IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Deep Fake Detection With Machine Learning Using CNN And Resnet Algorithm

[1]Mr.M.Sundaram, [2]Ms.J.Nithya, [3]Ms.K.Kanishka, [4]Ms.S.Selvasri

[1]Assistant professor, [2]Student, [3]Student, [4]Student

Department of Computer Science and Engineering

Pavai College of Technology, Pachal, Namakkal,Tamilnadu, India.

Abstract: Video forgery is continuously increasing in the digital world due to breaches of information security, consequently establishing a scenario for image and video content monitoring for forgery identification. The spread of fake videos raises security risks and anarchy in society. The reason for video forgery is to the augmentation in the malware, which has facilitated user (anyone) to upload, download, or share objects online comprising audio, images, or video. With the development of technology and ease of creation of fake content, the manipulation of media is carried out on a large scale in recent times. Video forgery detection has applications in media science, forensic analysis, digital investigations, and authenticity verification of a video. The purpose of video forensic technology is to extract features to distinguish fake content frames from original videos. Deepfake media has posed a great threat to media integrity and is being produced and spread widely across social media platforms, the detection of which is seen to be a major challenge. Proposed, an approach for Deepfake detection has been provided for forgery detection in video. ResNet, a Convolutional Neural Network (CNN) algorithm is used as an approach to detect the Deepfake videos. The model aims to enhance the performance of detecting forgery videos produced by a certain method as well as enhance the accuracy of the detector. The proposed approach only uses the deep features extracted from the ResNet CNN model then applies the conventional mathematical approach on these features to find the forgery in the video. It is the detector to be constantly updated with real-world data, and propose an initial solution in hopes of solving Deepfake video detection.

Index Terms **– Deep Fake, CNN, RESnet, Video Forgery Detection.**

1. INTRODUCTION

The rapid advent of computationally cheap and cross-platform video editing software has enabled the huge volume of video content available to a large number of users via the Internet. In recent years, the abundance of video data, AI techniques, and readily available, high-performance easy to use video editing tools have given rise to fake videos. Fraudulent activities are carried out using fake images and videos to bypass facial authentication, to publish fake news, and for entertainment as well. Video forgery is continuously increasing in the digital world due to breaches of information security, consequently establishing a scenario for image and video content monitoring for forgery identification. The spread of fake videos raises security risks and anarchy in society. Video forgery detection has applications in media science, forensic analysis, digital investigations, and authenticity verification of a video. The purpose of video forensic technology is to extract features to distinguish fake content frames from original videos. The developing computational force has made deep learning so incredible that would have been thought unthinkable just a modest bunch of years prior. Like any extraordinary innovation, this has made new difficulties. Purported "DeepFake" created by deep generative adversarial models that can control video and brief snippets.

For identification of the DF (Deep Fake), it is very essential to recognize the manner Generative Adversarial Network (GAN) creates the DF.GAN takes as input a video and a picture of a particular individual ('target'), and outputs some other video with the goal's faces changed with the ones of some other individual ('source'). The spine of DF is deep adversarial neural networks trained on face images and target films to robotically map the faces and facial expressions of the source to the target. With the right post processing, the ensuing films can gain an excessive degree of realism. The GAN breaks up the video into frames and replaces the input picture in each frame. Further, it reconstructs the video. This interaction is generally accomplished by utilizing auto encoders. We describe brand new deep learning-primarily based totally approach that could efficiently distinguish DF films from the actual ones. Copy move forgery problem is investigated a lot, however, inter-frame duplication is not explored much and still is not applicable in real-time due to computational limitations and robustness issues for real-time scenarios. Existing approaches in the literature suffer from low accuracy rates, low efficiency, and high computational complexity. Further, most of the existing approaches are trained on datasets with a limited sample size; which is not enough to unleash the potential of deep learning performance. Moreover, most of the existing work does not address variable frame rates for sample videos.

2. PROBLEM STATEMENT

The rapid advancement of digital manipulation techniques, particularly the emergence of deepfake technology, has led to an alarming increase in the production and dissemination of forged video content. Deepfake videos, created using sophisticated AI algorithms, pose a significant threat to the integrity of media content and societal trust in digital information. To address this critical issue, this project aims to propose a novel approach for deepfake video detection leveraging ResNet, a Convolutional Neural Network (CNN) architecture renowned for its effectiveness in image classification tasks. The primary objectives are:

Forgery Detection Enhancement: Enhance the performance of deepfake video detection by leveraging ResNet's capabilities to extract deep features from video frames and identify subtle discrepancies indicative of manipulation.

Accuracy Improvement: Improve the accuracy of the forgery detection model by integrating deep features extracted from ResNet with conventional mathematical analysis techniques, thereby enabling more precise identification of forged video content.

Real-World Adaptability: Develop a framework for continuous updating of the forgery detection model with real-world data to adapt to evolving deepfake techniques and maintain effectiveness in detecting newly emerging threats.

3. OBJECTIVE AND GOAL

Objectives

Develop a Deepfake Detection Model: Implement a robust deep learning model utilizing the ResNet architecture to effectively distinguish between authentic and manipulated video content.

Feature Extraction: Utilize ResNet to extract deep features from frames of both original and potentially forged videos, capturing subtle discrepancies indicative of manipulation.

Real-World Adaptability: Design a framework for continuous updating of the detection model with real-world data, enabling adaptation to evolving deepfake techniques and ensuring effectiveness in detecting newly emerging threats.

Goals:

- Achieve a high level of accuracy and reliability in distinguishing between authentic and forged videos, effectively mitigating the spread of deepfake content across digital platforms.
- Contribute to the mitigation of security risks and combatting misinformation by providing a reliable mechanism for identifying and flagging deepfake videos in online media.
- Uphold the integrity of digital media by providing media scientists, forensic analysts, and digital investigators with a powerful tool for verifying the authenticity of video content.

- Contribute to the advancement of video forensic technology by proposing a novel approach for deepfake detection that combines the power of deep learning with conventional analysis techniques.

4. PROJECT SCOPE AND LIMITATIONS

The project scope encompasses the development of a deep learning-based approach using the ResNet CNN architecture for detecting deepfake videos. This entails collecting and preparing a diverse dataset comprising authentic and manipulated videos, followed by training and optimizing the ResNet model to learn discriminative features for distinguishing between the two. Feature extraction and analysis will be performed on video frames to identify discrepancies indicative of manipulation, integrating conventional mathematical techniques for further refinement. Evaluation and validation of the detection algorithm will be conducted using standard metrics, with testing on both synthetic and real-world deepfake videos. Additionally, a framework for continuous updating of the algorithm with real-world data will be developed to adapt to evolving deep fake techniques.

Limitations:

- The effectiveness of the detection algorithm may be limited by the availability and diversity of the dataset, particularly in capturing the full range of deepfake techniques and variations.
- The computational requirements for training and testing the ResNet CNN model, especially on large-scale video datasets, may be substantial and could pose limitations on the scalability of the project.
- The detection algorithm may still produce false positives or false negatives, leading to potential misclassification of authentic videos as deepfakes or vice versa.

5. EXISTING SYSTEM

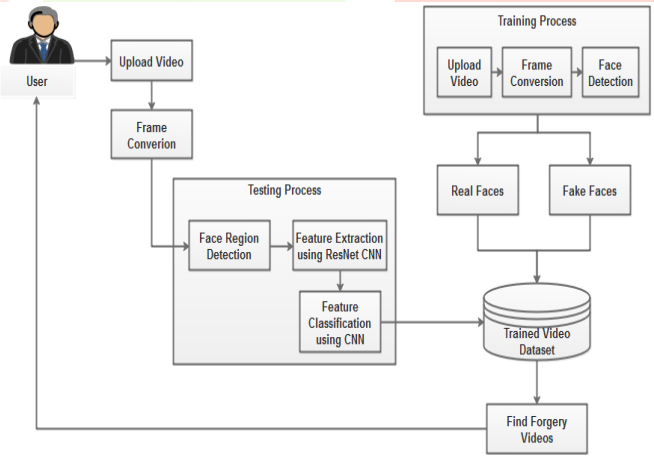
Exiting system implemented a method based on meta-learning called the meta-deepfake detection (MDD) algorithm. With a meta-optimization objective, in order to learn efficient face representations on both synthetic source and target domains. The MDD shifts the source domain to the target domain. So as to increase model generalization, the gradients from the meta-train and the meta-test are combined using meta-optimization. The MDD can handle unseen domains without model updating for unseen domains. Here separated the source domains into the meta-train domain Ttrains and the meta-test domain Ttests during training to obtain domain generalization. In order to simulate the domain shift problem that existed when used in real-world situations, the model is driven to acquire generalizable information about how to generalize well on the new domains with different distributions. We also create meta-batches for training and testing by randomly splitting N source domains of TS. These data contain both real and fake face pairs and these patterns are not duplicated across domains. These pairs increase collation and comparison of information between real and fake images. Therefore, it also increases inter-class separability, which can be interpreted as a distinct dispersion of the feature distribution of samples, increasing differentiation during training as well as enhancing the model's quality. More distinguishable characteristics may be learned by the network with less effort during optimization. The fact is that features learned by supervised learning have much less ability to generalize when subjected to unseen manipulation techniques. Therefore, the model is easier to generalize when the source domain is split into meta-train and meta-test. In addition, samples in the meta-train and meta-test are also shuffled and selected at random, which minimizes the problem of overfitting. Additionally, the data in the unseen domain is very diverse in reality, which the model has never seen or been trained in before. Thus, meta-splitting makes the model easier to train and also to generalize to unseen data.

6. PROJECT PERSPECTIVE

Implement a brand new deep learning based approach that could successfully distinguish AI-generated fake videos (DF Videos) from actual videos. It's incredibly essential to broaden technology that could spot fakes, so that the DF may be recognized and averted from spreading over the internet. Recently, Convolution Neural Network (CNN) has turn into a de-facto technique for classification of multi-dimensional data and it renders standard and also highly effectual network layer arrangements. But these architectures are limited by the speed due to massive number of calculations needed for training in addition to testing the network and also, it might render less accuracy. To trounce these issues, this paper proposed to ameliorate the image and video forgery detection's efficiency utilizing hybrid CNN. Initially, the intensive along with incremental learning phase is carried out. After that, the hybrid CNN is implemented to detect the image together with video forgery.

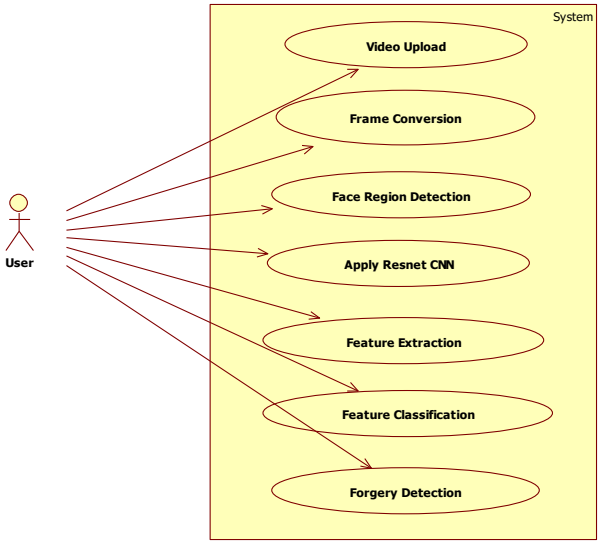
Recently, Deep learning methods and specifically Convolution Neural Networks (CNNs) have gained tremendous success due to its powerful ability of automatic learning of features for large-scale video classification. The purpose of this proposed system is to investigate deep learning method for video forgery detection on large datasets with varied frame rates. A deep neural network approach is presented to classify the forged videos by finding duplicated frames in a video. The first step is to design a Convolutional Neural Network (CNN) that can work as a backbone for feature extraction. The feature extractor generates feature maps of different sizes. Then, these extracted feature maps help the classifier (another neural network) predict the image's nature (i.e., real or fake). In this work, inverted residual blocks and linear bottlenecks are used as intermediate layers to prevent the loss of spatial information. The residual blocks are also memory efficient as they reduce the representations' dimensionality. The network performance is improved by adding different modern techniques and micro design changes during training. These changes help the model learn discriminative features faster with less inference time.

7. SYSTEM DESIGN MODEL



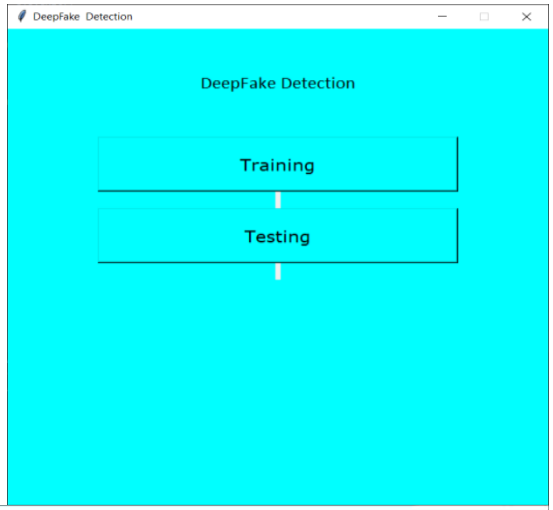
8. USECASE DIAGRAM

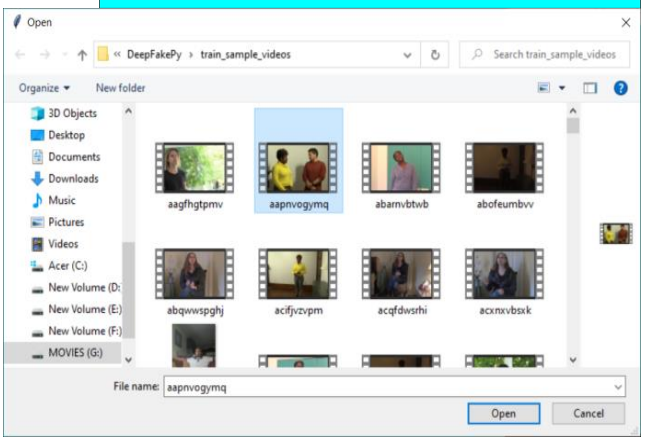
A use case is a list of steps, typically defining interactions between a role (known in Unified Modeling Language (UML) as an "actor") and a system, to achieve a goal. The actor can be a human, an external system, or time. In systems engineering, use cases are used at a higher level than within software engineering, often representing missions or stakeholder goals. Use Case Diagram has actors like sender and receiver. Use cases show the activities handled by both sender and receiver.



9. MODULES

- Video Upload and Training
- Input Video Processing
- Feature Extraction
- Forgery Detection





CNN

- A CNN consists of multiple hidden layer, input layers and an output layer. Hidden layers in a CNN consist of convolutional layers, pooling layers, fully connected layers and normalization layers. The input is the target image to be classified and the output is the context of the bird nest within the image. In addition, there is a cost function used to find the most fitted set of parameters and activation functions to determine the final output.
- In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called Residual Blocks. In this network, we use a technique called skipconnections. The approach behind this network is instead of layers learning the underlying mapping; we allow the network to fit the residual mapping. So, instead of say H(x), initial mapping, let the network fit,
  $F(x) := H(x) - x$ which gives $H(x) := F(x) + x$.
- The advantage of adding this type of skip connection is that if any layer hurt the performance of architecture then it will be skipped by regularization. So, this results in training a very deep neural network without the problems caused by vanishing/exploding gradient.

ResNet

- The Residual Blocks idea was created by this design to address the issue of the vanishing/exploding gradient. We apply a method known as skip connections in this network. The skip connection bypasses some levels in between to link-layer activations to subsequent layers. This creates a leftover block. These leftover blocks are stacked to create resnets. The strategy behind this network is to let the network fit the residual mapping rather than have layers learn the underlying mapping. Thus, let the network fit instead of using, say, the initial mapping of H(x),
  F(x) := H(x) - x which gives H(x) := F(x) + x.

10. SYSTEM TESTING

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program input produces valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application . This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Functional testing

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.
Invalid Input : identified classes of invalid input must be rejected.
Functions : identified functions must be exercised.
Output : identified classes of application outputs must be exercised. Systems/Procedures: interfacing systems or procedures must be invoked.

System testing

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre- driven process links and integration points. It ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

Performance testing

The Performance test ensures that the output be produced within the time limits, and the time taken by the system for compiling in fig(7.4), giving response to the users and request being send to the system for to retrieve the results.

Integration testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to applications, shown in fig(7.5) e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Acceptance testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

- The Acknowledgements will be received by the Sender Node after the Packets are received by the Destination Node
- The Route add operation is done only when there is a Route request in need
- The Status of Nodes information is done automatically in the Cache Updating process

11. CONCLUSION AND FUTURE SCOPE

Conclusion

- In this work implemented DeepFake detection model to detect video forgery using DeepFake video dataset. A deepfake detection method using convolutional neural network (CNN) and ResNet to extract temporal features of a given video sequence is presented, in which features are represented via the sequence descriptor. The detection network consisting of fully connected layers is employed to take the sequence descriptor as input and calculate probabilities of the frame sequence belonging to either authentic or deepfake class. Proposed model when trained with a large data set gave quiet impressive results compared to other deep learning models. Novel approach of combining feature extraction capability of Convolutional Neural Network and ResNet gave better results when compared with existing models. This shows that a simple model which uses ResNet convolutional neural networks can be used to check if a video has been subjected to any kind of manipulation. Proposed work can offer a first line defence in the detection of AI based fake videos.

Future Scope

- Deepfakes are constantly evolving, and new techniques are being developed to create more convincing fakes. Need to develop ResNet CNNs that are capable of detecting a wider range of deepfakes, including those created using new techniques. Real-time detection of deepfakes is critical for preventing the spread of misinformation on social media platforms. Future researchwork focus on developing ResNet CNNs that can detect deepfakes in real-time, using limited computational resources.

12. ACKNOWLEDGEMENT

We are grateful to Mr.M.Sundaram., Assistant Professor, Department of Computer Science and Engineering, Pavai College of Technology (Anna University) for mentoring us to present the paper successfully.

13. REFERENCES

[1] Wu, Rongliang, Gongjie Zhang, Shijian Lu, and Tao Chen. "Cascade ef-gan: Progressive facial expression editing with local focuses." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5021-5030. 2020.

[2] Li, Yuezun, Xin Yang, Pu Sun, Honggang Qi, and SiweiLyu. "Celeb-df: A large-scale challenging dataset for deepfake forensics." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3207-3216. 2020.

[3] Shen, Yujun, JinjinGu, Xiaoou Tang, and Bolei Zhou. "Interpreting the latent space of gans for semantic face editing."In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9243-9252. 2020.

[4] Nirkin, Yuval, Yosi Keller, and Tal Hassner. "FSGANv2: Improved subject agnostic face swapping and reenactment." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 1 (2022): 560-575.

[5] Nguyen, ThanhThi, Quoc Viet Hung Nguyen, Dung Tien Nguyen, DucThanh Nguyen, Thien Huynh-The, SaeidNahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, and Cuong M. Nguyen. "Deep learning for deepfakes creation and detection: A survey." Computer Vision and Image Understanding 223 (2022): 103525.

[6] Huang, Yihao, Felix Juefei-Xu, Qing Guo, Yang Liu, and GeguangPu. "Fakelocator: Robust localization of GAN-based face manipulations." IEEE Transactions on Information Forensics and Security 17 (2022): 2657-2672.

[7] Wang, Zhi, YiwenGuo, and WangmengZuo. "Deepfake forensics via an adversarial game." IEEE Transactions on Image Processing 31 (2022): 3541-3552.

[8] Rana, MdShohel, Mohammad NurNobi, BeddhuMurali, and Andrew H. Sung. "Deepfake detection: A systematic literature review." IEEE Access (2022).

[9] Wubet, WorkuMuluye. "The deepfake challenges and deepfake video detection." Int. J. Innov. Technol. Explor. Eng 9 (2020).

[10] TRIPATHY, SOUMYA, JUHOKANNALA, AND ESARAHTU. "FACEGAN: FACIAL ATTRIBUTE CONTROLLABLE REENACTMENTGAN."

IN PROCEEDINGS OF THE IEEE/CVF WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION, PP.