Novelty Assessment Report

Paper: InfoTok: Adaptive Discrete Video Tokenizer via Information-Theoretic Compression

PDF URL: https://openreview.net/pdf?id=JEYWpFGzvn

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2025-12-27

Abstract

Accurate and efficient discrete video tokenization is essential for long video sequences processing. Yet, the inherent complexity and variable information density of videos present a significant bottleneck for current tokenizers, which rigidly compress all content at a fixed rate, leading to redundancy or information loss. Drawing inspiration from Shannon's information theory, this paper introduces \alg, a principled framework for adaptive video tokenization. We rigorously prove that existing data-agnostic training methods are suboptimal in representation length, and present a novel evidence lower bound (ELBO)-based algorithm that approaches theoretical optimality. Leveraging this framework, we develop a transformer-based adaptive compressor that enables adaptive tokenization. Empirical results demonstrate state-of-the-art compression performance, saving 20% tokens without influence on performance, and achieving $2.3\times$ compression rates while still outperforming prior heuristic adaptive approaches. By allocating tokens according to informational richness, \alg enables a more compressed yet accurate tokenization for video representation, offering valuable insights for future research.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Adaptive Video Tokenization via Information-Theoretic Compression**

A total of **18 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Information-Theoretic Adaptive Compression Frameworks**
- **Spatiotemporal Adaptive Token Reduction for Video-Language Models**
- **Neural Discrete Representation Learning for Video Tokenization**
- **Generative Video Tokenization with Adaptive Spatial Allocation**
- **Attention-Based Token Reduction for Video Understanding**
- **Content-Adaptive Learned Video Compression**
- **Inference-Time Adaptive Tokenization via Online Compression**
- **Application-Specific Adaptive Video Compression**

Complete Taxonomy Tree

- Adaptive Video Tokenization via Information-Theoretic Compression Survey Taxonomy
- Information-Theoretic Adaptive Compression Frameworks
  - Entropy-Based Optimal Tokenization ★ (1 papers)
  - [0] InfoTok: Adaptive Discrete Video Tokenizer via Information-Theoretic Compression (Anon et al., 2026) View paper
  - Information Uniqueness-Driven Compression (1 papers)
  - [12] UniComp: Rethinking Video Compression Through Informational Uniqueness (Chao Yuan, 2025) View paper
- Spatiotemporal Adaptive Token Reduction for Video-Language Models
  - Cross-Modal Query-Guided Compression (1 papers)
  - [1] LongVU: Spatiotemporal Adaptive Compression for Long Video-Language Understanding (Shen, 2024) View paper
  - Progressive Multi-Frame Compression (1 papers)
  - [5] PVC: Progressive Visual Token Compression for Unified Image and Video Processing in Large Vision-Language Models (Chen-Yu Yang, 2024) View paper
  - Dynamic Token Grouping and Pruning (1 papers)
  - [6] DynTok: Dynamic Compression of Visual Tokens for Efficient and Effective Video Understanding (Zhang Hong-zhi, 2025) View paper
  - Plug-and-Play Inference Acceleration (1 papers)
  - [4] Video Compression Commander: Plug-and-Play Inference Acceleration for Video Large Language Models (Xuyang Liu, 2025) View paper
- Neural Discrete Representation Learning for Video Tokenization
  - Extreme Token Reduction via Discrete Codebooks (1 papers)
  - [2] VQToken: Neural Discrete Token Representation Learning for Extreme Token Reduction in Video Large Language Models (Zhang Haichao, 2025) View paper
  - Multi-Stage Transformer Generation with Vector Quantization (1 papers)
  - [3] Unifying generation and compression: Ultra-low bitrate image coding via multi-stage transformer (Naifu Xue, 2024) View paper
  - Domain-Adaptive Codebook Learning (1 papers)
  - [16] Adaptive Human-Centric Video Compression for Humans and Machines (Wei Jiang, 2023) View paper

- Generative Video Tokenization with Adaptive Spatial Allocation
  - Gaussian Splatting-Based Tokenization (1 papers)
  - [8] Versatile Video Tokenization with Generative 2D Gaussian Splatting (Chen Zhenghao, 2025) View paper
  - Dynamic Latent Frame Rate Adaptation (1 papers)
  - [7] DLFR-VAE: Dynamic Latent Frame Rate VAE for Video Generation (Zhihang Yuan, 2025) View paper
  - Granularity-Adaptive Spatial Tokenization (1 papers)
  - [10] Granularity-Adaptive Spatial Evidence Tokenization for Video Question Answering (Hao Jiang, 2025) View paper
- Attention-Based Token Reduction for Video Understanding (1 papers)
  - [11] Motion Guided Token Compression for Efficient Masked Video Modeling (Feng, 2024) View paper
- Content-Adaptive Learned Video Compression
  - Online Motion Rate Adaptation (1 papers)
  - [15] Content-Adaptive Motion Rate Adaption For Learned Video Compression (Chih-Hsuan Lin, 2022) View paper
  - Joint Multi-Frame Training for Error Propagation Mitigation (1 papers)
  - [17] Content Adaptive and Error Propagation Aware Deep Video Compression (Guo Lu, 2020) View paper
  - Multi-Rate Transform Coding (1 papers)
  - [13] Multi-Rate Adaptive Transform Coding for Video Compression (Lyndon R. Duong, 2023) View paper
- Inference-Time Adaptive Tokenization via Online Compression (1 papers)
  - [9] zip2zip: Inference-Time Adaptive Tokenization via Online Compression (Geng, 2025) View paper
- Application-Specific Adaptive Video Compression
  - Semantic-Guided Edge-Based Compression (1 papers)
  - [14] Poster: "Semantic-Guided Skip Sampling and Soft Edge Compression for Real-Time Edge-Based Traffic Video Analytics" (Deng Pan, 2025) View paper
  - Block-Based Compressed Sensing with Saliency Detection (1 papers)
  - [18] Adaptive Block-Based Compressed Video Sensing Based on Saliency Detection and Side Information. (Wei Wang, n.d.) View paper

Narrative

Core task: adaptive video tokenization via information-theoretic compression. The field addresses how to efficiently represent video data by dynamically allocating tokens based on content complexity and information density. The taxonomy reveals several complementary directions: information-theoretic frameworks that optimize compression through entropy measures and rate-distortion principles; spatiotemporal adaptive methods that reduce tokens specifically for video-language models; neural discrete representation learning approaches that build codebooks for video tokenization; generative methods with adaptive spatial allocation; attention-based token reduction for understanding tasks; learned compression systems that adapt to content statistics; inference-time adaptive schemes that compress on-the-fly; and application-specific compression tailored to particular downstream tasks. Works like LongVU[1] and VQToken[2] exemplify spatiotemporal and discrete representation approaches respectively, while Ultra-low Bitrate Transformer[3] and PVC[5] demonstrate learned compression strategies that adapt to varying content characteristics.

Particularly active themes include the trade-off between compression efficiency and downstream task performance, the challenge of handling temporal redundancy versus spatial detail, and the question of whether to optimize tokenization jointly with task objectives or as a separate preprocessing step. InfoTok[0] sits within the information-theoretic adaptive compression frameworks branch, specifically focusing on entropy-based optimal tokenization. This positions it closely with works that use rate-distortion theory to guide token allocation, contrasting with purely learned approaches like DynTok[6] that adapt tokens through neural architectures without explicit information-theoretic objectives. Compared to application-specific methods such as Video Compression Commander[4] or Human-Centric Video Compression[16], InfoTok[0] appears to pursue a more general compression principle grounded in entropy optimization, aiming for broader applicability across tasks rather than tuning for specific downstream applications.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

Both subtopics address video tokenization through information-theoretic principles, aiming to reduce redundancy while preserving essential information. The original leaf focuses on entropy minimization and ELBO optimization for deriving optimal compression rates, while the sibling emphasizes measuring intrinsic token redundancy through information uniqueness metrics and conditional entropy minimization under budget constraints. The key distinction lies in the optimization framework: ELBO-based global optimization versus uniqueness-driven local redundancy measurement.

Similarities: - Both employ information-theoretic foundations (entropy, conditional entropy) for compression - Both aim to achieve optimal tokenization by reducing redundancy in video representations - Both operate within the context of adaptive video tokenization for efficient processing

Differences: - Original leaf uses ELBO optimization and entropy minimization as primary objectives; sibling uses information uniqueness metrics to identify redundant tokens - Original leaf derives optimal compression rates theoretically; sibling enforces budget constraints explicitly during token selection - Original leaf's exclusion of conditional entropy for redundancy measurement contrasts with sibling's core use of conditional entropy under budget constraints - Sibling explicitly excludes attention-based methods, while original leaf does not mention attention mechanisms in its scope

Suggested Search Directions: - Investigate hybrid approaches combining ELBO optimization with information uniqueness metrics - Explore the relationship between optimal compression rates from entropy minimization and budget-constrained conditional entropy approaches - Examine whether information uniqueness can serve as a proxy or complement to ELBO-based objectives

Sibling Subtopics

- **Information Uniqueness-Driven Compression** (leaves: 1, papers: 1)
- Scope: Methods measuring intrinsic token redundancy via information uniqueness to minimize conditional entropy under budget constraints.
- Exclude: Excludes ELBO-based optimization or attention-driven approaches; see Entropy-Based Optimal Tokenization or Attention-Based Token Reduction.

Contributions Analysis

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Theoretical proof of suboptimality in existing tokenizers

Description: The authors provide rigorous theoretical proofs demonstrating that both fixed-compression tokenizers and existing adaptive tokenizers using data-agnostic routers (such as uniform sampling) are suboptimal in terms of expected token length compared to information-theoretic optimality. They show these methods fail to achieve near-optimal compression rates.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. MOAT: Revealing the Task-Optimality Gap in Adaptive Tokenization

URL: View paper

Brief Assessment

MOAT[25] focuses on task-optimality gaps in adaptive tokenization for language processing, not video tokenization. The original paper proves suboptimality in video tokenizers using Shannon's information theory and ELBO-based methods, while MOAT[25] addresses conflicts between information-theoretic optimality and task-specific utility in text representation.

2. Unpacking tokenization: Evaluating text compression and its correlation with model performance

URL: View paper

Brief Assessment

Unpacking Tokenization[21] focuses on text tokenization (BPE) and compression as an intrinsic quality metric for NLP models, not on proving theoretical suboptimality of video tokenizers with fixed or adaptive compression rates using information-theoretic frameworks.

3. Training llms over neurally compressed text

URL: View paper

Brief Assessment

Neurally Compressed Text[24] focuses on training LLMs over neurally compressed text using arithmetic coding and equal-info windows, not on proving information-theoretic suboptimality of fixed or adaptive tokenizers with data-agnostic routers.

4. Single-pass adaptive image tokenization for minimum program search

URL: View paper

Brief Assessment

Adaptive Image Tokenization[20] focuses on single-pass adaptive image tokenization using Kolmogorov complexity principles for minimum program search. It does not provide theoretical proofs about the suboptimality of fixed-compression or data-agnostic adaptive tokenizers in terms of information-theoretic optimality or expected token length, which is the core novelty claim of the original paper.

5. Leveraging Information Theoretic ToolsFor Foundation Model Analysis

URL: View paper

Brief Assessment

Information Theoretic Tools[26] focuses on tokenization from an OOD adaptation and compression efficiency perspective, not on proving suboptimality of fixed or adaptive tokenizers with data-agnostic routers in terms of expected token length compared to information-theoretic optimality.

6. Tokenization and the noiseless channel

URL: View paper

Brief Assessment

Noiseless Channel[22] focuses on NLP tokenization for text sequences using information-theoretic compression principles, while the original paper addresses video tokenization with adaptive compression rates. The domains and technical approaches differ fundamentally.

7. Language modeling is compression

URL: View paper

Brief Assessment

Language Modeling Compression[19] focuses on the equivalence between prediction and compression in language models, demonstrating compression capabilities across modalities. It does not address the specific theoretical analysis of tokenizer suboptimality in terms of fixed vs. adaptive compression rates that the original paper claims as novel.

8. WSDL term tokenization methods for IR-style Web services discovery

URL: View paper

Brief Assessment

WSDL Term Tokenization[27] focuses on tokenizing WSDL terms for web service discovery using compression methods like PPM, not on video tokenization or information-theoretic optimality proofs for adaptive tokenizers.

9. HutterX â□□ Omniscientrix Hybrid Compressor (vÎ© Unified Informationalâ□□Awareness Framework Build) by Cornelius Aurelius

URL: View paper

Brief Assessment

HutterX Omniscientrix[28] focuses on text compression for the Hutter Prize using hybrid compression techniques, not on video tokenization or theoretical proofs about tokenizer optimality in visual domains.

10. Emergent architectural dynamics of neural token compression in large language models

URL: View paper

Brief Assessment

Neural Token Compression[23] focuses on emergent architectural dynamics in LLMs rather than information-theoretic proofs about tokenizer optimality. The candidate does not provide theoretical analysis of fixed-compression or data-agnostic adaptive tokenizers.

Contribution 2: INFOTOK framework with ELBO-based router and adaptive compressor

Description: The authors introduce INFOTOK, a novel framework for adaptive video tokenization that uses an Evidence Lower Bound (ELBO)-based router to determine token sequence lengths based on video information complexity, combined with a transformer-based adaptive compressor that efficiently compresses embeddings into variable-length token sequences.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 3: Empirical validation of superior token efficiency

Description: The authors conduct comprehensive experiments showing that INFOTOK achieves state-of-the-art compression performance, saving approximately 20% tokens without performance loss and achieving 2.3× better compression rates compared to prior adaptive approaches while maintaining or improving reconstruction quality.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. LTX-Video: Realtime Video Latent Diffusion

URL: View paper

Brief Assessment

LTX-Video[30] focuses on video generation using a transformer-based latent diffusion model with a high-compression video-VAE (1:192 ratio), not on adaptive video tokenization methods or token efficiency improvements in the context of discrete video tokenizers for reconstruction tasks.

2. Tokenlearner: Adaptive space-time tokenization for videos

URL: View paper

Brief Assessment

TokenLearner[31] focuses on adaptive spatial tokenization for video transformers using learned attention mechanisms, while INFOTOK uses information-theoretic principles (ELBO-based routing) for adaptive compression. The technical approaches and theoretical foundations differ fundamentally.

3. Don't Look Twice: Faster Video Transformers with Run-Length Tokenization

URL: View paper

Brief Assessment

Run-Length Tokenization[35] focuses on removing temporally redundant patches in videos through run-length encoding, not on adaptive video tokenization based on information-theoretic compression. The candidate addresses a different technical problem (temporal redundancy) using a different approach (run-length encoding) than INFOTOK's information-theoretic adaptive compression framework.

4. Image and Video Tokenization with Binary Spherical Quantization

URL: View paper

Brief Assessment

Binary Spherical Quantization[37] focuses on a different quantization method (binary spherical quantization on hyperspheres) rather than adaptive tokenization based on information theory. Their compression achievements stem from their novel quantization approach, not from adaptive token allocation based on video complexity as in the original paper.

5. Adaptive token sampling for efficient vision transformers

URL: View paper

Brief Assessment

Adaptive Token Sampling[29] focuses on reducing computational costs in vision transformers through adaptive token selection for image/video classification tasks, not on adaptive video tokenization methods for reconstruction. The candidate addresses token sampling efficiency in transformer architectures, while the original paper addresses token efficiency in discrete video tokenization with reconstruction objectives.

6. ADAPTOR: Adaptive Token Reduction for Video Diffusion Transformers

URL: View paper

Brief Assessment

ADAPTOR[36] focuses on token reduction for video diffusion transformers in generation tasks, while the original paper addresses adaptive video tokenization for reconstruction. These are fundamentally different applications with distinct objectives and methodologies.

7. Keyframe-oriented Vision Token Pruning: Enhancing Efficiency of Large Vision Language Models on Long-Form Video Processing

URL: View paper

Brief Assessment

Keyframe-oriented Pruning[34] focuses on vision token pruning for VLMs processing long-form videos, achieving 80% token reduction. The original paper addresses adaptive video tokenization for reconstruction tasks with 20% token savings and 2.3× compression rates. These are distinct technical domains with different objectives and evaluation metrics.

8. Rethinking video tokenization: A conditioned diffusion-based approach

URL: View paper

Brief Assessment

Conditioned Diffusion Tokenization[33] focuses on diffusion-based video reconstruction quality rather than adaptive token efficiency. The paper does not address adaptive tokenization methods or compression rate optimization based on video complexity, which are central to INFOTOK's contribution.

9. Dense Video Understanding with Gated Residual Tokenization

URL: View paper

Brief Assessment

Gated Residual Tokenization[32] focuses on dense video understanding through motion-compensated gating and scene merging for high-fps video processing, not on adaptive tokenization methods that adjust compression rates based on information complexity as in INFOTOK.

10. LongVU: Spatiotemporal Adaptive Compression for Long Video-Language Understanding

URL: View paper

Brief Assessment

LongVU[1] focuses on spatiotemporal compression for video-language understanding tasks using cross-modal queries and inter-frame dependencies, while INFOTOK addresses adaptive video tokenization for reconstruction based on information theory. These are fundamentally different approaches serving different purposes.

Appendix: Text Similarity Detection

Textual similarity detection checked 20 papers and found 1 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Tokenization and the noiseless channel

Detected in: Contribution: contribution_1

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] InfoTok: Adaptive Discrete Video Tokenizer via Information-Theoretic Compression View paper
- [1] LongVU: Spatiotemporal Adaptive Compression for Long Video-Language Understanding View paper
- [2] VQToken: Neural Discrete Token Representation Learning for Extreme Token Reduction in Video Large Language Models View paper
- [3] Unifying generation and compression: Ultra-low bitrate image coding via multi-stage transformer View paper
- [4] Video Compression Commander: Plug-and-Play Inference Acceleration for Video Large Language Models View paper
- [5] PVC: Progressive Visual Token Compression for Unified Image and Video Processing in Large Vision-Language Models View paper
- [6] DynTok: Dynamic Compression of Visual Tokens for Efficient and Effective Video Understanding View paper
- [7] DLFR-VAE: Dynamic Latent Frame Rate VAE for Video Generation View paper
- [8] Versatile Video Tokenization with Generative 2D Gaussian Splatting View paper
- [9] zip2zip: Inference-Time Adaptive Tokenization via Online Compression View paper
- [10] Granularity-Adaptive Spatial Evidence Tokenization for Video Question Answering View paper
- [11] Motion Guided Token Compression for Efficient Masked Video Modeling View paper
- [12] UniComp: Rethinking Video Compression Through Informational Uniqueness View paper
- [13] Multi-Rate Adaptive Transform Coding for Video Compression View paper
- [14] Poster: "Semantic-Guided Skip Sampling and Soft Edge Compression for Real-Time Edge-Based Traffic Video Analytics" View paper
- [15] Content-Adaptive Motion Rate Adaption For Learned Video Compression View paper
- [16] Adaptive Human-Centric Video Compression for Humans and Machines View paper
- [17] Content Adaptive and Error Propagation Aware Deep Video Compression View paper
- [18] Adaptive Block-Based Compressed Video Sensing Based on Saliency Detection and Side Information. View paper
- [19] Language modeling is compression View paper
- [20] Single-pass adaptive image tokenization for minimum program search View paper
- [21] Unpacking tokenization: Evaluating text compression and its correlation with model performance View paper
- [22] Tokenization and the noiseless channel View paper
- [23] Emergent architectural dynamics of neural token compression in large language models View paper
- [24] Training llms over neurally compressed text View paper
- [25] MOAT: Revealing the Task-Optimality Gap in Adaptive Tokenization View paper
- [26] Leveraging Information Theoretic ToolsFor Foundation Model Analysis View paper
- [27] WSDL term tokenization methods for IR-style Web services discovery View paper
- [28] HutterX â Omniscientrix Hybrid Compressor (vÎ© Unified InformationalâAwareness Framework Build) by Cornelius Aurelius View paper
- [29] Adaptive token sampling for efficient vision transformers View paper
- [30] LTX-Video: Realtime Video Latent Diffusion View paper
- [31] Tokenlearner: Adaptive space-time tokenization for videos View paper
- [32] Dense Video Understanding with Gated Residual Tokenization View paper
- [33] Rethinking video tokenization: A conditioned diffusion-based approach View paper
- [34] Keyframe-oriented Vision Token Pruning: Enhancing Efficiency of Large Vision Language Models on Long-Form Video Processing View paper
- [35] Don't Look Twice: Faster Video Transformers with Run-Length Tokenization View paper
- [36] ADAPTOR: Adaptive Token Reduction for Video Diffusion Transformers View paper
- [37] Image and Video Tokenization with Binary Spherical Quantization View paper