International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Fake and Image Manipulation

Jasmine Praful Bharadiya

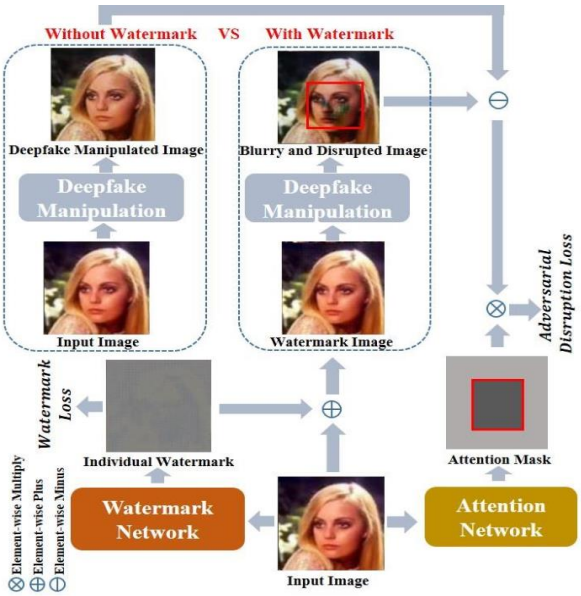
Submitted: 08/12/2023 **Revised**: 19/01/2024 **Accepted**: 29/01/2024

Abstract: It has become more difficult to detect the difference between real and false media because to the rapid advancements in computer graphics and highly artificial intelligence in the production of realistic photos and videos in recent years. complicated. Even if these computer-generated images or movies have practical uses, they can also present privacy and security risks. Deepfake is one method that carries these hazards. Combining the terms "deep learning" and "fake" yields the term "deepfake." Anyone can use Deepfake to edit or erase another person's face from a photo or video. The speech and facial emotions of an original image or video can likewise be altered using deepfakes. These days, deepfake algorithms use artificial intelligence and deep learning to replace the original voice, face, or emotions. It's difficult to tell whether the content has been modified by deepfake techniques. Deepfakes are altered and retouched using deep learning algorithms, which makes it more difficult to distinguish between actual and fake images. Generative adversarial neural networks (GANs) are used to create deepfakes, which may be dangerous for the public. It's imperative to spot fake graphic content with great attention. Several investigations have been carried out to identify deepfakes in photo manipulation. The two main problems with the existing approaches are their time consumption and accuracy.

Keywords: *Generative Adversarial Networks (GAN), Computer Graphics, Watermarking, Deepfake, Artificial Intelligence, And Detection Techniques.*

1. Introduction

Deep learning has become more and more popular in recent years and is utilized in many everyday applications. Conversely, deep learning puts data security and privacy at risk. Deepfake is a popular application of the generative model to replace facial identities or modify facial features in Internet photographs. The edited photos have amazing details and look natural. Deepfake poses a serious threat to public perception of leaders, celebrities and ordinary people due to its ease of use. Generating adversarial attacks has recently been suggested as a way to prevent Deepfake image manipulation. After Deepfake processing, the resulting photos become blurred, making them easy to identify. Nevertheless, the previously stated approach has obvious drawbacks because it uses the adversarial assault strategy, which is intended for classification tasks rather than visual translation. First, only a small percentage of photos have been retouched. Second, most retouched photographs have a color change that is undetectable to the human eye. Third, it takes a long time to generate the protective watermark, which makes it unsuitable for use. We propose a concept called Smart Watermark to overcome the limitations of current technique and manage the risk of facial manipulation. In order to avoid manipulation, it is planned to apply imperceptible watermarks and transform watermarked photos into contradictory instances of the Deepfake model.



1.1 To Guard Your Photos Against Deepfake Image Manipulation, Use a Smart Watermark

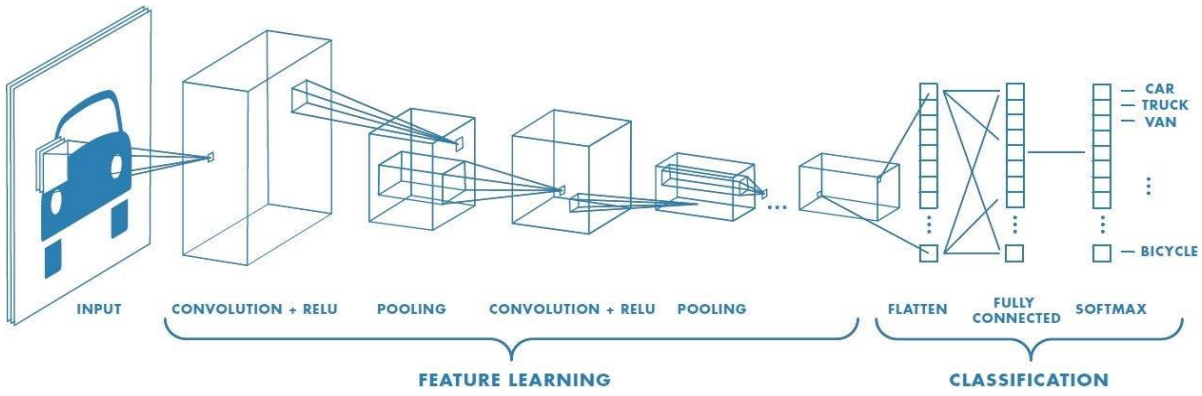
As deep learning techniques have become more prevalent, innovative methodologies have also impacted the creation and identification of deepfakes. As we will see in the next part, this strategy mainly uses AI-based techniques such as CNN architecture, GAN and others. Plus, it's incredibly

[1]* *Department of Information and Technology, University of the Cumberlands, Williamsburg, KY, USA*
Email: jasminbharadiya92@gmail.com,
ORCID: https://orcid.org/0009-0002-4264-6005

simple and time-consuming to create our own user-friendly software thanks to the abundance of code snippets available. This process requires very little professional knowledge. Several research projects have been undertaken to improve the methods and make them more adapted to user expectations. The following part discusses the basic methods of the state-of-the-art methodology used in this work:

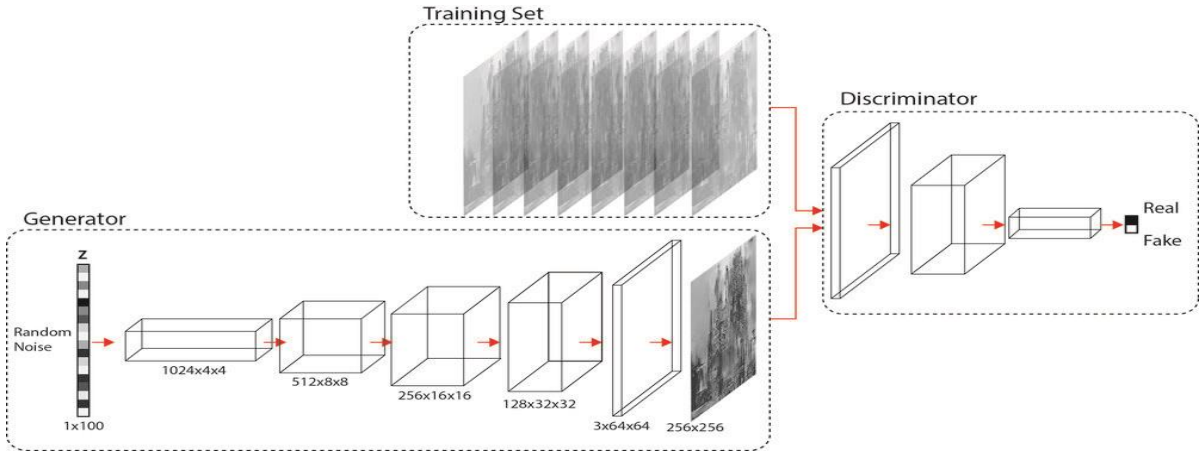
2. CNN Network

Convolutional neural networks, or CNNs for short, are deep learning algorithms that use point differences between images to discern between several categories. This technique is mainly used for the creation and identification of deepfakes. Additionally, the preprocessing required for ConvNet is much lower than that of other classification techniques. The convolution tools layer, which separates the image's many features for processing, makes up half of the CNN network. It is made up of an activation layer, max pooling, and convolution. Based on the output of the convolution tool layer, the second completely linked layer predicts what description of the image would be most appropriate. A vector of probability scores describing the unique qualities of each class and indicating the likelihood based on the predictions produced by the layers is the final output that CNN produces.



3. Generative Adversarial Network (GAN)

A powerful family of neural networks called GANs is used in unsupervised learning. Artificial intelligence systems built on GAN networks—which are made up of a discriminator and a generator—are typically used to produce deepfake movies. The discriminator's task is to distinguish between authentic and fraudulent video samples, whilst the generator creates false ones. Upon successfully detecting a bogus sample, the discriminator provides the generator with guidance on how to avoid creating bogus video samples in the future. On the other hand, the generator gets better at producing fake video samples as the discriminator gets better at identifying them. When the discriminator and generator work together, a generative adversarial network (GAN) is created.



In the future recommendations

1. Enhancing Algorithm for the deep detection:

In order to accurately detect altered data, detection algorithms must advance along with the complexity of deepfake creation techniques. Researchers looking into cutting edge deep learning architectures and techniques can create more robust models that can handle various complex types of deepfakes.

2. Techniques fore multimodal detection and fusion:
Deepfake identification accuracy can be increased by integrating many modalities, such as visual and audio input. Fusion approaches combine the benefits of multiple modalities in an effort to increase the detection system's overall robustness and accuracy.

3. Few-Shot Learning and Transfer Learning:
Transfer learning improves detection performance even with less labeled data by leveraging pre-trained models' expertise gained from large dataset training. Few-step learning facilitates the training of detection models with small sample sets, it can be essential to the creating and adopting to novel of entire deepfake methods.

4. AI explain for the credibility of deepfake:
Operating systems for detecting deepfakes operate as "Black Box" samples of models, making it difficult for users to understand the decisions they make. Explainable AI techniques increase trust and transparency by allowing people to understand the reasoning behind a media's authenticity or falsity categorization.

5. Real-Time Deepfake Detection:
To quickly detect and remove deepfake content from the Internet as soon as it appears, real-time detection is essential. Rapid protection against fraudulent media can be provided by real-time deepfake detection systems by optimizing algorithms and deploying them on efficiently functioning hardware.

6. Diverse and Adversarial Dataset Creation:
Training detection algorithms that can perform effectively in a range of scenarios requires the construction of enormous datasets using a number of deepfake manipulation approaches. Adversarial training attempts to strengthen the defenses of detection systems against hostile attacks by intentionally creating challenges.

7. Collaborative Research and Benchmarking:
Cooperation between academics, businesses, and government organizations promotes information sharing and data exchange, which advances the field of the deep fake detection and collaboration. Equities reference criteria contain impartial distribution and compensation of desire the different detections techniques.

8. User Friendly and deepfake detection tool:
Easy-to-use apps and browser extensions that enable proactive action against potentially infringing content allow users to independently confirm the legitimacy of media they encounter online.

9. Regulations and Policy Development:
To reduce the risk of misuse of deepfake technology, rules and guidelines for its ethical use need to be developed. Working collaboratively with social media companies can make it easier to identify and remove deepfake information from online spaces.

10. AI Ethics and Responsible Use:
Alignment with business principles is ensured by incorporating ethical considerations into the creation of deepfake detection algorithms. Encouraging responsible research and the deployment of deepfake detection technologies requires transparency in the disclosure of results and limitations.

11. Continued Research and Education:
To stay abreast of the development of deepfake techniques and any adversarial attacks, continued research is necessary. Improving general knowledge and preparedness involves educating the public and industry experts about deepfake threats and detection techniques.

Importance of Deepfake image manipulation

Misinformation and Fake Content: Deepfakes can spread false information and fake documents on a never-before-seen scale. These technologies enable the creation of fascinating films and images featuring famous figures, politicians, and celebrities saying or doing things they have never done before. These images and videos are likely to confuse viewers and influence public opinion.

Consent and Privacy Violation: Deepfake technology can be used to create explicit or sensitive content about a subject without that subject's knowledge or consent, which can be upsetting to them and cause emotional distress.

Danger to National Security and Reputation: Deepfakes can be used to disseminate false material that could be detrimental to international relations, national security, and individual reputations by posing as well-known individuals or elected authorities.

Fake Evidence in Legal Proceedings: Deepfakes could be used to fabricate evidence in court, undermining the legal system and possibly leading to incorrect convictions or acquittals.

Erosion of Trust in Media: People may have more difficulty distinguishing between real and fake news due to the spread of deepfake content, which can potentially undermine public trust in media and other trusted sources of information.

Economic and Financial Consequences: Since edited content can be used to spread misleading information about companies, thereby causing stock price fluctuations or harming companies' reputations, deepfake attacks can potentially result in significant financial damage.

Harm to Individuals and Relationships: If false or compromising content is widely disseminated, victims of deepfake manipulation may experience emotional pain, injury to their personal relationships, and damage to their professional lives.

Challenges for Content Moderation: Deepfake content is difficult to identify and remove from social media and content-sharing platforms because sophisticated detection systems are needed to track new alteration techniques.

Protecting Freedom of Expression: Misuse of deepfake technology can lead to issues with censorship and restriction of free speech, as people may be reluctant to communicate honestly and openly for fear of manipulating content.

Advancing AI Research: The development of effective deepfake detection algorithms also advances artificial intelligence research, fostering the search for new approaches to solve this difficult problem.

4. Conclusion

In conclusion, the field of “deepfake and image manipulation detection” is becoming increasingly important due to the exponential growth of digital technology and the possibility of its misuse. Deepfake poses serious risks to data security and privacy as well as the spread of false information. People's ability to differentiate between genuine and fraudulent media is becoming increasingly difficult due to advances in deepfake creation techniques.

In order to address these issues and continually improve deepfake detection systems, researchers need to study complex deep learning architectures, multi-modal detection strategies, and transfer learning methodologies. To build user trust and transparency in detection systems, explanatory AI is crucial. Ethical considerations, user-friendly tools, and real-time detection capabilities are essential to protect people from the dangers of deepfakes. To prevent the malicious use of deepfake technology, policymakers, industry and researchers must work together to establish uniform standards and laws. Continued research and education are needed to stay current with advances in deepfake techniques and to increase public and professional knowledge.

With technology rapidly evolving, ethical and proactive approaches to identifying deepfakes are essential to uphold societal norms, protect the integrity of digital media, and protect people from potential harm. to distorted information.

5. Recommendation

In order to stay up with the rapid progress of deepfake technology, research and development are necessary. Bringing together experts in audio analysis, computer vision, natural language processing, and psychology can result in more comprehensive and dependable deeper fake detection systems. Increasing public awareness of deepfakes, educating people about them, and giving them the tools to identify and respond to manipulated media will lead to a safer and more knowledgeable digital society.

References

[1] Rossler, A., Cozzolino, D., Verdoliva, L., et al. (2019). "FaceForensics++: Learning to Detect Manipulated Facial Images." In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[2] Li, Y., Yang, X., Sun, P., et al. (2018). "Exposing DeepFake Videos by Detecting Face Warping Artifacts." In arXiv preprint arXiv:1811.00656.

[3] Marra, A., Gragnaniello, D., & Verdoliva, L. (2019). "Detection of GAN-Generated Fake Images Over Social Networks." In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[4] Hsu, C., Wu, W., & Kirchherr, J. (2020). "Deep Learning-Based Image Forensics: A Comprehensive Review." IEEE Access, 8, 187760-187782.

[5] Cozzolino, D., Gragnaniello, D., & Verdoliva, L. (2018). "Sparsity-Based Forensic Analysis of Deep Learning Models." In Proceedings of the European Conference on Computer Vision (ECCV).

[6] Nguyen, A., Yosinski, J., & Clune, J. (2015). "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[7] Bayar, B., & Stamm, M. C. (2016). "A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer." In Proceedings of the IEEE Workshop on Information Forensics and Security (WIFS).

[8] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). "MesoNet: A Compact Facial Video Forgery Detection Network." In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec).

[9] Zhou, Y., Ye, Q., Qiu, W., et al. (2017). "Two-Stream Inception Network for Detection of Faked Images and Videos." In Proceedings of the IEEE

International Workshop on Information Forensics and Security (WIFS).

[10] Güera, D., Erdenebat, M., & Erdenebayar, U. (2019). "Deepfake Video Detection Using Recurrent Neural Networks." In Proceedings of the IEEE International Conference on Image Processing (ICIP).

[11] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., et al. (2020). "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." Information Fusion, 64, 131-148.

[12] Zhao, Y., Zheng, L., Zheng, Z., et al. (2020). "Detecting Deepfake Videos from the Clues Left in the Deep Learning Models." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[13] Frid-Adar, M., Diamant, I., Goldberger, J., & Greenspan, H. (2018). "GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification." Neurocomputing, 321, 321-331.

[14] Attias, D., Michaeli, T., & Irani, M. (2019). "Logical Adversarial Networks for Active Visual Testing." In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[15] Matern, F., Riess, C., Stotzka, R., et al. (2019). "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations." In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[16] Rossler, A., Cozzolino, D., Verdoliva, L., et al. (2019). "FaceForensics++: Learning to Detect Manipulated Facial Images." In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[17] Li, Y., Yang, X., Sun, P., et al. (2018). "Exposing DeepFake Videos by Detecting Face Warping Artifacts." In arXiv preprint arXiv:1811.00656.

[18] Marra, A., Gragnaniello, D., & Verdoliva, L. (2019). "Detection of GAN-Generated Fake Images Over Social Networks." In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[19] Hsu, C., Wu, W., & Kirchherr, J. (2020). "Deep Learning-Based Image Forensics: A Comprehensive Review." IEEE Access, 8, 187760-187782.

[20] Cozzolino, D., Gragnaniello, D., & Verdoliva, L. (2018). "Sparsity-Based Forensic Analysis of Deep Learning Models." In Proceedings of the European Conference on Computer Vision (ECCV).

[21] Nguyen, A., Yosinski, J., & Clune, J. (2015). "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[22] Bayar, B., & Stamm, M. C. (2016). "A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer." In Proceedings of the IEEE Workshop on Information Forensics and Security (WIFS).

[23] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). "MesoNet: A Compact Facial Video Forgery Detection Network." In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec).

[24] Zhou, Y., Ye, Q., Qiu, W., et al. (2017). "Two-Stream Inception Network for Detection of Faked Images and Videos." In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS).

[25] Güera, D., Erdenebat, M., & Erdenebayar, U. (2019). "Deepfake Video Detection Using Recurrent Neural Networks." In Proceedings of the IEEE International Conference on Image Processing (ICIP).

[26] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., et al. (2020). "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." Information Fusion, 64, 131-148.

[27] Zhao, Y., Zheng, L., Zheng, Z., et al. (2020). "Detecting Deepfake Videos from the Clues Left in the Deep Learning Models." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[28] Frid-Adar, M., Diamant, I., Goldberger, J., & Greenspan, H. (2018). "GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification." Neurocomputing, 321, 321-331.

[29] Attias, D., Michaeli, T., & Irani, M. (2019). "Logical Adversarial Networks for Active Visual Testing." In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[30] Matern, F., Riess, C., Stotzka, R., et al. (2019). "Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations." In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).