IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 15, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

DETECTION OF DEEPFAKE VIDEOS USING LONG DISTANCE ATTENTION

[1]Mr.V. Rajashekhar,[2]P. Srujana,[3] P. Aishwarya,[3] N.Anuja
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT

Recent years have seen a significant advancement in deepfake methods, which has led to the creation of very convincing video content and serious security risks via facial video forgeries. Additionally, it is even more difficult and urgent to discover such phony films. The majority of detection techniques in use today approach the issue as a standard binary classification problem. Since the distinctions between fake and genuine faces are quite small, this work treats the topic as a specific fine-grained classification problem. A number of typical artifacts, such as generative faults in the spatial domain and inter-frame inconsistencies in the temporal domain, are shown to have been left behind by the majority of face forgeries techniques now in use. Additionally, a two-component spatial-temporal model is put out to capture, in a global viewpoint, the temporal and spatial forgery traces, respectively. A revolutionary long-distance attention mechanism is used in the design of the two components. Artifacts may be captured in a single frame using the one spatial domain component and in subsequent frames using the other temporal domain component. They produce patches that represent attention maps. With a more expansive perspective, the attention approach helps to better compile global data and extract local statistical data. Ultimately, similar to previous fine-grained categorization techniques, the attention maps are used to direct the network to concentrate on important facial features. The state-of-the-art performance of the suggested technique is shown by the experimental results on several public datasets, and the proposed long distance attention method can successfully capture crucial elements for face forging.

1. INTRODUCTION

The purpose of the deepfake films is to swap out one person's face with another. With the development of generative models [1]–[4], deepfake films are becoming more lifelike. In the meanwhile, anybody may create very convincing faked films thanks to the development of many face forgery applications [5]–[7]. The Internet is now overflowing with deepfake videos. Such technology is readily utilized in the internet age to disseminate hate and misinformation, which is very detrimental to society. Thus, researchers are interested in the high quality deepfake films that are indiscernible to the human sight. There is an immediate need for an efficient detection approach.

Fig. 1 illustrates the typical procedure for creating deepfake films. The video is first split up into frames, and each frame's face is identified and cropped. Next,

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

a generative model is used to transform the original face into the target face, which is then spliced into the appropriate frame. Ultimately, every frame is serialized in order to create the deepfake video. Two types of faults are typically introduced in these procedures. The flawed generation model introduces visual aberrations in the spatial domain during the process of creating forged faces. The absence of global limits results in frame discrepancies when assembling frame sequences into videos.

Based on the faults in the spatial domain, several detection techniques have been presented [8]–[10]. Because generative models lack global restrictions during the fake face generation process, certain approaches exploit the flaws in face semantics in deepfake films. This results in the introduction of some anomalous face components and mismatched features in the face from a global viewpoint. As an example, consider asymmetric faces [11], asymmetric facial parts [10], and colored eyes [8]. But depending only on these semantics is risky. Performance will drop dramatically as the deepfake movies lose the particular semantic flaws on which the approach relies.

Additionally, several "deep" techniques exist [9], [12], and [13], which aim to uncover spatial flaws based on the characteristics of the deepfake generators. Convolutional networks, on the other hand, often extract features from the picture content rather than the forgery traces, which are rather weak in the spatial domain when compared to image contents [14]. Therefore, relying just on deep learning to detect bogus information is not particularly successful [15].

There will be discrepancies in the time domain since the deepfake video is synthesized frame by frame and there is no exact restriction between the frame sequences. Certain techniques take use of these temporal domain flaws. In [16], the eye movements are taken advantage of. The human blink frequency in the film is used by Li et al. [17] to identify deepfake movies. In the temporal domain, the movement of the lip [18] and the heart rate [19] are also used as the foundation for distinguishing between real and deepfake films. The genuine face and the synthetic face's optical flows and movement patterns are categorized in [20] and [21], respectively.

The deep fake detection issue is treated as a standard binary classification problem by all of the previously discussed approaches. Nevertheless, as counterfeit goods get more lifelike, the distinctions between the genuine and imitation will become more minute and specific, which will hinder the effectiveness of global feature-based vanilla solutions [22].

Similar issues have been researched in the fine-grained categorization sector. Classifying relatively similar categories, such bird species, vehicle models, and aircraft kinds, is the goal of fine-grained categorization [23]. In [22], the deep fake detection problem is framed as a fine-grained classification challenge since both

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

deep fake detection and fine-grained classification include learning subtle and discriminative characteristics. Additionally, a network is able to concentrate on the subtle but crucial areas by using a convolutional attention module with a 1 x 1 size.

But integrating global semantics is as crucial as concentrating on local domains. Due to the fact that many flaws seem normal when seen locally or isolated, but aberrant when viewed globally. Uncoordinated head postures [24], mismatched head motions and facial expressions [25], and mismatched eye features [26] are a few examples. These types of long-distance abnormalities occur between various facial areas. Stated differently, the global semantics should guide the determination of the local regions of emphasis [27], and it is crucial to characterize long-distance relationships in both the spatial and temporal domains. However, it is not directly related to the convolutional attention mechanism—particularly at small kernel sizes. Global pooling might be an option for putting together global information, however this procedure would average the weak forgery hints, making it harder to discern between them [22].

2. LITERATURE SURVEY

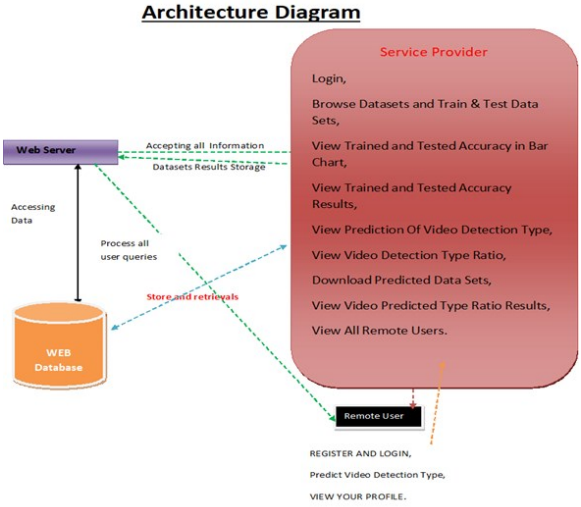
In our novel approach to generative model estimation using adversarial nets, we simultaneously train two models: a discriminative model D that calculates the likelihood that a sample originated from the training data instead of G, and a generative model G that represents the data distribution. G's training process aims to increase the likelihood that D will make a mistake. This framework is equivalent to a two-player minimax game. There is a unique solution in the space of arbitrary functions G and D, where G recovers the distribution of the training data and D is always equal to 1/2. Backpropagation may be used to train the whole system when G and D are specified by multilayer perceptrons. Neither Markov chains nor unrolled approximation inference networks are required for sample creation or training. Experiments show the framework's potential by evaluating the produced samples both qualitatively and quantitatively.

We provide a novel approach to generative adversarial network training. The important concept is to gradually develop the discriminator and generator: we start at a low resolution and add successive layers as training goes on, each of which models progressively finer features. This significantly stabilizes and accelerates the training process, enabling us to generate pictures of previously unheard-of quality, such as 1024^2 CelebA images. Additionally, we provide a straightforward method to boost the variety of pictures produced and get a record inception score of 8.80 in unsupervised CIFAR10. We also provide an overview of many implementation features that are crucial to deterring harmful rivalry between the discriminator and generator. Lastly, we propose a novel measure for assessing the

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

quality and diversity of images produced by GANs. To further contribute, we build an improved version of the CelebA dataset.

3. SYSTEM ARCHITECTURE



4. EXISTING SYSTEM

The performance on generic picture classification tasks has increased dramatically in the last several years. Since Alexnet's phenomenal debut in Imagenet [31], deep learning-based techniques have almost completely taken over the Imagenet competition [32]. However, there are still many obstacles to overcome in the area of fine-grained object identification [33]–[37]. The primary explanation for this is because, from both an apparent and global point of view, the two things are almost identical. Consequently, one of the main themes of fine-grained identification is learning to distinguish the minute variations in certain important components.

Previous efforts [38], [39] make effective use of the human-annotated bounding box of important components. However, the drawback is that it requires costly manual annotation, and depending entirely on the annotator's cognitive ability, the manual annotation's position may not always be the ideal region for distinction [40], [41].

As concentrating on increasingly discriminative local regions is a crucial step in fine-grained categorization [42], several weakly supervised learning techniques have been put forward [23], [40], and [43]. To identify the critical components for detection, the majority of them use various convolutional attention methods. Recurrent attention convolutional neural networks (RA-CNNs) are used by Fu et al. [43] to teach discriminative area attention. A channel-wise attention strategy is proposed by Hu et al. [44] to simulate interdependencies between channels. More granular features may be learnt by using a multi-attention convolutional neural network, as shown in [40]. A poorly supervised data augmentation network using attention dropping and cropping is proposed by Hu et al. [23].

Fine-grained classification and deepfake detection both aim to categorize very similar objects. As a result, we draw on our knowledge in this area and use the attention maps created using long-range information to direct the networks' attention to crucial areas.

Disadvantages

- The system has not been put into place.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

- The spatial attention model was not made to capture artifacts in the spatial domain with only one frame. System performance degraded due to poor spatial-temporal model.

5. PROPOSED SYSTEM

Presented here is an innovative long-range attention mechanism that has the potential to guide by integrating information from across the world, and the fine-grained classification field's experience is introduced. It shows that the attention mechanism with a longer attention span is better at acquiring global information and concentrating local regions. In addition, attention maps may be made using the non-convolution module.

• An integrated spatial-temporal model is created to include the defects in both the spatial and temporal domains. The model constructs a multi-level semantic guiding system using deepfake movie characteristics and long-distance attention as its main mechanism. The results of the experiments show that it is quite efficient.

Advantages

- A short description of the suggested model follows an explanation of the rationale for using long-distance attention in the proposed system. In the face of global forgeries, local areas are constantly at odds with one another due to the deepfake generation model's lack of exact global constraints.
- Since deepfake films are created frame by frame, there are imperfections in each frame as well as inconsistencies (such unsmooth lip movement) across sequences of frames. A spatial-temporal model is suggested to capture these faults; the model consists of two parts, one for spatial defects and the other for temporal defects. A unique long-distance attention mechanism is built into every component, allowing for the assembly of global information to emphasize local locations.

6. IMPLEMENTATION

Modules Description

Service Provider

To access this module, the Service Provider has to provide a valid username and password. Once he's logged in, he'll have access to certain features, including the ability to browse datasets and perform tests and training on them. Take a look at the video's projected type ratio, see the results of the trained and tested accuracy in a bar chart, see all the users from afar, and extract the predicted data sets.

View and Authorize Users

All users who have registered for this module may be seen by the administrator. Users' names, email addresses, and physical addresses are viewable to the administrator, who may also authorize users.

Remote User

In this module, you will find n users. The user must register before they may begin any activity. Upon registration, the user's details are stored in the database. He will be

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

prompted to enter his approved username and password after he successfully registers. When a user logs in, they will be able to access features like "View Your Profile," "Predict Video Detection Type," and "Register and Login."

7. CONCLUSION

Since the distinction between fake and genuine faces is quite slight, we approach the detection of deepfake video from the standpoint of fine-grained classification in this study. A spatial temporal attention model is created to make the network concentrate on the important local areas based on the generation faults of the deep fake generation model in the spatial domain and the inconsistencies in the time domain. Additionally, a brand-new long-distance attention technique is put forward to capture deep fake's global semantic inconsistency. We separate the image into tiny patches and recalculate the relative relevance of each patch in order to more effectively extract the texture and statistical information from the picture. Numerous tests have been conducted to prove that our approach achieves state-of-the-art performance and that the long-distance attention mechanism that we have developed is capable of producing global guidance. In addition to the long-distance attention mechanism and the spatial-temporal model, we believe that this paper's primary contribution is its confirmation that, in addition to being crucial to concentrating on significant locations, integrating global semantics is also essential. This is an important observation that may help advance the state-of-the-art models.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems, vol. 27, Montreal, CANADA, 2014.

[2] D. P. Kingma and M.Welling, “Auto-Encoding Variational Bayes,” 2014.

[3] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” in International Conference on Learning Representations, Vancouver, Canada, 2018.

[4] Q. Duan and L. Zhang, “Look More Into Occlusion: Realistic Face Frontalization and Recognition With BoostGAN,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 214–228, 2021.

[5] “deepfake,” http://www.github.com/deepfakes/ Accessed September 18, 2019.

[6] “fakeapp,” http://www.fakeapp.com/ Accessed February 20, 2020.

[7] “faceswap,” http://www.github.com/MarekKowalski/ Accessed September 30, 2019.

[8] F. Matern, C. Riess, and M. Stamminger, “Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations,” in IEEE Winter Applications of Computer Vision Workshops, Waikoloa, USA, 2019, pp. 83–92.

[9] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “Mesonet: a Compact Facial

ISSN 2347–3657
Volume 15, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v15.i3.pp846-852

Video Forgery Detection Network," in IEEE International Workshop on Information Forensics and Security, Hong Kong, China, 2018, pp. 1–7.

[10] X. Yang, Y. Li, H. Qi, and S. Lyu, "Exposing GAN-Synthesized Faces Using Landmark Locations," in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Paris, France, 2019, p. 113–118.

[11] D.-T. Dang-Nguyen, G. Boato, and F. G. De Natale, "Discrimination between computer generated and natural human faces based on asymmetry information," in Proceedings of the 20th European Signal Processing Conference, Bucharest, Romania, 2012, pp. 1234–1238.

[12] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Los Angeles, USA, June 2019.

[13] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-Stream Neural Networks for Tampered Face Detection," in IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017, pp. 1831–1839.

[14] B. Bayar and M. C. Stamm, "A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer," in Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, Vigo, Spain, 2016, pp. 5–10.

[15] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, doi:10.1109/TPAMI.2020.3009287.