POLITECNICO DI TORINO
Repository ISTITUZIONALE

Deepfake Detection and Temporal Localization in HTTP Adaptive Streaming via Segment-Level Analysis

Original
Deepfake Detection and Temporal Localization in HTTP Adaptive Streaming via Segment-Level Analysis / Coduri, C., Sacco, A., Marchetto, G.. - (In corso di stampa). (ACM CoNEXT 2026 Student Workshop Utrecht, The Netherlands December 7-10, 2026) [10.1145/3843179.3858046].

Availability:
This version is available at: 11583/3017824 since: 2026-11-11T21:02:36Z

Publisher:
ACM

Published
DOI:10.1145/3843179.3858046

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
ACM postprint/Author's Accepted Manuscript

(Article begins on next page)

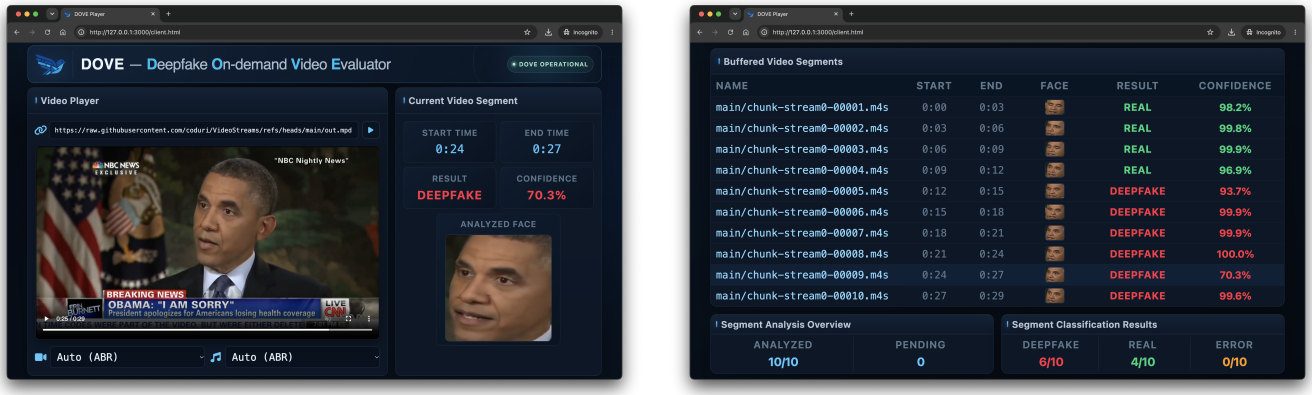
13 November 2026

Deepfake Detection and Temporal Localization in HTTP Adaptive Streaming via Segment-Level Analysis

Christian Coduri
christian.coduri@polito.it
Politecnico di Torino
Italy

Alessio Sacco
alessio_sacco@polito.it
Politecnico di Torino
Italy

Guido Marchetto
guido.marchetto@polito.it
Politecnico di Torino
Italy



(a) Video playback and statistics for the current segment

(b) Deepfake classification results for buffered segments

Figure 1: DOVE Player during on-demand MPEG-DASH streaming

Abstract

The spread of deepfake videos poses growing challenges for online content authenticity. Existing detection methods typically process complete videos offline or analyze frames only as they are displayed, making them poorly suited to video-on-demand playback. We present DOVE, a user-side system that exploits buffered segments in HTTP-based adaptive streaming to detect and temporally localize deepfake content ahead of playback. For each buffered segment, DOVE extracts and analyzes one or more frames and associates the detection result with that segment. This look-ahead approach can flag manipulated content before it reaches the viewer, localize short manipulations, and identify fully manipulated videos early, without requiring cooperation from the streaming platform. Our prototype demonstrates the feasibility of integrating user-controlled deepfake analysis into adaptive video streaming.

Keywords

Deepfake detection, Deepfake temporal localization, HTTP Adaptive Streaming, Video on demand, MPEG-DASH

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CoNEXT-SW '26, Utrecht, Netherlands
© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-2976-8/2026/12
https://doi.org/10.1145/3843179.3858046

ACM Reference Format:
Christian Coduri, Alessio Sacco, and Guido Marchetto. 2026. Deepfake Detection and Temporal Localization in HTTP Adaptive Streaming via Segment-Level Analysis. In *ACM CoNEXT 2026 Student Workshop (CoNEXT-SW '26), December 07–10, 2026, Utrecht, Netherlands.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3843179.3858046

1 Introduction

AI-based tools can generate increasingly realistic manipulated videos, including deepfakes, making online content more difficult to authenticate [2]. To address this challenge, content provenance standards such as C2PA record a video's origin, editing history, and potential AI involvement [4]. However, their effectiveness depends on adoption and remains subject to known security limitations [1].

This motivates the use of content-based deepfake detection. Yet, most detectors assume access to the complete, pre-recorded video, either to assign a single label [2] or to temporally localize manipulated intervals [6], making them unsuitable for streaming detection. Other near-real-time approaches use screen-capture-based methods to analyze frames as displayed. While well-suited to live video and videoconferencing, where future content is unavailable [3], in video-on-demand (VoD) they may expose viewers to manipulations that could have been detected earlier.

VoD services rely on adaptive streaming protocols such as HTTP Live Streaming (HLS) and MPEG-DASH, which divide each video representation into short segments. Clients request and buffer these segments over HTTP(S), adapting the selected quality to network conditions [8]. This buffering creates a look-ahead window in which upcoming segments can be analyzed before they are displayed.

To exploit this opportunity, we present DOVE, an MPEG-DASH-based **D**eepfake **O**n-demand **V**ideo **E**valuator, whose interface is shown in Figure 1. DOVE analyzes buffered segments on the fly, effectively "*analyzing the future*" of playback and enabling deepfake content to be flagged, paused, or skipped before reaching the viewer.

2 Design Overview

A conventional VoD architecture consists of a content-delivery infrastructure and a client-side player. Among the possible VoD standards, we consider MPEG-DASH [8], in which the player retrieves a Media Presentation Description (MPD), then requests and buffers initialization and media segments over HTTP(S), adapting the requested quality to network conditions. DOVE extends the playback pipeline (buffer, decode, render) with segment processing and a local Inspector, as shown in Figure 2.

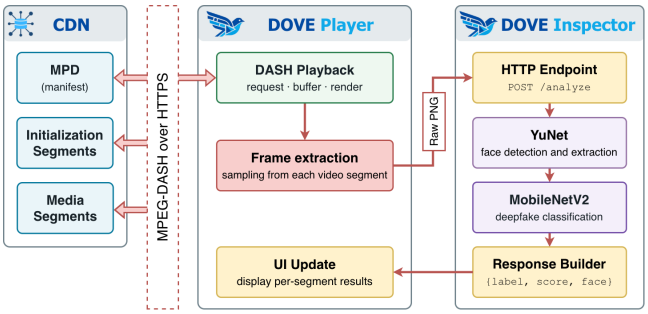


Figure 2: DOVE Architecture

DOVE Player is a web application that provides video playback and visualizes analysis results for all buffered segments. Alongside its user interface and DASH player, it contains hidden video and canvas elements for frame extraction. Each downloaded segment is copied and appended to the hidden player's media buffer, where it is decoded. The hidden player then applies a predefined frame-extraction strategy to determine the temporal positions of the frames to be analyzed within each segment. Based on this strategy, it seeks the selected positions and captures the corresponding frames through the hidden canvas.

DOVE Inspector is a local backend running on the viewer's device that receives the extracted frames, detects and extracts faces, classifies them as real or deepfake, and returns the results to the Player for visualization. Since the DOVE Inspector is intended to run locally in a CPU-only environment, both inference models must be computationally efficient. This prototype therefore uses *YuNet* [9] for face extraction and *MobileNetV2* [7], trained on *FaceForensics++* [5], for deepfake classification.

3 Discussion

By analyzing each segment as it enters the playback buffer, DOVE can provide *timely feedback* during playback, potentially warning the viewer before manipulated content is displayed, and *temporal localization*, since each segment classification maps directly to the corresponding interval in the complete video. This temporal granularity is especially useful for long videos containing short manipulations that sparse offline sampling may miss. In the example shown in Figure 1, analyzing one frame from each three-second segment can already be sufficient to distinguish 13 seconds of authentic content from the following 16 seconds of manipulated content. Conversely, when the entire video is manipulated, the initial segments may provide sufficient evidence for *early detection*, thereby avoiding unnecessary processing of the remainder of the stream.

Another advantage is *independence from the streaming platform*. Server or CDN-side analysis could process a video once and distribute labels for its segments. However, this requires providers to deploy a detector and disclose its results, which some platforms may have little incentive to do. DOVE instead lets interested viewers analyze the video locally, at the cost of each user analyzing the same content independently. To use DOVE with an existing service, the stream must play in DOVE's browser application; its MPD and segments must be publicly accessible, and decoded frames must be readable through a canvas. In Figure 1, GitHub serves only as a static, CDN-like source.

This design also has limitations. For detection, the current prototype samples only one frame per segment and may miss manipulations elsewhere within the same segment. For localization, manipulated intervals may not align with segment boundaries, so results can only approximate the start and end times of deepfake video parts. Segment duration, compression, and requested quality may further affect evidence for analysis, while running the Inspector locally consumes client resources and may affect playback.

4 Conclusion & Future Work

We introduced DOVE, a user-side system that analyzes buffered video segments ahead of playback to flag deepfakes, associate detections with segment intervals, and warn viewers before manipulated content is displayed. Future work will evaluate the trade-offs among detection accuracy, computational cost, energy use, and playback performance across detectors, segment configurations, and client devices. We will develop duration-aware strategies that adapt frame sampling to segment length, and investigate spatiotemporal and audio analysis to reduce sensitivity to frame selection. We also plan to extend this approach beyond deepfake detection, including harmful-content detection and parental controls.

References

[1] Enis Golaszewski et al. 2026. Verifying Provenance of Digital Media: Security Analysis of C2PA and its Implementation. *Cryptology ePrint Archive* (2026).
[2] Ammarah Hashmi et al. 2026. Understanding Audiovisual Deepfake Detection: Techniques, Challenges, Human Factors, and Perceptual Insights. *IEEE Computational Intelligence Magazine* 21, 2 (2026), 38–54.
[3] Vineet Mehta et al. 2021. Fakebuster: a deepfakes detection tool for video conferencing scenarios. In *Companion proceedings of the 26th international conference on intelligent user interfaces*. 61–63.
[4] Stefano Petrangeli et al. 2024. Integrating content authenticity with dash video streaming. In *Proceedings of the 15th ACM multimedia systems conference*. 492–498.
[5] Andreas Rossler et al. 2019. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1–11.
[6] Sanjay Saha et al. 2023. Undercover deepfakes: Detecting fake segments in videos. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 415–425.
[7] Mark Sandler et al. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *IEEE/CVF conference on computer vision and pattern recognition*. 4510–4520.
[8] Iraj Sodagar. 2011. The mpeg-dash standard for multimedia streaming over the internet. *IEEE multimedia* 18, 4 (2011), 62–67.
[9] Wei Wu et al. 2023. Yunet: A tiny millisecond-level face detector. *Machine Intelligence Research* 20, 5 (2023), 656–665.