electronics

MDPI

Article

Video Detection Method Based on Temporal and Spatial Foundations for Accurate Verification of Authenticity

Chin-Yuan Lin [1], Jen-Chun Lee [2], Shuenn-Jyi Wang [1], Chung-Shi Chiang [2] and Chao-Lung Chou [3,*]

[1] Department of Computer Science and Information Engineering, CCIT, National Defense University, Taoyuan 33551, Taiwan; ichilin929@gmail.com (C.-Y.L.); sjwang.jason@msa.hinet.net (S.-J.W.)
[2] Department of Telecommunication Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 81157, Taiwan; i923002@nkust.edu.tw (J.-C.L.); g950302@gmail.com (C.-S.C.)
[3] Department of Information Engineering and Computer Science, Feng Chia University, Taichung 40724, Taiwan
* Correspondence: chaolung.chou@gmail.com

Abstract: With the rapid development of deepfake technology, it is finding applications in virtual movie production and entertainment. However, its potential for malicious use, such as generating false information, fake news, or synthetic pornography, poses significant threats to national and social security. Various research disciplines are actively engaged in developing deepfake video detection technologies to mitigate the risks associated with malicious deepfake content. Therefore, the importance of deepfake video detection technology cannot be overemphasized. This study addresses the challenge posed by images in nonexistent datasets by analyzing deepfake video detection methods. Using temporal and spatial detection techniques and employing 68 facial landmarks for alignment and feature extraction, this research integrates the attention-guided data augmentation (AGDA) strategy to enhance generalization capabilities. The detection performance is evaluated on four datasets: UADFV, FaceForensics++, Celeb-DF, and DFDC, with superior results compared to alternative approaches. To evaluate the study's ability to accurately discriminate authenticity, detection experiments are conducted on both genuine and deepfake videos synthesized using the DeepFaceLab and FakeApp frameworks. The experimental results show better performance in detecting deepfake videos than other methods compared.

Keywords: deepfake video detection; security; temporal–spatial analysis; 68 facial landmarks; attention-guided data augmentation

Citation: Lin, C.-Y.; Lee, J.-C.; Wang, S.-J.; Chiang, C.-S.; Chou, C.-L. Video Detection Method Based on Temporal and Spatial Foundations for Accurate Verification of Authenticity. *Electronics* **2024**, *13*, 2132. https://doi.org/10.3390/electronics13112132

Academic Editors: Jun Feng, Changqing Luo and Mamoun Alazab

Received: 14 April 2024
Revised: 25 May 2024
Accepted: 28 May 2024
Published: 30 May 2024

1. Introduction

With the rapid development of generative modeling technologies, deepfake technology has been widely applied, creating numerous impressive virtual movie productions. For example, following the car accident death of Paul Walker, a star of the "Fast and Furious" series, the production team transferred his facial features onto his brother's face, allowing Paul Walker to "reappear" on screen, as shown in Figure 1 [1]. Additionally, there are lively and interesting video production apps such as ZAO and FaceApp, enabling anyone to easily create deepfake videos [2]. However, deepfake technology can also be used for malicious purposes, leading to fabricated pornography, false news, financial fraud, pranks, and cognitive warfare [3], causing serious trust crises and even affecting national and societal security.

To combat the cunning and varied deepfake technologies that disrupt order, various academic fields have also been actively developing detection methods for deepfake videos. The National Institute of Standards and Technology (NIST) initiated the Open Media Forensic Challenge (OpenMFC) platform in 2017, which is open to the public for researchers to engage in media forensic challenges and evaluate the capabilities of media forensic algorithms and systems, thus promoting the research and development of media

forensics [4]. In 2018, the Defense Advanced Research Projects Agency (DARPA) launched the Media Forensics (MediFor) program, developing tools to identify tampered images to combat the increasingly prevalent fake news [5]. In 2020, the well-known social media company Facebook organized a deepfake detection challenge, offering a prize of up to USD one million and providing a dataset of 100,000 deepfake images to encourage researchers to develop more accurate deepfake detection techniques [6].

Figure 1. A demonstrative illustration of deepfake videos from the "Fast and Furious" series [1].

To mitigate the damage caused by deepfake videos, the technology for detecting deepfakes has become critically important. Among these detection technologies, the most commonly used method is modeling through deep learning based on deepfake video datasets and inputting the dataset into a binary classifier to determine authenticity. However, as fake videos become increasingly realistic, the distinction between real and fake videos becomes increasingly difficult to discern, rendering this solution less effective [7], especially for images not present in the dataset. Some scholars have detected deepfakes through spatial artifacts within deepfake videos, achieving notable results but overlooking the temporal continuity of videos. Therefore, others have utilized temporal clues to address the issue of discontinuities in deepfakes without mastering the concept of time, yet they failed to fully discover spatial-related artifacts. Thus, scholars have attempted to detect deepfakes by capturing both spatial and temporal artifacts [8], ultimately recognizing deepfake videos not present in the dataset. However, due to the nature of deep learning training modes, while the accuracy of deepfake detection can be continually enhanced, the developed deepfake detection tools may neglect the issue of identifying real videos.

To address the limitations of existing deepfake video detection models, which heavily rely on deep learning trained on dataset-specific images, resulting in poor detection performance for images outside the dataset and failing to accurately pinpoint key facial features to uncover crucial clues in forged videos, this study develops a video detection method based on temporal and spatial foundations. This method can accurately distinguish between genuine and forged videos. The primary contributions are as follows:

1. Utilizing a cross-temporal and spatial detection method as the foundation, this study incorporates the attention-guided data augmentation (AGDA) mechanism to unearth more useful facial information. By employing the 68 facial landmarks method for facial marking and feature extraction, we have developed a deepfake video detection approach in this study.
2. For the commonly used datasets, including UADFV, FaceForensics++, Celeb-DF, and DFDC, the detection outcomes of this research, when compared with binary classifier detection methods, such as "FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception", and the "AltFreezing" temporal and spatial detection approach, all demonstrate superior effectiveness.

3. In detecting real videos and deepfake videos created using frameworks such as "DeepFaceLab" and "FakeApp", our study's detection results also show better performance compared to methods like "FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception", and "AltFreezing". This confirms our method as a viable approach for accurately identifying the authenticity of videos.

2. Related Work

Various styles of deepfake videos are currently being produced, among which face manipulation deepfake videos can be categorized into six types: entire face synthesis, identity swap, face morphing, attribute manipulation, expression swap (also known as face reenactment), and audio and text conversion. Entire face synthesis deepfakes refer to the use of generative adversarial networks (GANs) to create a nonexistent face [9]; identity swap deepfakes involve replacing the face of a person in a video with another person's face [10]; face morphing deepfakes mix two or more faces to create a new face [11]; attribute manipulation deepfakes alter certain facial features, such as age, gender, hairstyle, etc. [12]; expression swap deepfakes transfer the expression of one person to another's face [13]; audio and text conversion deepfakes generate corresponding videos from voice or text inputs [14]. This study will explore the detection technology for identity swap deepfakes, as illustrated in Figure 2.

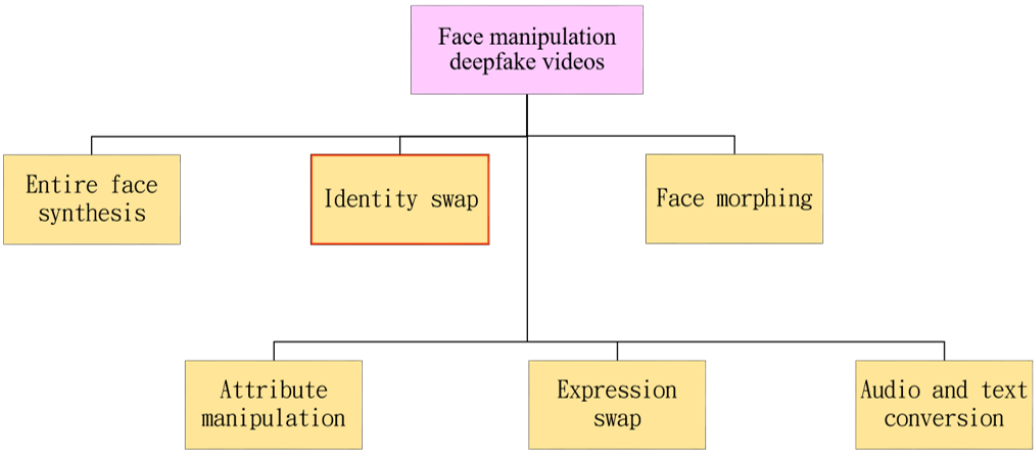

Figure 2. Illustration of face manipulation deepfake video categories.

Given the significant security threats posed by malicious deepfake videos, developing effective deepfake detection technologies is crucial. Initially, detection was conducted through biometric or visual methods. For instance, Yang et al. used 3D modeling and compared it with real human facial images using an SVM classifier to identify potential errors in deepfake head poses [15]; Li et al. employed deep neural networks (DNNs) to detect poorly processed blinking in deepfake videos [16]; Haliassos et al. leveraged the characteristics of mouth movements in real videos to conduct lip forensics in deepfakes [17]. Subsequently, most methods have utilized convolutional neural networks (CNNs) for modeling to extract complex features and then detect them through binary classifiers, among which the use of the Xception depthwise separable convolution network is popular [18]. For example, the Spanish software company "BBVA Next Technologies" developed a deepfake detection method named FakeVideoForensics in 2019, which was modeled on the Xception network and trained on the FaceForensics++ dataset [19], using a binary classifier for judgment [20]; Chen et al., in 2021, developed an improved Xception model for detecting faces generated by local GANs by enhancing the Xception network model to capture multi-level features through a feature pyramid, and creating a local GAN-generated facial dataset, LGGF, also classified by a binary classifier [21]; Tolosana et al., in 2022, developed the DeepFakes_FacialRegions detection method, applying Xception, a capsule network [22], and DSP-FWA [23] among three network models for training on four datasets [24]: UADFV [16],

FaceForensics++, Celeb-DF [25], and DFDC [26], and implemented judgment with a binary classifier. However, increasingly realistic deepfake videos present significant challenges to such binary classifier detection methods.

Given the rapid advancements in facial manipulation technology, traditional binary classifier detection methods encounter difficulties in detecting images absent from datasets. Scholars have initiated research into spatial and temporal aspects to remedy this challenge. For instance, Li et al. proposed Face X-ray in 2020, a method that identifies mixed boundaries in images to ascertain if an image comprises blends from two distinct sources, thereby facilitating the detection of facial image forgeries [27]. In 2020, Qian et al. introduced the Frequency in Face Forgery Network (F^3-Net), which utilizes two different but complementary frequency-aware clues: frequency-aware decomposed image components and local frequency statistics, to deeply mine forgery patterns [28]. Building on the Face X-ray methodology, Shiohara et al. unveiled the SBI detection approach in 2022, discovering enhanced training outcomes with deepfake videos created from a singular original image [29]. In 2023, Ju et al. introduced the GR-PSN network framework, composed of two subnetworks, GeometryNet and ReconstructNet, which learn surface normals from photometric stereo images and generate photometric images under distant illumination from various lighting directions and surface materials [30]. In 2021, Zheng et al. introduced the Comprehensive Temporal Convolution Network (FTCN), employing temporal convolution kernels to probe long-term temporal coherence independently of any external dataset [31]. Furthermore, in 2023, Wang et al. developed the AltFreezing detection method, leveraging a spatiotemporal model (3D ConvNet) that alternately freezes spatiotemporal network weights, as depicted in Figure 3, to detect both temporal and spatial forgeries [8].

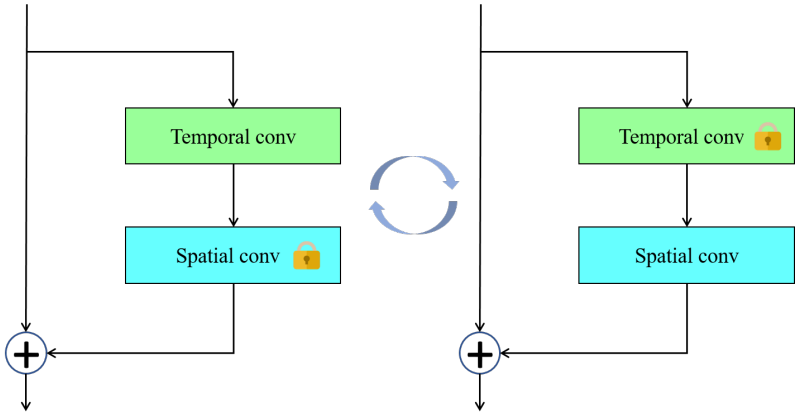

Figure 3. Illustration of AltFreezing spatiotemporal network weights.

Data augmentation mechanisms are strategies designed to increase the diversity and size of datasets, thereby enhancing the model's generalization capabilities. For example, in 2019, Li et al. synthesized a deepfake dataset by creating blurred images from original ones, reducing the need for extensive resources [23]. In 2022, Kong et al. utilized meaningful high-level semantic segmentation images to locate manipulated areas in order to detect forged faces in images [32]. In 2023, Luo et al. introduced the Critical Forgery Mining (CFM) framework, which can be flexibly integrated with various backbone networks to enhance their generalization and robustness of performance [33]. In 2021, Zhao et al. introduced a multi-attentional deepfake detection technique that focuses on local features in different regions, enhances texture feature blocks, aggregates low-level texture features with high-level semantics, and incorporates an attention-guided data augmentation (AGDA) strategy, as depicted in Figure 4. This approach directs the model's focus to important parts of the input data while disregarding irrelevant sections. It includes adjustments to the size of convolution kernels, the dilation rate of dilated convolutions, the standard deviation for Gaussian blur processing, the range of binary threshold values, the scope of magnification operations, scaling factors, noise ratio, and modes [7].

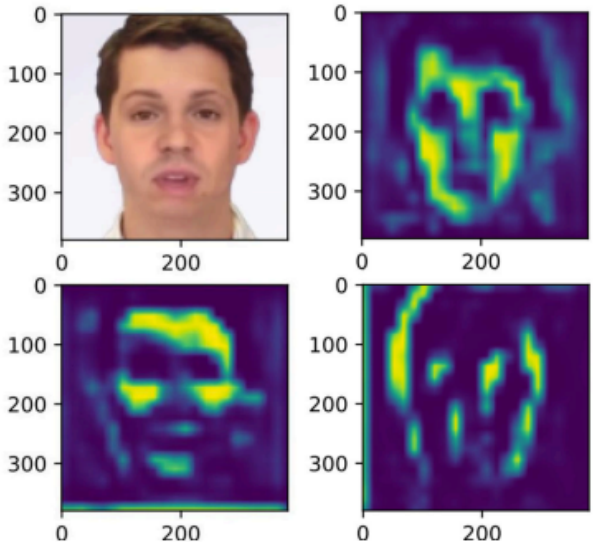

Figure 4. Illustration of the multi-attentional deepfake detection technique [7].

Baltrusaitis et al. launched the OpenFace tool in 2016, making it accessible for researchers in fields such as computer vision, machine learning, and affective computing for analytical applications. The tool's 68 facial landmarks method is capable of performing tasks like head pose estimation, facial landmark detection, and facial unit recognition [34]. In 2021, Li et al. introduced a deepfake detection method that leverages biometric features, using the 68 facial landmarks to differentiate between the central facial area and the entire facial area, thereby constructing facial vectors to determine the authenticity of videos [35]. In 2022, Tolosana et al. utilized the 68 facial landmarks to identify areas such as the eyes, nose, mouth, and rest (non-feature facial areas) for deepfake video detection, as shown in Figure 5. Their approach, which involved a fusion method, produced superior results [24].

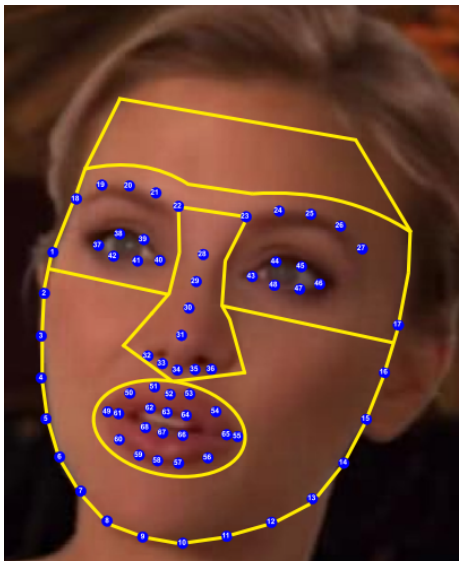

Figure 5. Illustration of the deepfake detection technique using 68 facial landmarks [24].

In their 2022 investigation of deepfake videos, Xu et al. [36] identified the most popular datasets as UADFV, FaceForensics++, Celeb-DF, and DFDC. In 2018, Li et al. developed the UADFV dataset to detect blinking frequency in deepfake videos. It comprises a total of 98 videos, including 49 real videos downloaded from YouTube and 49 deepfake videos created using the FakeApp tool [16]. In 2019, Rossler et al. collected videos from the internet, specifically from YouTube, and using four deepfake video creation meth-

ods—DeepFakes, Face2Face, FaceSwap [37], and NeuralTextures [38]—developed the FaceForensics++ dataset, which includes 1000 real videos and 4000 deepfake videos [19]. In 2020, Li et al. selected public videos of 59 celebrities from the online video platform YouTube and utilized the DeepFake video creation tool to develop the large and challenging Celeb-DF dataset, which contains 890 real videos and 5639 deepfake videos [25]. In 2020, Facebook Inc. organized a deepfake detection challenge, paid for the collection of private videos, and developed the DFDC dataset using eight deepfake video creation methods: DF-128, DF-256, MM/NN [39], NTH [13], FSGAN [40], StyleGAN [41], refinement, and audio swaps. This dataset includes 23,654 real videos and 104,500 deepfake videos [26].

3. Proposed Method

3.1. Deepfake Video Detection Method

Given the challenges that binary classifier deepfake detection methods face in distinguishing images not present in datasets, the inability of temporal-level detection methods to capture spatial-level mixed forgeries, and the limitations of spatial-level detection methods in detecting temporal discontinuities in forgeries, this study adopts the temporal and spatial AltFreezing detection method as its foundational approach. It integrates the 68 facial landmarks to align faces and extract features and introduces an attention-guided data augmentation (AGDA) strategy to highlight significant parts of the detection target, thereby enhancing the model's generalization capability. This approach results in the development of a video detection method that is grounded in both temporal and spatial dimensions and can accurately identify authenticity, as illustrated in Figure 6. The method has been tested on today's most popular datasets: UADFV, FaceForensics++, Celeb-DF, and DFDC. It is compared with binary classifier detection methods, such as "FakeVideoForensics", "DeepFakes_FacialRegions", and "Improved Xception", as well as with the "AltFreezing" temporal and spatial detection method. The method is compatible with operating systems like Windows and Linux and requires a computer equipped with a GPU.

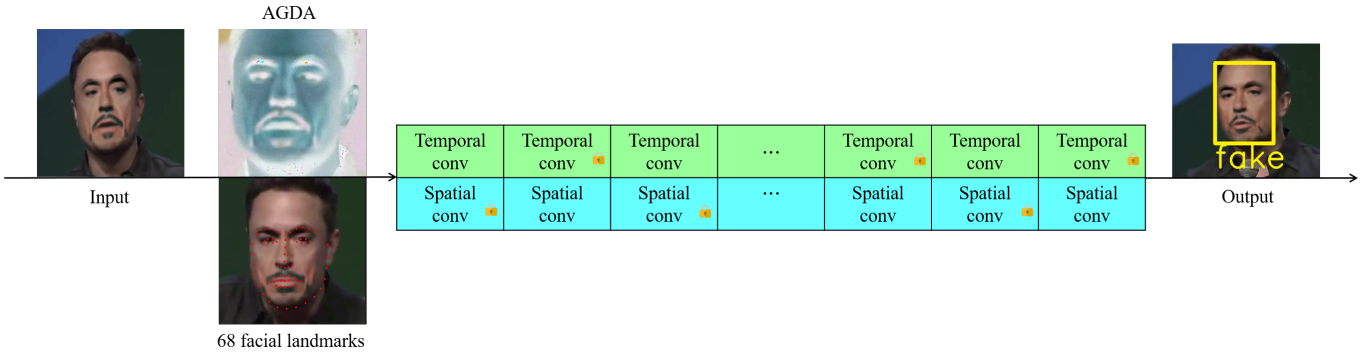

Figure 6. Diagram of the video detection method architecture based on temporal and spatial dimensions for accurate authenticity identification.

The method for calculating temporal weights and spatial weights in the detection model can be represented by the following two equations:

$$\theta_S \leftarrow \theta_S - \alpha \cdot \nabla_{\theta_S}(\text{Loss}(\text{Model}(\text{input}; \theta_S, \theta_T), \text{labels})) \quad (1)$$

$$\theta_T \leftarrow \theta_T - \alpha \cdot \nabla_{\theta_T}(\text{Loss}(\text{Model}(\text{input}; \theta_S, \theta_T), \text{labels})) \quad (2)$$

where θ_S represents the spatial weight and θ_T represents the temporal weight, both parameters being variables that need to be optimized during the learning process; α represents the learning rate, and the greater the learning rate, the larger the change in parameters; ∇ represents the gradient of the loss function with respect to the parameters, a key factor used for parameter updates in the gradient descent algorithm; Loss is the loss function, which is used to measure the difference between the model output and the actual labels;

Model(input; θ_S, θ_T) is the model function, which receives input data and parameters (θ_S and θ_T) and produces output, which is used to calculate the loss function; finally, input is the image input to the model; labels are the actual labels or data.

The gradient of the loss function for Equation (1) is calculated based on the current values of θ_S and θ_T, and the parameters θ_S are updated according to the learning rate α. Similarly, the gradient of the loss function for Equation (2) is calculated based on the current values of θ_S and θ_T, and the parameters of θ_T are updated according to the learning rate α. This approach employs the gradient descent method for dual parameter updates, minimizing the loss function to update both sets of model parameters θ_S and θ_T. The purpose of this design is to capture both spatial and temporal artifacts simultaneously by updating the weights associated with spatial and temporal dimensions, respectively. This compensates for the limitations of focusing solely on spatial weights and ignoring temporal coherence, or focusing solely on temporal weights and ignoring spatial discontinuities, making it more suitable for detecting faces in deepfake videos.

The network model that calculates weights through the interplay of time and space operates by utilizing two distinct perceptual frequencies for detection. For the detector, each image is unprecedented, not subject to the limitations observed in traditional binary classifier deepfake detection methods, which rely on dataset-specific images for deep learning and fail to recognize new images. Additionally, it resolves the issue of disparate levels of artifacts that arise when calculating weights separately in time and space. Furthermore, the incorporation of 68 facial landmarks and an attention-guided data augmentation (AGDA) strategy facilitates the localization of manipulated areas and the extraction of key clues from deepfake images, thereby enhancing detection accuracy.

To test the capability of this study's method in accurately identifying authenticity, detection was conducted on real videos and on deepfake videos created using "DeepFaceLab" (as shown in Figure 7) and "FakeApp" (as shown in Figure 8) tools [42]. These were then compared with methods such as "FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception", and "AltFreezing". The "DeepFaceLab" deepfake video creation tool, released by Iperov in 2019 on the GitHub software source code hosting platform, utilizes Google's TensorFlow open-source machine learning library. It is compatible with operating systems like Windows and Linux, with superior results achieved on computers equipped with GPUs. The "FakeApp" deepfake video creation tool, introduced by a user named Deepfakes in 2018 on the Reddit social networking site, also makes use of TensorFlow and artificial neural networks for its production process. It supports operating systems such as Windows and Linux and requires a GPU-equipped computer for optimal functionality. Additionally, versions suitable for iOS and Android mobile phones have been developed.

Figure 7. Illustration of DeepFakeLab's deepfake video production.

Figure 8. Illustration of FakeApp's deepfake video production.

3.2. Deepfake Video Detection Process

The detection process of this study's video detection method, which is based on temporal and spatial dimensions for accurate authenticity identification, is described as follows:

1. Configuration preparation: Specifies the paths for detection and output videos' and processes' video parameters, such as adjusting the video size (clip_size), batch size (batch_size), and image dimensions (imsize).
2. Face positioning: Utilizes the 68 facial landmarks method to extract key facial points from every frame of the video.
3. Alignment cutting: Performs size alignment and cropping as pre-processing for each frame's facial image.
4. Image processing: Enhances specific features in each frame's facial image by amplifying attention through attention-guided data augmentation (AGDA).
5. Input model: Inputs the pre-processed images into the network model, which is based on temporal and spatial dimensions.
6. Make predictions: Utilizes the loaded model to predict the authenticity of each frame and calculates the overall authenticity of the test video.
7. Result output: Compares and computes authenticity scores for each frame, producing the final detection outcome and exporting the detection result video.

4. Experimental Results and Analysis

4.1. Experimental Procedure

Utilizing equipment such as a computer with an Intel Core i7-10875H CPU, an NVIDIA GeForce RTX 2070 GPU, and 64 GB of RAM, this study employs a video detection framework based on temporal and spatial dimensions to accurately identify authenticity. Detection was carried out on four datasets: UADFV, FaceForensics++, Celeb-DF, and DFDC. Additionally, real videos and deepfake videos created with the "DeepFaceLab" and "FakeApp" tools were tested. In the detection results, a value of 1 represents a deepfake video, while 0 indicates a real video; the higher the score, the greater the likelihood of being identified as a deepfake video.

4.2. Deepfake Video Detection and Analysis

In regards to the detection of deepfake videos, this study, along with four other detection methods—"FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception",

and "AltFreezing"—was applied across four datasets: UADFV, FaceForensics++, Celeb-DF, and DFDC. The method developed in this study, which leverages temporal and spatial dimensions to accurately identify authenticity, demonstrated superior detection results compared to the methods of "FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception", and "AltFreezing". Furthermore, it exhibited better average performance, as detailed in Table 1.

Table 1. Analysis results of detecting existing deepfake datasets.

Methods	UADFV	FaceForensics++	Celeb-DF	DFDC	Avg
FakeVideoForensics	0.9546	0.8637	0.9559	0.9083	0.9206
DeepFakes_FacialRegions	0.2400	0.4000	0.2040	0.2000	0.2610
Improved Xception	0.5625	0.8605	0.6906	0.7755	0.7222
AltFreezing	0.9637	0.9494	0.7390	0.9029	0.8888
Ours	0.9813	0.9794	0.9787	0.9861	0.9814

Through this study's deepfake video detection framework, detection was conducted on real videos and deepfake videos created using two methods: "DeepFaceLab" and "FakeApp". Both real videos and deepfake videos produced by "DeepFaceLab" and "FakeApp" were successfully detected by the study's deepfake detection model, with authenticity accurately identified (as shown in Figure 9).

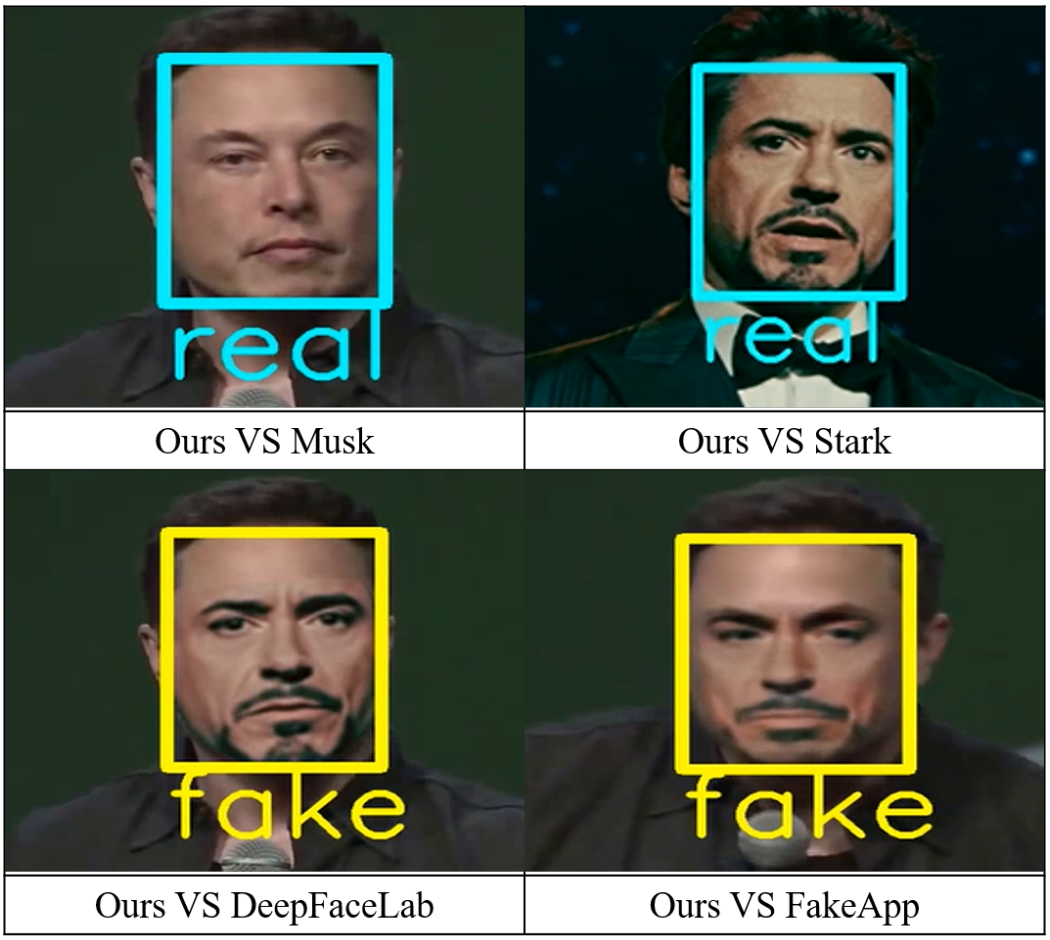

Figure 9. Detection results of this study's method.

Additionally, applying five detection methods—"FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception", "AltFreezing", and this study's approach—to

detect real videos and deepfake videos created using "DeepFaceLab" and "FakeApp", this study's method, based on temporal and spatial dimensions for accurately identifying authenticity, demonstrated superior performance in detecting both real and deepfake videos compared to the four other methods, namely "FakeVideoForensics", "DeepFakes_FacialRegions", "Improved Xception", and "AltFreezing" (as shown in Table 2).

Table 2. Analysis results of real and deepfake video detection.

Methods	Real	DeepFaceLab	FakeApp
FakeVideoForensics	0.8583	0.9654	0.9509
DeepFakes_FacialRegions	0.2105	0.2635	0.2278
Improved Xception	0.5839	0.4964	0.4599
AltFreezing	0.1188	0.9577	0.8966
Ours	0.0988	0.9727	0.9638

Compared to other methods, this study's detection approach, grounded in temporal and spatial dimensions, incorporates the 68 facial landmarks and introduces the attention-guided data augmentation strategy (AGDA). It is capable of recognizing images not present in datasets and accurately identifying authenticity, achieving optimal detection results (as illustrated in Table 3).

Table 3. Comparison table of detection methods.

Methods	Model Features	Comparison
FakeVideoForensics	Xception	Capable of detecting sequences in videos but struggles to identify images not present in datasets.
DeepFakes_FacialRegions	Xception, Capsule, DSP-FWA, 68 Facial Landmarks	Capable of detecting specific facial features but struggles to identify images not present in datasets.
Improved Xception	Improved Xception	The network model performs well but struggles to identify images not present in datasets.
AltFreezing	3D ConvNet	Based on temporal and spatial dimensions, it can recognize images not present in datasets, yet it does not prioritize the detection results of real videos.
Ours	3D ConvNet, 68 Facial Landmarks, AGDA	Based on temporal and spatial dimensions, incorporating the 68 facial landmarks and introducing the attention-guided data augmentation strategy (AGDA), it can recognize images not present in datasets and accurately identify authenticity.

5. Conclusions

This study develops a detection model suitable for deepfake videos that is grounded in temporal and spatial dimensions by incorporating the 68 facial landmarks method and the attention-guided data augmentation (AGDA) mechanism. This method, based on temporal and spatial dimensions for accurately identifying authenticity, was applied to four datasets: UADFV, FaceForensics++, Celeb-DF, and DFDC. Compared to binary classifier detection methods like "FakeVideoForensics", "DeepFakes_FacialRegions", and "Improved Xception", as well as the "AltFreezing" temporal and spatial detection approach, it demonstrates superior effectiveness. Additionally, when detecting real videos and deepfake videos produced using the "DeepFaceLab" and "FakeApp" frameworks, it accurately distinguishes authenticity with better performance than the four aforementioned methods. This approach addresses the various security issues posed by contemporary deepfake videos.

Author Contributions: Investigation, C.-Y.L.; methodology, C.-Y.L.; validation, C.-Y.L. and C.-L.C.; writing—original draft preparation, C.-Y.L.; writing—review and supervision, J.-C.L., S.-J.W., C.-S.C., and C.-L.C.; editing, C.-Y.L. and C.-L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Science and Technology Council of Taiwan, grant No. NSTC 112-2221-E-606-009-MY2.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yadav, D.; Salmani, S. Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network. In Proceedings of the 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 15–17 May 2019; pp. 852–857.
2. de Seta, G. Huanlian, or changing faces: Deepfakes on Chinese digital media platforms. *Int. J. Res. New Media Technol.* **2021**, *27*, 935–953. [CrossRef]
3. Raja, K.; Ferrara, M.; Batskos, I.; Barrero, M.G.; Scherhag, U.; Venkatesh, S.K.; Singh, J.M.; Ramachandra, R.; Rathgeb, C.; Busch, C. Morphing Attack Detection-Database, Evaluation Platform, and Benchmarking. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 4336–4351. [CrossRef]
4. Tyagi, S.; Yadav, D. A detailed analysis of image and video forgery detection techniques. *Vis. Comput.* **2023**, *39*, 813–833. [CrossRef]
5. Jain, R. Detection and Provenance: A Solution to Deepfakes? *J. Stud. Res.* **2023**, *12*. [CrossRef]
6. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C.C. The DeepFake Detection Challenge (DFDC) Dataset. *arXiv* **2020**, arXiv:2006.07397.
7. Zhao, H.; Zhou, W.; Chen, D.; Wei, T.; Zhang, W.; Yu, N. Multi-attentional Deepfake Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 2185–2194.
8. Wang, Z.; Bao, J.; Zhou, W.; Wang, W.; Li, H. AltFreezing for More General Video Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 4129–4138.
9. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8107–8116.
10. Jiang, L.; Li, R.; Wu, W.; Qian, C.; Loy, C.C. DeeperForensics-1.0: A Large-Scale Dataset for Real-World Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 2889–2898.
11. Neubert, T.; Makrushin, A.; Hildebrandt, M.; Kraetzer, C.; Dittmann, J. Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biom.* **2018**, *7*, 325–332. [CrossRef]
12. Dang, H.; Liu, F.; Stehouwer, J.; Liu, X.; Jain, A. On the Detection of Digital Face Manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5780–5789.
13. Zakharov, E.; Shysheya, A.; Burkov, E.; Lempitsky, V. Few-Shot Adversarial Learning of Realistic Neural Talking Head Models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9459–9468.
14. Thies, J.; Elgharib, M.; Tewari, A.; Theobalt, C.; Nießner, M. Neural Voice Puppetry: Audio-driven Facial Reenactment. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 716–731.
15. Yang, X.; Li, Y.; Lyu, S. Exposing Deep Fakes Using Inconsistent Head Poses. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8261–8265.
16. Li, Y.; Chang, M.C.; Lyu, S. In ictu oculi: Exposing AI created fake videos by detecting eye blinking. In Proceedings of the IEEE International Workshop on Information Forensics and Security(WIFS), Hong Kong, China, 11–13 December 2018; pp. 1–7.
17. Haliassos, A.; Vougioukas, K.; Petridis, S.; Pantic, M. Lips don't lie: A Generalisable and Robust Approach to Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 5037–5047.
18. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
19. Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C. FaceForensics++: Learning to Detect Manipulated Facial Images. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1–11.
20. Munoz, A.; Hernandez, M. FakeVideoForensics. Available online: https://github.com/afuentesf/fakeVideoForensics (accessed on 13 March 2022).
21. Chen, B.; Ju, X.; Xiao, B.; Ding, W.; Zheng, Y.; de Albuquerque, V.H.C. Locally GAN-generated face detection based on an improved Xception. *Inf. Sci.* **2021**, *572*, 16–28. [CrossRef]
22. Nguyen, H.; Yamagishi, J.; Echizen, I. Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2307–2311.

23. Li, Y.; Lyu, S. Exposing DeepFake Videos by Detecting Face Warping Artifacts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seoul, Republic of Korea, 27 October–2 November 2019.
24. Tolosana, R.; Romero-Tapiador, S.; Vera-Rodriguez, R.; Gonzalez-Sosa, E.; Fierrez, J. DeepFakes detection across generations: Analysis of facial regions, fusion, and performance evaluation. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104673. [CrossRef]
25. Li, Y.; Yang, X.; Sun, P.; Qi, H.; Lyu, S. Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3207–3216.
26. Trabelsi, A.; Pic, M.M.; Dugelay, J.L. Improving Deepfake Detection by Mixing Top Solutions of the DFDC. In Proceedings of the European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 29 August–2 September 2022; pp. 643–647.
27. Li, L.; Bao, J.; Zhang, T.; Yang, H.; Chen, D.; Wen, F.; Guo, B. Face X-ray for More General Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5001–5010.
28. Qian, Y.; Yin, G.; Sheng, L.; Chen, Z.; Shao, J. Thinking in Frequency: Face Forgery Detection by Mining Frequency-aware Clues. In Proceedings of the Computer Vision–ECCV 2020, Glasgow, UK, 23–28 August 2020; pp. 86–103.
29. Shiohara, K.; Yamasaki, T. Detecting Deepfakes with Self-Blended Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 18720–18729.
30. Ju, Y.; Shi, B.; Chen, Y.; Zhou, H.; Dong, J.; Lam, K.-M. GR-PSN: Learning to Estimate Surface Normal and Reconstruct Photometric Stereo Images. *IEEE Trans. Vis. Comput. Graph.* **2023**, 1–16. [CrossRef] [PubMed]
31. Zheng, Y.; Bao, J.; Chen, D.; Zeng, M.; Wen, F. Exploring Temporal Coherence for More General Video Face Forgery Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 15044–15054.
32. Kong, C.; Chen, B.; Li, H.; Wang, S.; Rocha, A.; Kwong, S. Detect and Locate: Exposing Face Manipulation by Semantic- and Noise-Level Telltales. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 1741–1756. [CrossRef]
33. Lou, A.; Kong, C.; Huang, J.; Hu, Y.; Kang, X.; Kot, A.C. Beyond the Prior Forgery Knowledge: Mining Critical Clues for General Face Forgery Detection. *IEEE Trans. Inf. Forensics Secur.* **2023**, *19*, 1168–1182.
34. Baltrušaitis, T.; Robinson, P.; Morency, L.-P. Openface: An open source facial behavior analysis toolkit. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10.
35. Li, M.; Liu, B.; Hu, Y.; Zhang, L.; Wang, S. Deepfake Detection Using Robust Spatial and Temporal Features from Facial Landmarks. In Proceedings of the IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, 6–7 May 2021; pp. 1–6.
36. Xu, F. J.; Wang, R.; Huang, Y.; Guo, Q.; Ma, L.; Liu, Y. Countering Malicious DeepFakes: Survey, Battleground, and Horizon. *Int. J. Comput. Vis.* **2022**, *130*, 1678–1734.
37. Faceswap. Available online: https://github.com/deepfakes/faceswap (accessed on 27 May 2023).
38. Thies, J.; Zollhofer, M.; Nießner, M. Deferred Neural Rendering: Image Synthesis using Neural Textures. *ACM Trans. Graph.* **2019**, *38*, 1–12. [CrossRef]
39. Huang, D.; Torre, F.D.L. Facial Action Transfer with Personalized Bilinear Regression. In Proceedings of the European Conference on Computer Vision (ECCV), Florence, Italy, 7–13 October 2012; pp. 144–158.
40. Nirkin, Y.; Keller, Y.; Hassner, T. FSGAN: Subject Agnostic Face Swapping and Reenactment. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7184–7193.
41. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4401–4410.
42. Guera, D.; Delp, E.J. Deepfake Video Detection Using Recurrent Neural Networks. In Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 27–30 November 2018.