures. If the request's local node is unavailable, the request is rerouted to an eligible neighbor when possible and otherwise to the cloud. Service availability is the fraction of requests that deliver at least the base layer within the 2.5 s segment budget; recovery latency is the additional source-discovery and rerouting delay.

With 20% random node failures, Full ALT-PACC retains a 70.60% cache hit rate, compared with 62.58% for ALT+LRU and 51.48% for CCo-Fog-adapted. Service availability is 98.89% for all three methods within rounding because cloud fallback serves most cache misses caused by failed nodes. The near-identical availability values show why cooperative hit rate should not be read as end-to-end resilience: cooperation improves cache-level accessibility and can reduce fallback, but deterministic availability is not demonstrated. The modeled added recovery/source-discovery delay is about 16.0 ms. The five controlled failure conditions are summarized in Table 7, and Figure 4 visualizes their cache-hit results.

Table 7. Failure-injection results (20 runs per condition)

Scenario	Full hit%	LRU hit%	Avail. %	Recov. ms
1 node	71.6	64.7	99.13	15.7
10% random	71.4	64.4	99.13	15.7
20% random	70.6	62.6	98.89	16.0
10% hotspot	71.0	63.8	99.14	15.9
2 adjacent	71.2	64.2	99.14	15.7

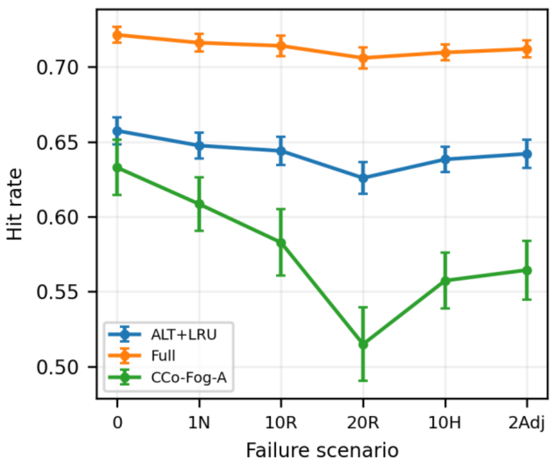

Figure 4. Cache hit rate under controlled node failures; error bars are 95% confidence intervals

The mobility workload changes the active worker zone every 500 post-warm-up requests in a repeated Welding→Assembly→Repair sequence. Full ALT-PACC reaches a 71.90% hit rate, compared with 71.33% for Global-PACC and 65.10% for ALT+LRU. The 0.57-point gain over Global-PACC is modest, which is consistent with the role of zone-category context as an incremental signal rather than the main source of the cooperative-caching gain. Figure 5 shows the worker zone-transition comparison with 95% confidence intervals.

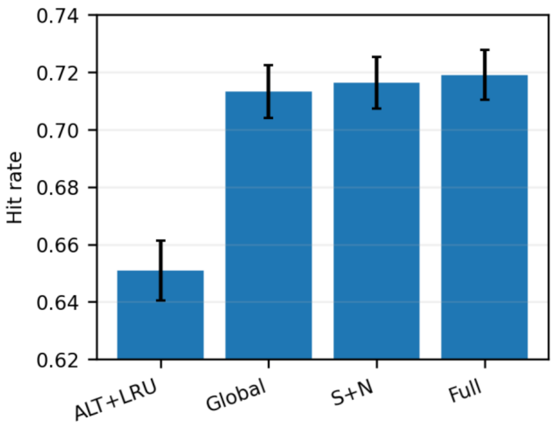

Figure 5. Worker zone-transition workload; 95% confidence intervals over 20 seeds

5.6. Codec-Level JSVM Validation

A codec-level check was run with JSVM 9.19.15, the H.264/AVC SVC reference software [29]. The source is a 12-s public construction-work video released under CC BY-SA 4.0 [30]; the first 10 s were converted from 2560×1440/30 fps VP9 to 256×144/24 fps YUV420p. The three same-resolution CGS layers used QP 36/32/28, GOPSize=8, IntraPeriod=32, one reference frame, BaseLayerMode=1, CgsSnrRefinement=0, SearchMode=4, SearchRange=4, and adaptive inter-layer prediction for the enhancement layers. LowComplexityMbMode=1 was used to keep reference-software execution tractable. The experiment contains 240 coded frames.

Measured cumulative rates are 141.82 kb/s for BL, 319.16 kb/s for BL+E1, and 606.34 kb/s for Full. The corresponding Y-PSNR values are 29.57, 32.21, and 35.03 dB, with Y-SSIM of 0.845, 0.899, and 0.939. Thus the cumulative-rate shares are 23.4%, 52.6%, and 100%, close to the 20%, 50%, and 100% shares used by the network abstraction. The measured enhancement gains are +2.64 and +2.81 dB. This single-clip experiment supports the scale of the abstraction but is not presented as a universal codec calibration. Table 8 reports the measured rate–distortion values, and Figure 6 plots cumulative bitrate against Y-PSNR.

The same auxiliary run recorded 15.23 s wall-clock encoding time for the 240-frame sequence and 0.71 s to decode the full scalable bitstream in the x86_64 evaluation container. These timings document reference-software execution only and are not used as ARM, gateway, or real-time deployment benchmarks.

Table 8. Measured JSVM rate-distortion results on the public construction clip

Representation	Cum. rate	Y-PSNR (95% CI)	Y-SSIM (95% CI)
BL	141.82 kb/s	29.57 (29.52–29.62)	0.845 (0.844–0.846)
BL+E1	319.16 kb/s	32.21 (32.17–32.26)	0.899 (0.899–0.900)
Full	606.34 kb/s	35.03 (34.98–35.08)	0.939 (0.939–0.940)

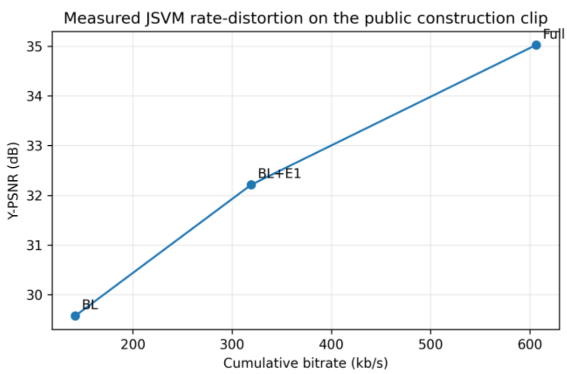

Figure 6. Measured cumulative bitrate and Y-PSNR for the three JSVM layer representations

5.7. Task-Oriented PPE Semantic Validation

Task-level semantic retention was evaluated on the official 141-image test split of Construction-PPE [31]. Each annotated image was letterboxed to 256×144 and coded as an independent JSVM access unit (GOPSize=1, IntraPeriod=1) with the same QP 36/32/28 layer structure; the all-intra setting avoids imposing artificial temporal relations on still-image annotations. SafetyVision YOLOv8 v2 [32] served as an external detector. Evaluation was restricted to eight semantically aligned classes shared by the dataset and detector: gloves, goggles, helmet, no_gloves, no_goggles, no_helmet, person, and vest. The detector was not trained or fine-tuned on the test images in this study.

Original images yield mAP@0.5=0.391 (95% bootstrap CI: 0.349–0.450). Confidence intervals use 500 paired image-bootstrap resamples with fixed seed 20260825. BL yields 0.390 (0.345–0.429), BL+E1 0.369 (0.340–0.407), and Full 0.403 (0.354–0.462). Relative to the original condition, the mAP@0.5 point estimates are 99.9%, 94.5%, and 103.2%; all paired-bootstrap retention intervals include 100%. Macro F1 at a 0.25 confidence threshold is 0.241, 0.198, 0.225, and 0.246, respectively. The non-monotonic detector response is itself informative: higher PSNR does not translate mechanically into a higher task score. Under this external test, the base layer does not show a detectable mAP@0.5 loss, but the result should not be generalized to worker hazard perception or learning. Table 9 summarizes the four PPE-detection conditions, and Figure 7 shows mAP@0.5 with image-bootstrap confidence intervals.

Table 9. PPE detection under original and JSVM-coded conditions (141 test images; eight shared classes)

Condition	PSNR / SSIM	mAP@0.5 (95% CI)	mAP@0.5:0.95	Retention
Original	–	0.391 (0.349–0.450)	0.194	100.0%
BL	33.55 dB / 0.937	0.390 (0.345–0.429)	0.201	99.9%
BL+E1	36.33 dB / 0.963	0.369 (0.340–0.407)	0.192	94.5%
Full	39.54 dB / 0.979	0.403 (0.354–0.462)	0.210	103.2%

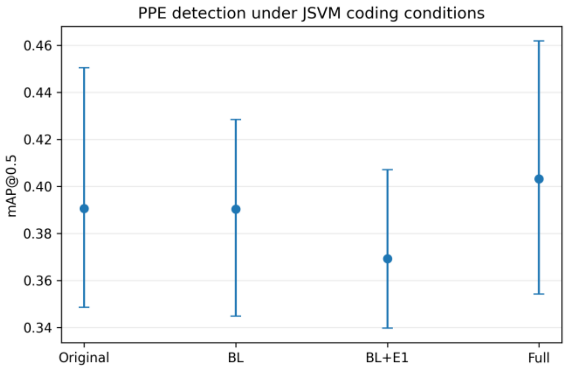

Figure 7. PPE mAP@0.5 under original and JSVM-coded conditions; error bars are 95% image-bootstrap confidence intervals

5.8. Statistical Reproducibility and External-Trace Scope

Primary network results are stored one row per run before aggregation. Supplementary Package S1 contains the simulator, benchmark configuration, seed list, raw run-level CSV files, summary tables, figure scripts, the CCo-Fog adaptation note, and the UMass preprocessing script. The auxiliary-validation folder adds the JSVM configuration files, source hashes, codec logs, frame-level PSNR/SSIM tables, PPE class-alignment record, detector predictions, task-metric tables, and scripts used to reproduce Sections 5.6–5.7. Large public source files and the third-party detector weights are referenced by URL and SHA256 rather than redistributed.

The raw UMass trace is not redistributed in S1. The present evaluation therefore reports no UMass-derived performance values; the supplied parser is included so that the public trace can be processed independently if an external popularity stress test is desired. Such a test would address only robustness to a non-industrial popularity marginal and cannot validate factory-zone labels, instructional categories, worker mobility, or industrial video semantics.

6. Discussion

The ablation makes the source of the improvement clearer. At 10 MB per node, Full ALT-PACC is 6.40 percentage points above ALT+LRU in hit rate and reduces modeled energy by 7.7% and mean delay by 54.4 ms. The difference from Global-PACC is only 0.48 points. Most of the gain therefore comes from layer-aware cooperation, while zone-category conditioning and the diversity/redundancy terms provide a smaller increment under tight storage.

The failure experiments also narrow the resilience claim. Full ALT-PACC maintains a higher cache hit rate than ALT+LRU as nodes fail, yet end-to-end service availability remains almost unchanged across methods because cloud fallback handles most disrupted requests. Cooperative caching is therefore best described as improving cache-level redundancy and reducing fallback dependence. The reported recovery delay belongs to the simulator's rerouting model and is not a measured TSN/5G failover bound.

The intended data plane remains private 5G or Wi-Fi, with TSN as a possible integration path; LPWAN is not used for video payloads. The JSVM check adds measured codec evidence without turning one public clip into a deployment calibration: its 23.4%/52.6%/100% cumulative-rate shares are close to the network model's 20%/50%/100% split, while its quality increments are of similar magnitude. The PPE experiment also changes the semantic claim. PSNR is still not a proxy for worker performance, but safety-object detection can now be examined directly. The coded conditions do not show a monotonic mAP response, and their retention intervals overlap 100%, so the evidence supports semantic robustness in this external test rather than a guarantee of hazard recognition.

7. Conclusion

Industrial training-video delivery is modeled here as a spatial-narrative-aware, layer-level cooperative caching problem rather than a video-segmentation task. The method combines a normalized zone-category demand model, SVC-layer placement with dependency constraints, a bounded diversity/redundancy priority, and channel-adaptive ALT delivery. Evaluation uses 20 fixed random seeds with per-request Rayleigh fading and matched request/channel conditions across policies.

At 10 MB per node, Full ALT-PACC reaches a 72.14% hit rate, versus 65.74% for ALT+LRU and 63.28% for CCo-Fog-adapted. It uses 7.7% less modeled energy than ALT+LRU and reduces mean delay by 54.4 ms, with a small PSNR trade-off. Under 20% random node failures, its cache hit rate remains 70.60%, while service availability stays near 99% across methods because of cloud fallback. The resilience claim is therefore limited to cache-level accessibility.

The combined evidence supports the network-caching and layered-delivery contribution under the stated simulation assumptions and adds two external checks: measured H.264/SVC rate-quality behavior on a public construction clip and PPE-detection retention on a public safety dataset. The latter is a machine-vision task metric, not a measure of worker learning, human hazard perception, or task completion. Deterministic TSN/5G recovery and measured edge-hardware feasibility likewise remain outside the evidence claimed here.

Data Availability Statement

Supplementary Package S1 contains all study-specific code, configurations, seeds, derived data, detector predictions, and figure-source tables. The public construction video is available from Wikimedia Commons [30], Construction-PPE from Ultralytics [31], JSVM from Fraunhofer HHI [29], and the SafetyVision detector from Hugging Face [32]. The model SHA256 used in this study is ea18ae903a566e8fa76f3ee1c503075522dca269269315e9c862efa170430b35. Raw UMass traces are not redistributed and remain limited to the optional popularity-marginal preprocessing path described in Section 5.8.

Supplementary Package S1 contains the network simulator, request generator, benchmark configuration, 20 fixed random seeds, run-level outputs, figure-source tables, baseline-adaptation note, UMass preprocessing script, and the scripts/configuration/results for the JSVM and PPE auxiliary validations. Public video, Construction-PPE images, JSVM source, and SafetyVision weights are not redistributed; S1 records their source locations and integrity hashes where applicable.

References

[1] Nahavandi S. Industry 5.0 - A human-centric solution. Sustainability. 2019;11(16):4371.

[2] Longo F, Padovano A, Umbrello S. Value-oriented and ethical technology engineering in industry 5.0: A human-centric perspective for the design of the factory of the future. Appl Sci. 2020;10(12):4182.

[3] Sukhmani S, Sadeghi M, Erol-Kantarci M, El Saddik A. Edge caching and computing in 5G for mobile AR/VR and tactile internet. IEEE Multimedia. 2019;26(1):21-30.

[4] Gao Y, Gonzalez VA, Yiu TW. The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. Comput Educ. 2019;90:1-14.

[5] Demir KA, Doven G, Sezen B. Industry 5.0 and human-robot co-working. Procedia Comput Sci. 2019;158:688-695.

[6] Maddikunta PKR, Pham QV, Prabadevi B, et al. Industry 5.0: A survey on enabling technologies and potential applications. J Ind Inform Integr. 2022;26:100257.

[7] Schwarz H, Marpe D, Wiegand T. Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans Circuits Syst Video Technol. 2007;17(9):1103-1120.

[8] Sun Y, Yin X, Jiang J, et al. CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction. In: Proc ACM SIGCOMM. 2016:272-285.

[9] Shanmugam K, Golrezaei N, Dimakis AG, et al. Femtocaching: Wireless content delivery through distributed caching helpers. IEEE Trans Inf Theory. 2013;59(12):8402-8413.

[10] Bastug S, Bennis M, Debbah M. Living on the edge: The role of proactive caching in 5G wireless networks. IEEE Commun Mag. 2014;52(8):82-89.

[11] Xu J, Ota K, Dong M. Energy efficient hybrid edge caching scheme for tactile internet in 5G. IEEE Trans Green Commun Netw. 2019;3(2):483-493.

[12] 3GPP. Service requirements for the 5G system. 3GPP TS 22.261, V18.0.0; 2022.

[13] Javaid M, Haleem A, Singh RP, et al. Understanding the adoption of Industry 4.0 technologies in improving

environmental sustainability. Sustain Oper Comput. 2022;3:203-217.
[14] Taleb T, Samdanis K, Mada B, et al. On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration. IEEE Commun Surveys Tutorials. 2017;19(3):1657-1681.
[15] Liu H, Chen Z, Tian X, et al. On content-centric wireless delivery networks. IEEE Wireless Commun. 2014;21(6):118-125.
[16] ElBamby MS, Bennis M, Saad W. Proactive edge computing in latency-constrained fog networks. In: Proc EuCNC. 2017:1-6.
[17] Chen B, Liu L, Sun M, Ma H. IoTCache: Toward data-driven network caching for Internet of Things. IEEE Internet Things J. 2019;6(6):10064-10076.
[18] Qiao G, Leng S, Maharjan S, et al. Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks. IEEE Internet Things J. 2020;7(1):247-257.
[19] Chaudhari B, Zennaro M, Borkar S. LPWAN technologies: Emerging application characteristics, requirements, and design considerations. Future Internet. 2020;12(3):46.
[20] Tao F, Zhang H, Liu A, Nee AYC. Digital twin in industry: State-of-the-art. IEEE Trans Ind Inform. 2019;15(4):2405-2415.
[21] Zink M, Suh K, Gu Y, et al. Characteristics of YouTube network traffic at a campus network. Comput Netw. 2009;53(4):501-514.
[22] Tran TX, Pompili D. Adaptive bitrate video caching and processing in mobile-edge computing networks. IEEE Trans Mobile Comput. 2019;18(9):1965-1978.
[23] Kim D, Seo DW, Choi M. Edge caching and computing of video chunks in multi-tier wireless networks. J Netw Comput Appl. 2024;226:103889. doi:10.1016/j.jnca.2024.103889.
[24] Sharifi F, Hessabi S, Lee YC. Cluster-based cooperative fog caching for scalable coded videos of multiple content providers. Internet Things. 2024;28:101380. doi:10.1016/j.iot.2024.101380.
[25] Uriol J, Yeregui I, Gabilondo A, Viola R, Angueira P, Montalban J. Context-Aware Adaptive Prefetching for DASH Streaming over 5G Networks. In: 2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). 2023. doi:10.1109/BMSB58369.2023.10211275.
[26] Shao J, Zhang X, Zhang J. Task-Oriented Communication for Edge Video Analytics. IEEE Trans Wireless Commun. 2024;23(5):4141-4154. doi:10.1109/TWC.2023.3314888.
[27] Xiao X, Zuo Y, Yan M, Wang W, He J, Zhang Q. Task-Oriented Video Compressive Streaming for Real-Time Semantic Segmentation. IEEE Trans Mobile Comput. 2024;23(12):14396-14413. doi:10.1109/TMC.2024.3446185.
[28] Satka Z, Mubeen S, Ashjaei M, Nordin D, Ragnarsson D. Enhancing Real-Time Networked Embedded Systems with End-to-end TSN-5G Networks. In: 2024 IEEE 14th International Symposium on Industrial Embedded Systems (SIES). 2024:104-111. doi:10.1109/SIES62473.2024.10768048.
[29] Fraunhofer Heinrich Hertz Institute. JSVM reference software for the scalable video coding extension of H.264/AVC. JSVM version 9.19.15. Available from: https://vcgit.hhi.fraunhofer.de/jvet/jsvm.
[30] 4MAPS, Welle D. Armature en acier de pliage [video]. Wikimedia Commons; 2024. CC BY-SA 4.0. Available from: https://commons.wikimedia.org/wiki/File:Armature_en_acier_de_pliage.webm.
[31] Dalvi M, Singh N, Bhingarde S, Chalke K. Construction-PPE: Personal Protective Equipment Detection Dataset. Ultralytics; 2025. Available from: https://docs.ultralytics.com/datasets/detect/construction-ppe/.
[32] Gupta A. SafetyVision YOLOv8: PPE and safety-violation detection model, v2 best_640.onnx. Hugging Face. Available from: https://huggingface.co/ayushgupta7777/safetyvision-yolov8.