ScienceDirect

Applied Soft Computing

Volume 174, April 2025, 112974

# Exploring coordinated motion patterns of facial landmarks for deepfake video detection

Yue Zhang [a], Run Niu [a], Xianlin Zhang [a], Siqi Chen [b], Mingdao Wang [c], Xueming Li [a]

Show more

Share Cite

https://doi.org/10.1016/j.asoc.2025.112974

Get rights and content

## Highlights

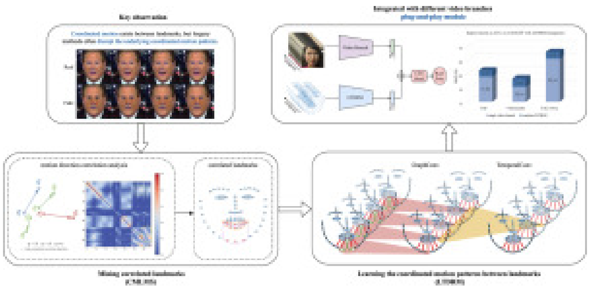
- Coordinated motion patterns between landmarks reveal potential abnormal motion.
- Motion direction correlation analysis uncovers related landmarks.
- Serves as a plug-and-play module, integrating with different types of video branches.

## Abstract

Due to the rich geometric and motion information they contain, recent studies indicate that facial landmark clues have significant potential for deepfake video detection. In this paper, we make a key observation that there exist coordinated motions among different facial landmarks for real individuals. While the forgery methods focus more on appearance realism, thus likely to disrupt the underlying coordinated motion patterns. Inspired by this observation, this paper explores how to leverage coordinated motion patterns among facial landmarks to enhance deepfake detection. First, we introduce a coordinated motion landmarks mining strategy (CMLMS), to effectively identify correlated landmarks. Utilizing these correlations, we propose a landmark temporal dynamic relation module (LTDRM), which focuses on the coordinated motion patterns between landmarks while extracting

their spatiotemporal features. Specifically, LTDRM constructs an adjacency matrix based on the correlated landmarks and uses graph convolution to selectively aggregate information between correlated landmarks. Additionally, LTDRM is a plug-and-play module and can boost the performance of existing deepfake detection methods with minimal computational overhead. Experimental results validate the effectiveness and generalizability of our method.

## Graphical abstract



Download: Download high-res image (224KB)

Download: Download full-size image

## Introduction

With the rapid development of deep generative models [1], [2], video editing and synthesis techniques [3], [4], [5] have continually improved, making the generated deepfake videos more realistic. Some malicious actors exploit deepfake technology to manipulate videos maliciously, disseminating false information and causing negative societal impacts [6]. As deep learning continues to play a crucial role in detecting manipulated content across various domains, from text correction [7] to multimedia forensics, researchers have developed numerous effective deepfake detection methods [8], [9]. However, deepfake detection remains an open challenge.

One promising direction in this field involves the use of facial landmarks, which provide rich geometric and motion information. Existing landmark-based methods [10], [11] have demonstrated the significant role of facial landmarks in identifying abnormal facial feature positions in forged images and detecting abnormal motion patterns in forged videos. These methods often have advantages such as low computational cost and good robustness to post-processing operations.

From a biomechanical standpoint, facial landmarks are driven by the movements of facial muscles, which produce coordinated motion patterns among various facial landmarks in real individuals. However, due to the inherent limitations of forgery methods, these coordinated motion patterns are often disrupted, leading to abnormal motion relations between landmarks. For instance, as shown in Fig. 1(a), the landmarks on the chin and lower lip exhibit strong motion correlation in real video. When the landmark on the chin moves downward, the landmark on the lower lip also exhibits a significant downward displacement. However, in forged video, although the chin landmarks move downwards similarly to real video, the displacement of the lower lip landmarks is minimal, resulting in an unreasonable reduction in motion correlation between these landmarks.

Inspired by the above observation, this paper explores the use of coordinated motion patterns among facial landmarks for deepfake detection. Fig. 1 illustrates our core idea. Taking LRNet [11] as an example, previous landmark-based methods often flatten landmark coordinates into feature vectors indiscriminately and overlook the relations among landmarks, as shown in Fig. 1(b). In contrast, we aggregate local motion information among correlated landmarks, explicitly model the relations between them, and use anomalies in their relative movements as a clue for detecting forged videos (Fig. 1(c)). Specifically, to identify which landmarks are correlated, we propose a coordinated motion landmarks mining strategy (CMLMS) based on motion direction correlation. The basic idea is that correlated landmarks typically exhibit synchronous or inverse movements between frames (e.g., the upper and lower lips often move oppositely during speech, whereas the lower lip and chin move in unison). Given that facial landmark movement is a mixture of both head-following motion and autonomous motion, our approach decomposes these motion types and mitigates the impact of head-following motion to better discern the correlated landmarks associated with autonomous motion (e.g., mouth, eyes, chin). Further, we propose a landmark temporal dynamic relation module (LTDRM). LTDRM can aggregate motion information among correlated landmarks using graph convolution and temporal convolution. Guided by CMLMS, LTDRM not only captures temporal inconsistencies between frames but also effectively models the coordinated motion patterns among landmarks.

LTDRM can serve as a plug-and-play component and flexibly integrate with off-the-shelf deepfake detection models. By integrating LTDRM with a video branch, complementary information from both data sources can be leveraged, resulting in more comprehensive multimodal spatio-temporal features. Video data provides rich contextual information and texture details, while facial landmark data offers precise geometric and motion descriptions. Combining these two types of information allows the method to more accurately capture forged evidence in deepfake videos.

In our experiments, we integrate LTDRM with various video-based methods, including CNN-based, Mamba-based, and Transformer-based, and assess their performance across a range of challenging scenarios. Compared with the original models, consistent improvements are achieved, with only a slight increase in FLOPs. Our contributions are summarized as follows:

- We make a key observation that there exist coordinated motions among different facial landmarks for real individuals. Building on this insight, we further explore using coordinated motion patterns among facial landmarks for deepfake detection.
- We propose a coordinated motion landmarks mining strategy (CMLMS) and a landmark temporal dynamic relation module (LTDRM). They explicitly model the coordinated motion patterns among facial landmarks and help uncover deepfake clues.
- Experiments demonstrate the effectiveness of LTDRM. When integrated with existing video-based methods of different architectures, including CNN-based, Mamba-based, and Transformer-based, it consistently delivers performance improvements with only a slight increase in FLOPs.

The structure of this paper is organized as follows: Section 2 reviews the related research on video and landmark-based deepfake detection methods. Section 3 provides a detailed explanation of the proposed method. Section 4 describes the datasets and experimental setup, and presents the results and analysis of various comparative and ablation experiments. Finally, Section 5 concludes the paper.

## Section snippets

### Video-based deepfake detection

Since most existing deepfake videos are generated on a frame-by-frame basis, there are often inconsistencies in lighting, texture, facial movement, and other information between different frames in deepfake videos. Video-level deepfake detection methods utilize these temporal inconsistencies compared to image-level methods. Early works often adopted architectures that combined CNN and RNN. Sabir et al. [12] first used CNN to extract facial features from each frame of the video and then utilized …

### Method

In this section, we first introduce the details of the proposed coordinated motion landmarks mining strategy (CMLMS) and the landmark temporal dynamic relation module (LTDRM). Then we describe how to integrate LTDRM with the video branch. The overall architecture of our model is shown in Fig. 2.

…

### Datasets

We evaluate our method on three widely used benchmarks: FaceForensics++ [27], Celeb-DF [29] and DFDC [30].

- FaceForensics++ (FF++) is the most commonly used benchmark dataset for deepfake detection, containing 1000 real videos and 4000 fake videos. Additionally, FaceForensics++ includes videos with three different compression levels. In our experiments, unless otherwise specified, we use the HQ version and follow the official dataset split. …
- Celeb-DF comprises 590 real videos and 5639 fake videos, …

…

### Conclusion

This paper is the first to explore using coordinated motion patterns among facial landmarks to enhance deepfake detection. We propose a coordinated motion landmarks mining strategy and a landmark temporal dynamic relation module. They can model the coordinated motion patterns among facial landmarks and help uncover deepfake clues. Experimental results demonstrate that integrating LTDRM into different video-based detection architectures consistently improves generalization, with AUC gains of …

## CRediT authorship contribution statement

**Yue Zhang:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Run Niu:** Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Xianlin Zhang:** Resources, Investigation. **Siqi Chen:** Resources, Investigation. **Mingdao Wang:** Resources, Investigation. **Xueming Li:** Resources, Investigation. …

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. …

## Acknowledgments

This work was supported by National Key R&D Program “Cultural Technology and Modern Service Industry” Key Special Project, China (#2021YFF0901700). …



## References (56)

R. Zhang *et al.*
Generalized face forgery detection with self-supervised face geometry information analysis network
Appl. Soft Comput. (2024)

M. Nawaz *et al.*
A deep learning model for FaceSwap and face-reenactment deepfakes detection
Appl. Soft Comput. (2024)

H. Ilyas *et al.*
AVFakeNet: A unified end-to-end dense swin transformer deep learning model for audio-visual deepfakes detection
Appl. Soft Comput. (2023)

Y. Li *et al.*

In ictu oculi: Exposing ai created fake videos by detecting eye blinking

I. Goodfellow *et al.*
Generative adversarial nets
Adv. Neural Inf. Process. Syst. (2014)

R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models,...

Y. Xu, Y. Yin, L. Jiang, Q. Wu, C. Zheng, C.C. Loy, B. Dai, W. Wu, Transeditor: Transformer-based dual-space gan for...

J. Sun, X. Wang, Y. Zhang, X. Li, Q. Zhang, Y. Liu, J. Wang, Fenerf: Face editing in neural radiance fields, in:...

G. Stanishevskii *et al.*
ImplicitDeepfake: Plausible face-swapping through implicit deepfake generation using NeRF and Gaussian splatting
(2024)

M. Boháček *et al.*
Protecting world leaders against deep fakes using facial, gestural, and vocal mannerisms
Proc. Natl. Acad. Sci. (2022)