AoI is Incomplete: Age of Semantics (AoS)-driven Adaptive Frame/Segment Control for Machine-centric Streaming Transmission

Ruichao Zhang[1,2], Lizhuang Tan[1,2], Maher Guizani[3], Wei Zhang[1,2], Hongxia Zhang[4], Peiying Zhang[4]
[1]*Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Ji'nan, China*
[2]*Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Ji'nan, China*
[3]*Department of Computer Science and Engineering, University of Texas Arlington, Arlington, United States*
[4]*College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China*
Email: tanlzh@sdas.org

***Abstract*—Machine-centric streaming transmission for inference must balance freshness and semantic availability, yet bandwidth variation and scene dynamics make this difficult. Optimizing only freshness-oriented metrics often favors Frame-by-Frame (FBF) transmission under congestion, causing intermediate frames to be dropped or invalid and undermining inference stability compared with Segment-by-Segment (SBS) transmission. We propose Age of Semantics (AoS) to jointly characterize freshness and semantic capture, with scene variation intensity modulating their relative importance. Based on AoS, we design a real-time adaptive control strategy that switches between FBF and SBS under bandwidth constraints using hysteresis to suppress oscillations near the switching boundary. Experiments on multiple datasets under different bandwidth levels demonstrate consistent gains over fixed modes and freshness baselines, reducing average AoS by ∼15% while maintaining stable real-time control behavior.**

***Index Terms*—Age of Information; Machine-centric Streaming Transmission; Real-time Control; Adaptive Transmission Control.**

I. Introduction

Driven by the rapid evolution of wireless communications, edge computing, and intelligent sensing technologies, streaming media has come to dominate modern network traffic [1]. As shown in Figure 1, the streaming transmission systems are evolving from human-centric to machine-centric. In the human-centric streaming system, we focus on the Quality of Experience (QoE) [2] during user viewing and interaction. Common optimization practices include employing improved encoding/decoding [3], adaptive transmission [4], and closed-loop control [5] mechanisms to reduce stuttering and improve clarity.

Meanwhile, in the machine-centric streaming system, the processing focus of streaming is on machine sensing and content analysis in applications like autonomous driving, intelligent inspection and machine vision. Pixel level restoration does not always improve analysis results in these tasks. Conversely it might bring extra transmission and computation overheads. More critically outdated information or missing key semantic information weakens the reliability of machine analysis directly. Therefore, the design goal of video communication systems changes from Quality of Experience to Quality of Analytics (QoA) [6].

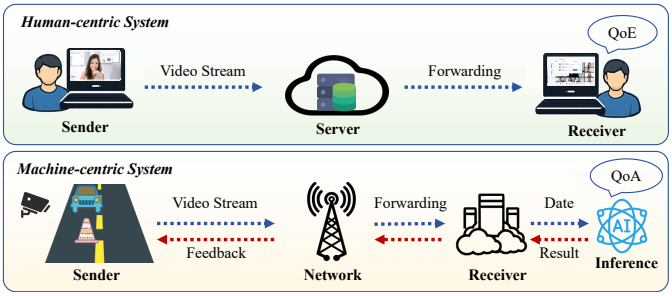

Fig. 1: Differences between Human-centric and Machine-centric Streaming Transmission Systems

In streaming system, Age of Information (AoI) [7] is often employed to measure the freshness of information. Building on this, Age of Semantic Information (AoSI) [8] incorporates semantic validity into the measurement framework to facilitate the design of analytics-oriented streaming systems. From QoA perspective, Age of Information alone cannot ensure continuous, inference-usable content. However, relying solely on AoSI as the optimization objective presents limitations in real-time video analytics. AoSI imposes substantial penalties on queuing delay and transmission delay. Consequently, systems tend to favor Frame-by-Frame (FBF) transmission to mitigate the age growth caused by queuing. The cost is that newly generated intermediate frames are prone to being discarded or expiring during link congestion. The information received at the receiver becomes sparse and exhibits semantic discontinuity. This weakens semantic information and affects inference

This work was supported by the research projects funded by the National Key R&D Program of China under Grant No.2024YFB2907000, the National Natural Science Foundation of China under Grant No.62471493, the Shandong Provincial Natural Science Foundation under Grant No.ZR2023LZH017, the Pilot Project for Integrated Innovation of Science, Education and Industry of Qilu University of Technology (Shandong Academy of Sciences) under Grant No.2025ZDZX01.

stability for analysis tasks that require continuous video context. In contrast, Segment-by-Segment (SBS) transmission preserves continuous content more effectively through grouping and inter-frame compression. However, it introduces additional encoding delay. SBS often resides in a disadvantaged position under evaluations based solely on AoSI. This phenomenon indicates that a single freshness metric struggles to characterize both freshness and completeness requirements simultaneously. Consequently, it fails to fully reflect the genuine needs of machine-centric streaming transmission system.

To address this, this paper proposes a new metric, Age of Semantics (AoS), which characterizes information freshness and semantic availability simultaneously. The formulation of AoS depends on two estimable quantities. The first is scene variation intensity, which controls the relative weight between the freshness and the completeness. The second is the semantic capture ratio, which describes the semantic completeness available within a given time window. AoS focuses more on freshness when scene variation is weak and it emphasizes maintaining a high semantic capture ratio more when scene variation is significant. It is necessary to emphasize that scene variation intensity only affects weight allocation. Mode control still needs to be considered jointly with conditions such as bandwidth to form stable decisions between Frame-by-Frame (FBF) transmission and Segment-by-Segment (SBS) transmission. Based on AoS this paper further designs the AoS-driven Adaptive Control Strategy (AAC). AAC estimates the AoS of the two transmission modes and dynamically switches between FBF and SBS. The strategy introduces a hysteresis mechanism to suppress frequent switching. Thus, it adapts better to complex environments where bandwidth fluctuation and scene variation coexist. The main contributions of this paper are as follows.

- Propose the Age of Semantics (AoS). This metric characterizes information freshness and semantic availability simultaneously and provides a construction method suitable for real time calculation.
- Propose the AoS-driven Adaptive Control Strategy (AAC). Adaptive control is performed between Frame-by-Frame (FBF) and Segment-by-Segment (SBS). A hysteresis mechanism is introduced to suppress frequent switching which ensures decision stability.
- Conduct experimental verification under machine analysis tasks. Experiments cover different bandwidth conditions including bandwidth limitation and time varying bandwidth while considering different levels of scene variation intensity. The results demonstrate that the proposed method has superior robustness and overall gain.

II. Related Works and Motivation

A. Machine-centric Streaming Transmission

In recent years, semantic communication has extended to image and video streaming to prioritize inference-usable content under constrained resources [9]–[11]. Here, semantics refers to the semantic capture rate for inference rather than semantic coding or communication paradigms. Against this backdrop, machine-centric streaming targets real-time inference [12] scenarios such as autonomous driving, intelligent inspection, and UAVs, where streaming transmission serves the inference pipeline rather than human viewing.

Xu *et al.* [13] proposed a systematic review of edge video analytics systems and summarized typical device-edge collaboration patterns, providing a general framework reference for the system design of machine-centric streaming. Zhang *et al.* [14] proposed FHVAC, an adaptive configuration method for machine-centric live streaming, which conducts online configuration and resource scheduling aiming at the delivery of information related to inference utility. Xiao *et al.* [15] proposed a task-oriented video transmission method for real-time semantic segmentation, which introduces inference goals into bitrate control and compression decisions to adapt to bandwidth-constrained and fluctuating scenarios. Chen *et al.* [16] proposed DeVA, a mobile-edge collaborative video analytics framework, which combines device-side lightweight processing and adaptive encoding to control analysis error while reducing transmission overhead. The aforementioned works collectively indicate that machine-centric streaming requires simultaneously balancing timeliness and inference availability. Optimizing solely for either dimension may lead to performance imbalance under complex network conditions and scene variations.

B. Age of Information

Kaul *et al.* [17] proposed Age of Information (AoI) to measure information freshness, forming a systematic optimization framework revolving around state updates and scheduling mechanisms. Chiariotti *et al.* [18] proposed Query Age of Information, which evaluates freshness at the query instant and discusses query-oriented scheduling strategies. Maatouk *et al.* [19] proposed Age of Incorrect Information, which characterizes cost based on the duration of incorrect states and constructs update strategies accordingly. Meesena *et al.* [20] proposed Age of Processed Information, which couples transmission and processing into a unified timeliness metric to support faster acquisition of available results. Huang *et al.* proposed Age of Semantic Information (AoSI), which incorporates valid semantic updates into the latency framework and designs transmission and scheduling methods minimizing AoSI to support analytics-oriented streaming systems.

C. Motivation

AoSI aligns more closely with analysis objectives, however, its cost function still imposes strong penalties on queuing and transmission latency. This makes the optimization process prone to prioritizing the suppression of age growth, thereby biasing the system towards Frame-by-Frame (FBF) transmission with less waiting time over Segment-by-Segment (SBS) transmission under constrained bandwidth. As shown in Figure 2, the structural consequences resulting from this are illustrated. In FBF mode, the continuous arrival of newly generated frames during link occupation leads to the discarding of intermediate

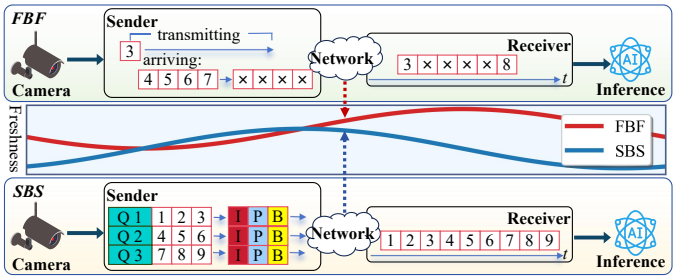

Fig. 2: Two transmission model: Frame-by-Frame (FBF) and Segment-by-Segment (SBS)

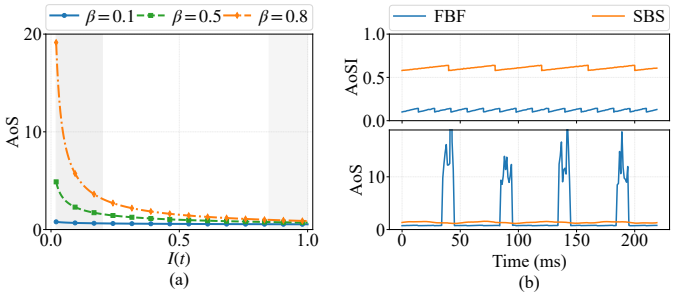

Fig. 3: Impact of β_t on AoS.

frames which results in a sparse sequence that degrades continuous frame information. In contrast, SBS uses inter-frame compression to enhance coding efficiency, making it easier to preserve continuous frame sequences under identical bandwidth conditions and thereby retaining more complete continuous frame information. This suggests that, when the system is optimized solely for Age of Semantic Information, it may achieve a lower instantaneous freshness in terms of the objective value, yet it is more likely to introduce missing intermediate frames under congestion, thereby undermining the stability of downstream inference.

Motivated by the aforementioned considerations, this paper establishes a unified metric that simultaneously characterizes freshness and the degree of semantic availability. Guided by this metric, a real-time adaptive control strategy is constructed. This strategy incorporates bandwidth and scene variation signals to make stable selections between FBF and SBS, while avoiding frequent oscillation near the boundaries.

III. Age of Semantics

In this section, we define Age of Semantics and introduce its key components, and then analyze how the scene variation factor affects the metric behavior.

A. AoS

In machine-centric streaming, the system must balance information freshness and inference-usable semantic coverage at the receiver. In the AoSI metric proposed by Huang *et al.*, the age grows over time between semantically valid updates and is reset only when a valid semantic update is successfully received. Building on this, we further introduce scene variation intensity and semantic capture rate to characterize semantic completeness. Defining Age of Semantics (AoS) as

$$\mathrm{AoS}(t) \triangleq \frac{(\mathrm{AoSI}(t))^{1-\beta_t}}{(I(t)+\epsilon)^{\beta_t}}, \tag{1}$$

where $\epsilon = 10^{-4}$ is a fixed numerical stabilizer used to avoid instability when $I(t) \to 0$. We use this default value throughout the experiments. $\beta_t \in [0,1]$ denotes the scene variation intensity, which serves solely to adaptively allocate weights between freshness and semantic capture. $I(t) \in [0,1]$ represents the semantic capture rate, which characterizes the proportion of timely available semantics within a sliding window. We adopt AoS as the optimization metric, where a smaller AoS indicates lower semantic cost and better overall system performance.

The working mechanism of Eq. (1) can be intuitively understood by examining two extreme scenarios. When scene variation is weak, AoS degenerates to AoSI, and the system cost is primarily determined by freshness. As scene variation intensifies, AoS becomes increasingly sensitive to the decline of $I(t)$, thereby penalizing semantic incompleteness more strongly. For efficient computation and threshold-based comparison, Eq. (1) is further rewritten in the log domain using the natural logarithm, which converts the multiplicative trade-off into an additive cost:

$$J(t) = (1-\beta_t)\ln \mathrm{AoSI}(t) - \beta_t \ln\left(I(t)+\epsilon\right). \tag{2}$$

B. Scene Variation Intensity

Scene variation intensity serves to characterize the relative level of current scene dynamics. As shown in Figure 3, it serves as a weighting factor in Eq. (1) driving the adaptive shift of cost emphasis between freshness and semantic capture. To satisfy the low overhead requirements for real-time control, this paper leverages compressed-domain motion cues to construct β_t. The encoder side directly parses block-level Motion Vectors (MVs) [21], where the magnitude of each block MV characterizes local motion amplitude, and subsequently aggregates the magnitudes of all blocks within a single frame to derive the frame-level motion intensity $\Delta_{mv}(t)$. To mitigate potential issues such as camera jitter or background noise, short-window smoothing is applied to $\Delta_{mv}(t)$ to derive $\bar{\Delta}_{mv}(t)$. Scene variation intensity is defined using the *Sigmoid* function as

$$\beta_t = \frac{1}{1+e^{-\left(\alpha\cdot\bar{\Delta}\mathrm{mv}(t)-\gamma\right)}}, \tag{3}$$

where $\bar{\Delta}_{mv}(t)$ is normalized to $[0,1]$, and $\alpha \in [6,14]$ and $\gamma \in [0.3,0.7]$ are fixed across datasets to control the slope and midpoint. Under this transformation, β_t approaches 0 when the scene is static or exhibits weak motion. Conversely, as motion significantly intensifies, β_t rapidly ascends towards 1. This characteristic provides a stable and computable scene variation signal for subsequent joint decision-making with network quantities such as bandwidth.

C. Semantic Capture Rate

The semantic capture rate $I(t)$ characterizes the inference-usable semantic coverage at the receiver at time t. The sliding window $W(t)$ covers the most recent K frames, where semantic units are indexed by k. For each unit, $q_k \in (0, 1]$ denotes its normalized task-relevance weight derived from downstream inference cues, where more important units are assigned larger weights; s_k denotes its encoded size; and $a_k(t) \in \{0, 1\}$ is an availability indicator. Specifically, $a_k(t) = 1$ only if unit k is successfully received and decoded in time for inference at time t; otherwise, $a_k(t) = 0$. The semantic capture rate is defined as

$$I(t) = \frac{\sum_{k \in W(t)} q_k \cdot s_k \cdot a_k(t)}{\sum_{k \in W(t)} q_k \cdot s_k}. \tag{4}$$

In Eq. (4), the numerator represents the total weighted semantic payload effectively delivered on time within the window. The denominator denotes the upper bound of the weighted payload of semantics generated within the same window. Therefore, the range of $I(t)$ is from 0 to 1. When bandwidth is constrained or congestion intensifies, the quantity of semantic units that can be timely acquired by the receiver and directly used for inference decreases, thereby causing $I(t)$ to decline. This decline manifests in Eq. (1) as a higher semantic cost, thereby making the system more sensitive to semantic gaps. As shown in Figure 3, it compares the differences between FBF and SBS in terms of freshness and semantic availability coverage illustrating that optimizing solely for freshness may lead to the accumulation of semantic gaps.

IV. AoS-driven Adaptive Control Strategy

Built on age of semantics, this section develops an AoS-driven adaptive control strategy for machine-centric streaming transmission systems. We present a hierarchical closed-loop framework and an online mode control rule based on cost difference and hysteresis.

A. Decision Space

The control state of AAC is composed of two types of quantities: network and content, namely $\{\beta_t, \hat{B}(t)\}$. $\hat{B}(t)$ reflects the link average resource level, and β_t reflects the scene variation intensity, which serves solely to allocate weights between the freshness term and the semantic capture term in AoS. As shown in Figure 4, with normalized bandwidth $B = \min(\hat{B}(t)/8 \text{ Mbps}, 1)$, we construct an interpretable decision map over the (B, β) plane by evaluating the mode-wise objective gap $\Delta J(B, \beta) = J_{\text{SBS}}(B, \beta) - J_{\text{FBF}}(B, \beta)$. For each mode, J is obtained by substituting the mode-dependent AoSI and semantic capture rate I into the log-domain objective in Eq. (2). The resulting sign of ΔJ partitions the plane into regions where one mode is preferable, while a transition band near $\Delta J \approx 0$ motivates the hysteresis design below. The upper left side of the boundary curve represents the region where $\Delta J < 0$, indicating that the cost of SBS is lower, which is suitable for scenarios with relatively low bandwidth and strong scene variation. The lower right side of the boundary

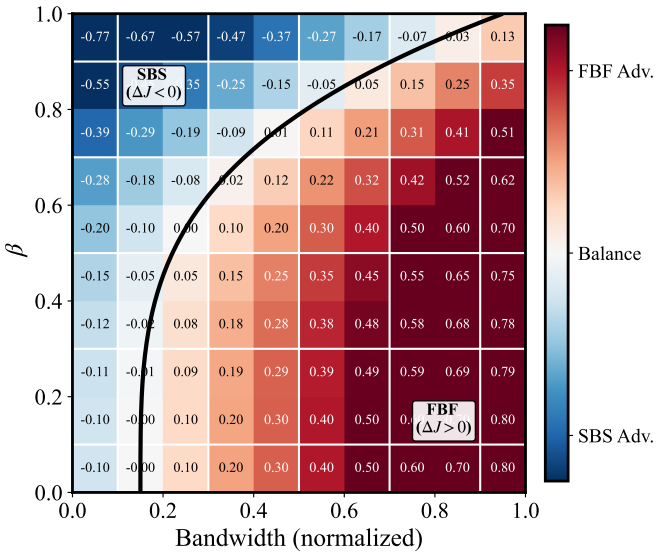

Fig. 4: Decision space.

curve represents the region where $\Delta J > 0$, indicating that the cost of FBF is lower, which is suitable for scenarios with relatively high bandwidth or weak scene variation. A transition band exists near the boundary curve where ΔJ approaches 0, indicating that the costs of both modes are similar, and the subsequent switching logic introduces hysteresis in this region to avoid frequent switching. Through this decision plane, AAC makes the influence of bandwidth and scene variation on mode selection explicit in an interpretable manner, providing a direct basis for real time control.

B. System Framework

As shown in Figure 5, AAC follows a two-layer feedback architecture with a control plane above a transmission loop. For each input frame x_i, the loop performs adaptive encoding, transmission, semantic decoding and recovery, and inference, then reports runtime measurements to the control plane. The control plane includes four modules, namely state estimation, cost evaluation, decision logic, and mode switching. State estimation extracts β_t from compressed-domain MVs and updates the bandwidth estimate $\hat{B}(t)$ using ACK arrival rate and packet-loss feedback. Cost evaluation uses a precomputed look-up table with inputs $\hat{B}(t)$ and β_t to obtain $J_{\text{FBF},t}$ and $J_{\text{SBS},t}$, and outputs ΔJ_t at each decision instant by default once per frame. Decision logic selects m_t based on ΔJ_t, m_{t-1}, and the hysteresis threshold, then mode switching applies m_t to the encoder to toggle between FBF and SBS for real-time adaptation to bandwidth fluctuations and content variations.

C. Mode Control Rule

The strategic core of AAC resides in the Decision Logic module, which utilizes the AoS cost difference as the unified criterion. The cost difference function is defined as $\Delta J(t) = J_{\text{SBS}}(t) - J_{\text{FBF}}(t)$. SBS is selected when $\Delta J(t) < 0$, while FBF

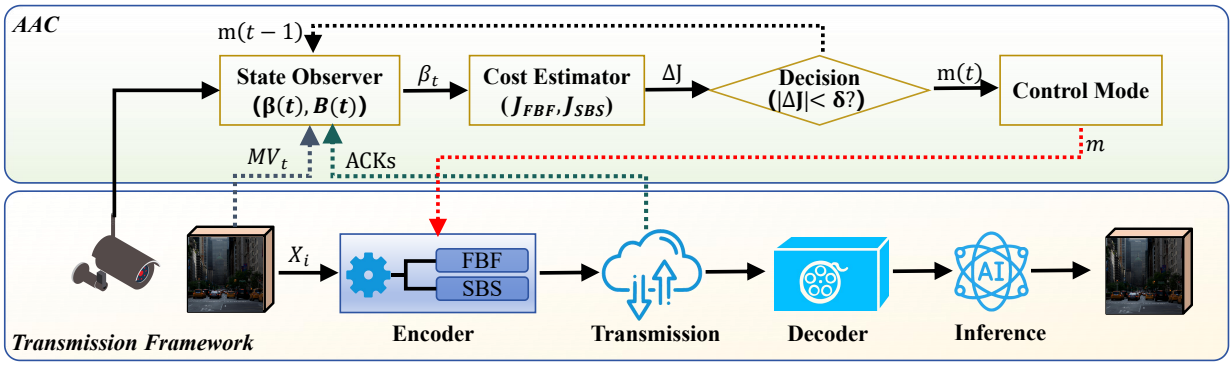

Fig. 5: Framework of AAC.

is selected when $\Delta J(t) > 0$. To suppress frequent switching near the boundary, we use a hysteresis threshold $\delta_{th} = 0.05$, chosen to exceed steady-state short-term fluctuations of $\Delta J(t)$. When $|\Delta J| \leq \delta_{\text{th}}$, the mode at the previous instant remains unchanged. At each decision instant, the system needs only update β_t and $B(t)$ based on motion cues and transmission feedback, and calculate the costs of the two candidate modes. Subsequently, the mode update is finalized via a single threshold comparison. The overall computational overhead is low, satisfying real-time control requirements.

V. Experimental Results

This section validates the proposed metric and control strategy via simulation. We describe the setup and baselines, and report results on mechanism complementarity, online switching behavior, overall performance trends, as well as cross-dataset and ablation results.

A. Experimental Setup

The experiments use *MOT17* [22] and *MOT20* [23] as benchmark datasets, and further include *SeaDronesSee* [24] and *VisDrone* [25] for cross-scenario validation. At the receiver, a pretrained *YOLOv10-s* detector runs at 640 input resolution with confidence threshold 0.25 and NMS IoU threshold 0.45 on a single NVIDIA GPU, achieving about 30 frames per second. The resulting detections are used to compute $I(t)$ and the AoS cost. On the network side, the simulated time-varying bandwidth $B(t)$ spans constrained to sufficient regimes. Each configuration is repeated for multiple independent runs to obtain stable statistics. We compare against fixed-mode FBF, fixed-mode SBS, and the AoSI baseline in [8].

B. Mechanism and Behavior

As shown in Figure 6, FBF and SBS exhibit complementarity at the semantic level. Under constrained bandwidth, FBF is more susceptible to a decline in the semantic capture rate, leading to reduced semantic availability. In contrast, SBS leverages inter frame compression to maintain stable semantic capture under identical resource conditions. As bandwidth increases, the freshness advantage of FBF gradually emerges, allowing it to dominate in the high-bandwidth region. Consequently,

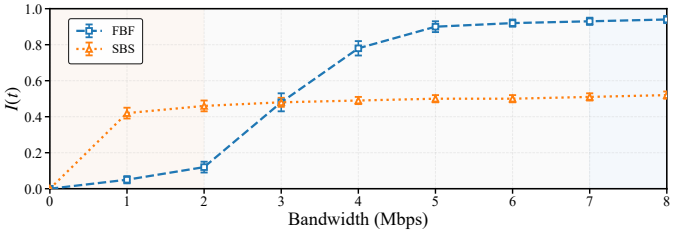

Fig. 6: Semantic complementarity of FBF and SBS.

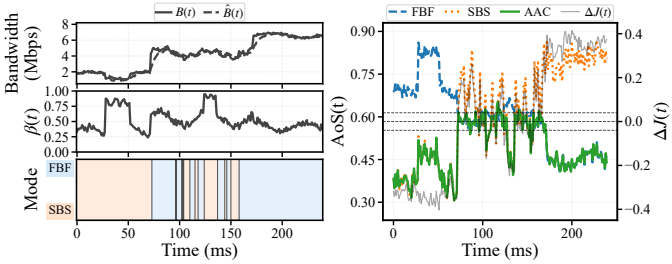

Fig. 7: Real time control under temporal variations.

a single mode cannot simultaneously balance freshness and semantic coverage across the entire bandwidth range.

As shown in Figure 7, AAC adaptively controls between the two modes according to variations in $B(t)$ and β_t, further validating its real time control behavior. In regions where the cost difference approaches 0, the hysteresis mechanism suppresses frequent switching, enabling the AoS trajectory to consistently track the superior fixed mode. Thereby, it maintains stable semantic availability coverage amidst the coexistence of dynamic bandwidth and temporal content variations.

C. Overall Performance

As shown in Figure 8, the average AoS results under different bandwidth levels are summarized to illustrate the overall performance trend of AAC. At lower bandwidths, AAC tends to select SBS to mitigate semantic gaps. Conversely, at higher bandwidths, it favors FBF to retain the freshness advantage. In the intermediate bandwidth region, it performs adaptive switching by incorporating scene variation intensity.

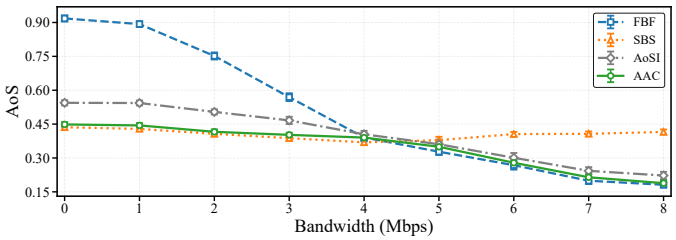

Fig. 8: Overall performance across different bandwidth levels.

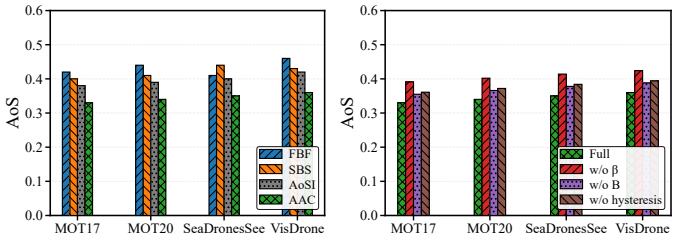

Fig. 9: Public dataset results and ablation results.

Consequently, it achieves a lower comprehensive cost across the entire range.

As shown in Figure 9, the cross-dataset results and ablation analysis indicate that the gains of AAC remain consistent across different datasets. Further ablation results demonstrate that the exclusion of β_t, $B(t)$, or the hysteresis mechanism leads to an increase in AoS and impairs control stability, thereby validating the essential contribution of each component to the overall performance.

VI. Conclusion

This paper proposes Age of Semantics (AoS), a unified metric that combining information freshness and semantic capture, where scene variation intensity adjusts their relative importance. Based on AoS, we develop a real-time adaptive control strategy that switches between Frame-by-Frame (FBF) and Segment-by-Segment (SBS) transmission with hysteresis to reduce oscillations. Experiment results on multiple datasets under different bandwidth conditions show consistent metric gains and stable real-time control. In future, we will improve semantic unit and correlation modeling, enhance bandwidth and scene variation estimation, and validate end-to-end performance using real traces and edge inference pipelines. We will also extend the framework to multi-stream and multi-user settings with scalable online control.

References

[1] A. Badalyan. (2024, Dec.) Top 40 mobile traffic statistics you should know. Digital Silk. Updated December 12, 2024. [Online]. Available: https://www.digitalsilk.com/digital-trends/top-mobile-traffic-statistics/

[2] W. Shi, Q. Li, Q. Yu, F. Wang, G. Shen, Y. Jiang, Y. Xu, L. Ma, and G.-M. Muntean, "A survey on intelligent solutions for increased video delivery quality in cloud-edge-end networks," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 2, pp. 1363–1394, 2024.

[3] Y. Duan, Q. Du, X. Fang, Z. Xie, Z. Qin, X. Tao, C. Pan, and G. Liu, "Multimedia semantic communications: Representation, encoding and transmission," *IEEE Network*, vol. 37, no. 1, pp. 44–50, 2023.

[4] L. Tan, X. Dong, X. Gao, P. Zhang, W. Su, and J. W.-K. Hong, "Enhancing QUIC Performance in Heterogeneous Networks: A Proactive Connection Migration Approach," *International Journal of Network Management*, vol. 35, no. 5, p. e70022, 2025.

[5] Y. Zhu, Y. Huang, X. Qiao, Z. Tan, B. Bai, H. Ma, and S. Dustdar, "A semantic-aware transmission with adaptive control scheme for volumetric video service," *IEEE Transactions on Multimedia*, vol. 25, pp. 7160–7172, 2022.

[6] S. Wang, S. Bi, and Y.-J. A. Zhang, "Edge video analytics with adaptive information gathering: A deep reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 9, pp. 5800–5813, 2023.

[7] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1183–1210, 2021.

[8] Z. Huang, W. Wu, K. Wu, G. Gao, and J. Wang, "Minimizing Age of Semantic Information for Analytics-Oriented Video Streaming Systems," *IEEE Transactions on Mobile Computing*, vol. 24, no. 12, pp. 12 885–12 902, 2025.

[9] T. M. Getu, G. Kaddoum, and M. Bennis, "Semantic communication: A survey on research landscape, challenges, and future directions," *Proceedings of the IEEE*, vol. 112, no. 11, pp. 1649–1685, 2024.

[10] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 230–244, 2022.

[11] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, and V. C. Leung, "Predictive and adaptive deep coding for wireless image transmission in semantic communication," *IEEE Transactions on Wireless Communications*, vol. 22, no. 8, pp. 5486–5501, 2023.

[12] J. Shao, X. Zhang, and J. Zhang, "Task-oriented communication for edge video analytics," *IEEE Transactions on Wireless Communications*, vol. 23, no. 5, pp. 4141–4154, 2023.

[13] R. Xu, S. Razavi, and R. Zheng, "Edge video analytics: A survey on applications, systems and enabling techniques," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2951–2982, 2023.

[14] Y. Zhang, W. Zhang, H. Du, C. Yan, L. Liu, and Q. Zheng, "Fhvac: Feature-level hybrid video adaptive configuration for machine-centric live streaming," *IEEE Transactions on Parallel and Distributed Systems*, vol. 35, no. 5, pp. 780–795, 2024.

[15] X. Xiao, Y. Zuo, M. Yan, W. Wang, J. He, and Q. Zhang, "Task-oriented video compressive streaming for real-time semantic segmentation," *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, pp. 14 396–14 413, 2024.

[16] S. Chen, J. Yin, R. Zhong, and F. Liu, "Deva: an edge-assisted video analytics framework for depth estimation," *IEEE Transactions on Mobile Computing*, vol. 24, no. 12, pp. 13 177–13 190, 2025.

[17] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proceedings of INFOCOM'12*. IEEE, 2012, pp. 2731–2735.

[18] F. Chiariotti, J. Holm, A. E. Kalør, B. Soret, S. K. Jensen, T. B. Pedersen, and P. Popovski, "Query age of information: Freshness in pull-based communication," *IEEE Transactions on Communications*, vol. 70, no. 3, pp. 1606–1622, 2022.

[19] A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: An enabler of semantics-empowered communication," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2621–2635, 2022.

[20] W. Meesena, C. Nikunram, S. J. Turner, and S. Supittayapornpong, "Minimizing age of processed information over unreliable wireless network channels," *IEEE Transactions on Mobile Computing*, vol. 24, no. 5, pp. 3567–3578, 2025.

[21] S. M. Ayyoubzadeh, W. Liu, I. Kezele, Y. Yu, X. Wu, Y. Wang, and T. Jin, "Test-time adaptation for optical flow estimation using motion vectors," *IEEE Transactions on Image Processing*, vol. 32, pp. 4977–4988, 2023.

[22] Y. Li, L. Zhan, L. Wu, H. Liu, J. Min, and X. Wang, "Re-identification assistance and multi-stage association for pedestrian multi-object tracking," *Scientific Reports*, vol. 15, no. 1, p. 22533, 2025.

[23] J. Alikhanov, D. Obidov, M. Abdurasulov, and H. Kim, "Practical evaluation framework for real-time multi-object tracking: Achieving optimal and realistic performance," *IEEE Access*, vol. 13, pp. 34 768–34 788, 2025.

[24] L. A. Varga, B. Kiefer, M. Messmer, and A. Zell, "Seadronessee: A maritime benchmark for detecting humans in open water," in *Proceedings of WACV'22*. IEEE, 2022, pp. 3686–3696.

[25] Y. Cao, Z. He, L. Wang, W. Wang, Y. Yuan, D. Zhang, J. Zhang, P. Zhu, L. Van Gool, J. Han, S. Hoi, Q. Hu, and M. Liu, "Visdrone-det2021: The vision meets drone object detection challenge results," in *Proceedings of ICCVW'21*, vol. 2021-October. IEEE/CVF, 2021, pp. 2847–2854.