IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

AI DEEP FAKE DETECTION RESEARCH PAPER

[1]Raghava M S , [2]Tejashwini S P, [3]Kavya Sree ,[4]Sneha A , [5]Naveen R
[1]Assistant Professor , Dept of Artificial Intelligence and Machine Learning, Dayananda Sagar Academy of Technology and Management
[2,3,4,5]Student, Dept of Artificial Intelligence and Machine Learning, Dayananda Sagar Academy of Technology and Management

Abstract

Deep learning has demonstrated remarkable success in solving complex problems across various domains, such as big data analytics, computer vision, and human-level control. However, the same advancements in deep learning have also given rise to applications that pose threats to privacy, democracy, and national security. One such application is deepfake technology, which leverages deep learning algorithms to create convincingly realistic fake images andvideos that are indistinguishable from authentic ones. Consequently, the need for technologies capable of automaticallydetecting and assessing the integrity of digital visual media has become imperative.

This paper aims to present a comprehensive survey of the algorithms employed to create deepfakes and, more importantly, the methods proposed in the literature for detecting deep fakes. The survey delves into extensive discussions on the challenges, research trends, and future directions concerning deepfake technologies. By reviewing the background of deepfakes and examining state-of-the-art deepfake detection methods, this study provides an inclusive overview of deepfake techniques, thereby facilitating the development of novel and robust methods to combat the increasingly sophisticated deep fake threats

In conclusion, this survey paper provides a comprehensive overview of deepfake techniques and detection methods. By synthesizing the existing literature and highlighting research trends and challenges, it aims to support the development of novel and effective approaches to combat the growing threat of deep fakes, ensuring the integrity, privacy, and security of digital visual media in an increasingly complex andinterconnected world.

Keywords

Python, generator,kp-detector,predictions

Introduction

Deepfakes, in a narrow definition, are a type of artificial content created using deep learning techniques that involve superimposing face images of a targetperson onto a video of a source person. This manipulation makes it appear as though the target person is performing

actions or saying things that the source person actually did. This specific category of deepfakes is commonly known as face swapping.

However, in a broader sense, deep fakes encompass other types of AI-generated content as well. Two additional categories of deepfakes are lip-sync and puppet-master.

Lip-sync deep fakes involve modifying videos to synchronize the movements of the subject's mouth with a particular audio recording. By altering the original video, the mouth movements of the person in the video are made consistent with the audio, creating a convincing lip-sync effect.Puppet-master deep fakes,on the other hand, consist of videos featuring a target person (the puppet) whose facial expressions, eye movements, and head movements are animated to mimic those of another person (the master) who is situated infront of a camera. The puppet follows theactions and expressions of the master,resulting in a video where the target person appears to be controlled by the movements of the master.

It is important to note that these categories of adeepfakes are not mutually exclusive, and a deepfake can incorporate elements from multiplecategories. The broader definition of deep fakes encompasses not only face swapping but also lip-sync andpuppet-master techniques, enabling awider range of AI-synthesized content that can potentially deceive viewers.

Related works

Driver drowsiness detection techniques can be categorized into three main groups: physiological measures, vehicle-based measures, and behavioral
measures. These methods aim to identifysigns of drowsiness in drivers and alert them to prevent accidents.

1. Survey Trends of Deepfakes :

In the realm of advanced artificial intelligence, the landscape of deepfake generation and detection methods is constantly evolving and becoming more sophisticated. The research community istirelessly working on improving deepfake detection algorithms and has published numerous findings in this area.There is an ongoing struggle between those who utilize advanced machine learning techniques to generate deep fakes and those who strive to identify and distinguish deep fakes from real videos.

In conclusion, the ever-evolving landscape of deepfake technology calls for ongoing research and development efforts to enhance deep fake detection systems. The utilization of Convolutional Neural Networks and other advanced AI techniques, combined with interdisciplinary collaborations, holds the potential to address the challenges posed by deep fakes and restore trust in the authenticity of digital visual media.

2. Tech facts:

The concept of creating fake images or manipulating images with different faces is not new, but recent technological advancements have significantly improved the accuracy and believability of such manipulations. However, generating high-quality deep fakes still poses a challenge. Training deepfake models using adversarial approaches canlead to a noticeable degradation in the quality of the synthesized images, as noted by Yang et al. (2020).

Furthermore, the generation of deepfakesoften requires substantial computational resources. The computing capacity needed for deepfake generation is typically quite demanding, and this can be a limitation for many state-of-the-art approaches.

3.Approaches : Kharbat et al. (2019) deviate from deep learning approaches and instead utilize machine learning algorithms for deepfake detection. Theypropose an algorithm that combines a Support Vector Machine (SVM) classifier with a Histogram of Oriented Gradients (HOG) feature point descriptor. Their approach demonstratesremarkable results in detecting deepfakes, showcasing the effectivenessof integrating machine learning algorithms with deep learning methodologies.

Another indicator used for identifying fake videos involves combining the DenseNet169 model with a facial warping artifact identification technique,as described by Maksutov et al. (2020).

The assumption underlying this inferenceis that most existing deep fake algorithms struggle to synthesize faces with high-quality resolution.

Consequently, to make the manipulated image appear more natural, these algorithms often employ affine transformations. However, these transformations can introduce recognizable visual artifacts, providing a potential marker for identifying deep fakes.

In summary, demonstrate the efficacy of machine learning algorithms, specificallyan SVM classifier with a HOG feature point descriptor, for deepfake detection. Additionally, Maksutov et al. (2020) propose utilizing the combination of a DenseNet169 model and facial warping artifact identification to identify deepfakes based on visible visual artifacts introduced during the synthesis process. These approaches showcase the potential of both machine learning and deep learning techniques in detecting deep fakes.

Methodology

The research in question focuses on utilizing a standard data mining process, namely the Cross-Industry Standard Process for Data Mining (CRISP-DM)methodology. CRISP-DM is a widely adopted methodology in data analyticsdue to its systematic and comprehensive approach. It involves a step-by-step process for carrying out data mining projects, ensuring a structured and well-executed workflow.

A successful data analytics project following the CRISP-DM methodology requires a thorough understanding of the business domain. This includes conducting an initial analysis of the business requirements, acquiring domain knowledge, and then applying theappropriate data mining techniques to gain insights and make informed decisions. This process is oftenconsidered a robust and well-planned strategy for conducting data analytics projects.

The CRISP-DM methodology consists ofseveral phases, and the relationships between each phase are typically depicted in Figure 2 of the research. These phases are as follows:

Business Understanding: This phase involves establishing clear business objectives and understanding the context and requirements of the project. It aimsto align the data mining goals with the overall business goals.

Data Understanding: In this phase, the focus is on acquiring and exploring the available data. It includes activities such as collecting and analyzing data, identifying data quality issues, and assessing the suitability of the data for analysis.

Data Preparation: The data preparation phase involves transforming and cleaning the data to make it suitable for analysis. This includes tasks like data cleaning, feature selection, dataintegration, and data transformation.

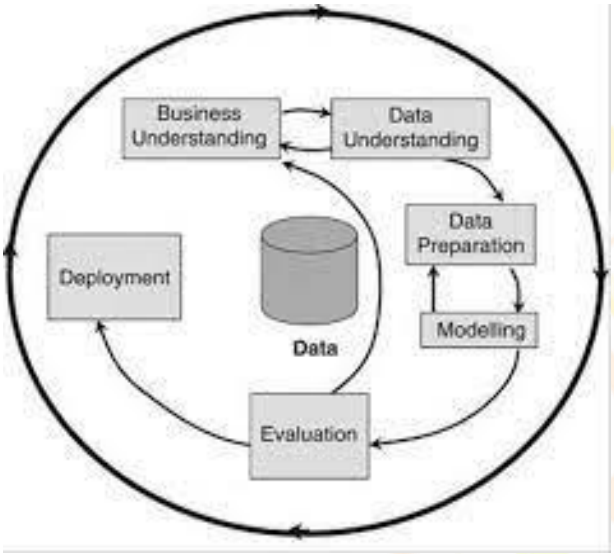
Modeling: In this phase, various data mining techniques and algorithms are applied to build models that can solve the specific business problem. This includes tasks such as selecting the appropriate modeling techniques, building and validating models, and fine-tuning the models for optimal performance.

Evaluation: The evaluation phase focuseson assessing the quality and effectiveness of the models developed in the previous phase. It includes evaluatingthe models' performance, interpreting theresults, and

determining their usefulness for the business objectives.

Deployment:

This includes creating a plan for model deployment, monitoring the model's performance, and establishing mechanisms for continuous CRISP-DM methodology; researchers and practitioners can ensure a comprehensiveand structured approach to data mining projects. The methodology covers essential phases, from understanding the business context to implementing the models in a real-world setting, ensuring that the data analytics process iswell-executed and aligned with the business goals.



Design

Transfer learning is a machine learning technique that involves utilizing apre-existing model trained on a specific task as a starting point for a new, related task. It is particularly prominent in the field of deep learning, where pre-trained models can be leveraged for computer vision and natural language processing tasks.

In transfer learning, the pre-trained model is typically trained on a large-scale dataset, often using high computational resources and time. This initial training enables the model to learngeneric features and patterns that are applicable to various tasks. Rather than starting from scratch, these pre-trainedmodels serve as a foundation for the newtask at hand.

Conclusion and Discussion

The issues surrounding deepfakes and their potential negative impacts are indeed significant in today's media landscape. As the technology for creatingdeep fakes becomes more accessible and social media platforms facilitate rapid dissemination of content, it becomes crucial to address the challenges associated with this phenomenon.

The survey you mentioned, whichprovides an overview of deepfakecreation and detection methods, can be a valuable resource for the artificialintelligence research community. By understanding the techniques used to generate deep fakes, researchers can develop effective methods to detect and mitigate their harmful effects.

Creating reliable and efficient deepfake detection methods is essential to combat the spread of disinformation, hate speech, and political tensions. By implementing robust detection mechanisms, it becomes possible to identify and flag manipulated media content, reducing the potential negative consequences associated with deep fakes.

Furthermore, it is important to raise awareness about deepfakes among the general public. By educating individuals about the existence and implications of deepfakes, they can become more critical consumers of media content and be better equipped to distinguish between real and manipulated information. This could potentially mitigate the erosion of trust caused by deep fakes.

In terms of future directions, ongoing research and development efforts are needed to stay ahead of the evolving deep fake technology. As creators of deep fakes become more sophisticated, detection methods must continually adapt to effectively identify manipulated content

Output

References

[1] G. Lee and M. Kim, "Deepfake detection using the rate of change between frames based on computer vision," Sensors, vol. 21, no. 3, pp. 1–14, 2021
.

[2] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt and M. Nießner, "Face2face: Real-time face capture and reenactment of RGB videos," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Juan,USA, pp. 2387–2395, 2016.

[3] I. Korshunova, W. Dambre and L. Theis, "Fast face-swap using convolutional neural networks," in Proc. IEEE Int. Conf. on Computer Vision, Cambridge, USA, pp. 3677–3685, 2017.

[4] A. Tewari, M. Zollhoefer, F. Bernard, P. Garrido, H. Kim et al., "High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 357–370, 2020.

[5] J. Lin, "FPGAN: Face de-identification method with generative adversarial networks for social robots," Neural Networks, vol. 133, no. 3, pp. 132–147, 2021.

[6] R. Chesney and D. Citron, "Deepfakes and the new disinformation war: The coming age of post-truth geopolitics," Foreign Affairs, vol. 13, no. 3, pp. 1–14, 2019.

[7] S. Lyu, "Deepfake detection: Current challenges and next steps," in Proc. IEEE Int. Conf. on Multimedia & Expo Workshops (ICMEW), London, United Kingdom, pp. 1–6, 2020.

[8] M. T. Jafar, M. Ababneh, M. A. Zoube and A. Elhassan, "Forensics and analysis of deepfake videos," in Proc. 11th Int. Conf. on Information and Communication Systems (ICICS), Jorden, pp. 53–58, 2020.

[9] M. A. Younus and T. M. Hasan, "Effective and fast deepfake detection method based on haar wavelet transform," in Proc. Int. Conf. on Computer Science and Software Engineering (CSASE), Kurdistan Region, Iraq, pp. 186–190, 2020.

[10] D. Afchar, V. Nozick, J. Yamagishi, I. Echizen and

A. MesoNet, “Compact facial video forgery detection network,” in Proc. IEEE Int. Workshop on InformationForensics and Security (WIFS), Hong Kong, China, pp.1–7, 2018.

[11] Y. Li, M. Chang and S. Lyu, “Exposing AI created fake videos by detecting eye blinking,” in Proc. IEEE Int.Workshop on Information Forensics and Security (WIFS), Hong Kong, China, pp. 1–7, 2018.

[12] D. Güera and E. J. Delp, “Deepfake video detectionusing recurrent neural networks,” in Proc. 15th IEEE Int.Conf. on Advanced Video and Signal Based Surveillance(AVSS), Auckland, New Zealand, pp. 1–6, 2018.

[13] S. Agarwal, H. Farid, O. Fried and M. Agrawala,“Detecting deep-fake videos from phoneme-viseme mismatches,” in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW),Seattle, WA, USA, pp. 2814–2822, 2020.

[14] L. Zheng, S. Duffner, K. Idrissi, C. Garcia and A.Baskurt, “Siamese multi-layer perceptrons for dimensionality reduction and face identification, Multimedia Tools and applications

[15] Bin Guo, Yasan Ding, Lina Yao, Yunji Liang, and Zhiwen Yu. The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys (CSUR), 53 (4):1–36, 2020.

[16] Patrick Tucker. The newest AI-enabled weapon: ‘deep-faking’ photos of the earth. https://www.defenseone.com/ t echnology/2019/03/next- phase- ai- deep-fakingwhole-world-and- china-ahead/155944/, March 2019.

[17] T Fish. Deep fakes: AI-manipulated media will be ‘weaponised’ to trick military. https://www.express. co . uk / news / science / 1109783 / deep - fakes - ai - artificial - intelligence - photos - video - weaponised-china, April 2019.

[18] B Marr. The best (and scariest) examples of AI-enabled deepfakes. https://www.forbes.com/sites/bernardmarr/ 2019/07/22/the-best-and-scariest-examples-ofai-enabled-deepfakes/, July 2019.

[19] Yisroel Mirsky and Wenke Lee. The creation and
detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1): 1–41, 2021.

[20] Luisa Verdoliva. Media forensics and deepfakes: an overview. IEEE Journal of Selected Topics in Signal Processing, 14(5): 910–932, 2020..