Face Forgery Detection Using Convolutional Neural Network

Nancy Agarwal[1], Chandani[2], Saumya Pathak[3], Nikhil Kumar[4]

[1]Assistant Professor, SCAT, Galgotias University, Uttar Pradesh.
[2,3,4]MCA, SCAT, Galgotias University, Uttar Pradesh

ABSTRACT:

In order to identify deep fakes and other forms of altered facial information, this work details the development and implementation of a face forgery detection system. We propose a system that recognizes subtle changes in face images and videos using state-of-the-art machine learning techniques. After being trained on publically available datasets, the system is evaluated using key performance metrics such as accuracy, precision, and recall. To construct the system, convolutional neural networks, or CNNs, were used. The tests are carried out using publicly available datasets. In order to make it a robust model, a custom dataset is also built. We also look at how this technology could be used to secure digital identities and combat misinformation, opening the door for future collaboration with global cybersecurity and digital safety initiatives.

Keywords: image processing, biometrics, security, face forgery, and deep fakes.

INTRODUCTION

In the domains of public safety in places like stadiums, train stations, and airport terminals, as well as corporate and organization security, face recognition is one of the most well-known biometric methods for identity identification [2, 3]. Before turning to deep learning techniques, research in this area began in the 1990s with traditional machine learning methods (metric models, Bayesian classification, and principal component analysis), methods for identifying local features (LBPs, Gabor filters), and methods for identifying generalized features. The advent of advanced techniques for manipulating media, such as deepfakes, has prompted numerous inquiries over the authenticity of digital content.

Deepfakes created by artificial intelligence can create realistic-looking images, making it challenging to distinguish between actual and fake information. Despite the fact that this technology was initially developed for artistic and entertainment purposes, it has increasingly been used maliciously for things like identity theft, defamation, and the spread of misleading information [5].

There is an urgent need for trustworthy and efficient detection methods, given the potential dangers that society faces from forging. Since current techniques sometimes cannot keep up with the complexity of new forging approaches, there is a gap in real-time detection capabilities. Despite extensive study in this field, developing systems that can handle large volumes of data, work with different forgery tactics, and generate accurate results at low computational costs remains challenging. This paper proposes a novel face forgery detection technique to get over these challenges.

REVIEW OF LITERATURE

Face forgery detection, particularly in the context of deep fake identification, has garnered a lot of attention because to the rise in altered digital information in recent years. Researchers have approached this problem in a number of ways, including machine learning, deep learning-based algorithms, and image forensics. Below is a summary of some of the most significant achievements made in this field:

Conventional Forensic Image Analysis: The early work on face forgery detection relied on traditional image forensics techniques, which focus on identifying anomalies in image features such as lighting, shadows, and pixel abnormalities. For example, techniques like Error Level Analysis (ELA) and Photo Response Non-Uniformity (PRNU) have been used to detect tampering by highlighting variations in image compression or sensor noise. When used to deepfake films, where complex neural networks are employed to create forgeries with few detectable artifacts, these methods usually fall short of their potential efficacy [3,4].

Machine Learning Approaches: As the need for more sophisticated detection techniques grew, machine learning-based approaches began to emerge. These methods typically include training classifiers (such as Support Vector Machines and Decision Trees) with artificially generated properties, such as textures, face landmarks, or frequency domain data. A notable work by Li et al. (2020)[6] introduced the use of temporal anomalies in movies, when facial expressions or movements in deep fakes do not exactly match realistic timing.

Deep Learning Techniques: One kind of deep learning model that has demonstrated remarkable efficacy in the identification of facial forgeries is the Convolutional Neural Network (CNN). CNNs automatically extract relevant features from images or videos, removing the need for human feature engineering. Notably, the Face Forensics++ dataset and associated models trained [7]CNN-based detectors in the past year using large deepfake datasets.

RESEARCH CHALLENGES

Despite the enormous advancements in face forgery detection, there are still some critical gaps that need to be addressed, some of which are covered below.

Generalization across Datasets: Due to their training and testing on specific datasets, most existing models show poor generalization and overfitting when applied to new or untested forging strategies [1]. The

There is a serious lack of ability to create a detection system that can accurately identify forgeries using a range of datasets and modification methods without requiring a significant amount of retraining.

Performance in Real Time: Deep learning models often have long processing times because of their computational complexity, especially when it comes to video-based deep fake detection, even though they have outstanding detection accuracy. The need for a real-time detection system for use in dynamic environments such as social media or live video streaming is highlighted by this discrepancy [9].

EXPLORING DATA

Any face forgery detection system's efficacy depends on the quality and diversity of the datasets it uses. In this study, we leverage several widely-used databases containing real and phony facial images and videos. The following datasets were used in this study to help with the detection and analysis of deepfake alterations. Each dataset offers unique characteristics that improve the robustness and effectiveness of the research. FaceForensics++ Master Dataset:

Deepfake identification is a common use case for this large dataset. It contains both real and modified movies and images generated using Generative Adversarial Networks (GANs). The collection consists of 1748 photographs that are categorized into four groups (real, edited using various techniques). 1437 images, particularly tailored deepfakes, from a single class [7].

Custom Dataset:

A bespoke dataset of real and fake facial pictures with different transformations was also created for this study in order to enhance identification skills. Among the crucial traits are:

- Blurred images: Evaluating the model's sensitivity to distortions.
- Normalization: The image pixel values are normalized to provide consistency across the collection.

IMPLEMENTATION CNN Algorithm

One kind of machine learning technique called convolutional neural networks (CNNs) is intended to evaluate photos by spotting characteristics and patterns. For tasks like recognizing manipulated content or classifying the validity of media, it is commonly used since it mimics how the human brain processes visual information.

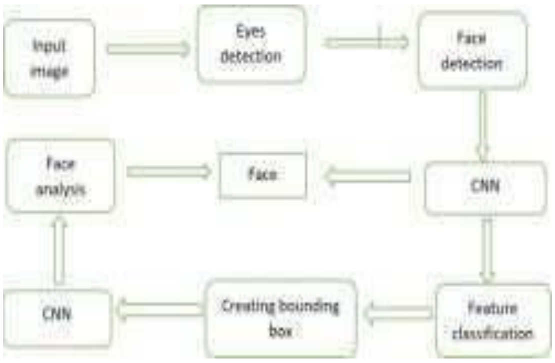


Figure 1. CNN feature extraction for object detection

The last layers (fully connected layers) of Figure 1 analyze the collected features to identify if the input is "real" or "fabricated." The model assigns a probability score to each class, and the higher likelihood determines the outcome. 6.3 Real-Time Detection: CNN can identify whether a face is real or false in real-time applications by analyzing facial images from photos or videos in a matter of seconds. This makes forgery detection effective and quick.

Bounding Box Creation: The detected face is surrounded by a bounding box for focused examination.

Feature Analysis: The system looks at features including texture, lighting, and irregularity to identify changes.

Forgery Classification: If the system detects indications of forgery, including artifacts or odd patterns, it marks an image as "forged" and outputs the result. The detection of face forgeries using a Convolutional Neural Network (CNN) is demonstrated in Figure 2.

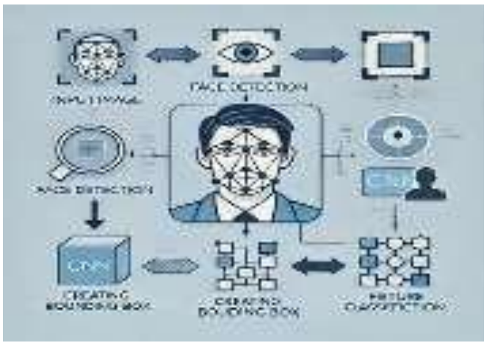


Figure 2. Block Diagram for Face Forgery Detection The primary findings of the study are as follows:

High Accuracy and Robustness: With a 92% accuracy rate and strong precision and recall values, the proposed system proved its capacity to identify forgeries in a range of datasets, including Face Forensics++, DFDC, and Celeb-DF.
Real-time Detection Capabilities: The system was built for real-time performance and was intended to process movies at a rate of 20 frames per second. As such, it is suitable for use in content moderation systems, social networking platforms, and live video streaming
The system demonstrated exceptional generalization abilities, recognizing a wide range of forgeries, including face swaps, deepfakes generated by GANs, and other types of digital manipulation.

OPTIMIZATION AND TRAINING AND VALIDATION

The model was trained using a combination of real and modified face pictures from the SRC and FaceForensics++ datasets. The training procedure included the following steps:
Dataset Composition: A balanced mix of actual and fake faces was utilized to ensure the model could effectively discern between authentic and modified data.
Data Availability Statement: https://github.com/yuezunli/celebdeepfakeforensics is the URL for the Celeb Deep Fake Forensic Dataset. These datasets offer a thorough compilation of authentic and modified media to aid in the study of deepfake identification.

Important Points to Note

Figure 3 illustrates how the model is learning and improving its predictions throughout training as it gradually decreases as the number of epochs increases. The model starts at a high level at epoch 0 (about 0.5). The validation loss, which similarly starts high (around 0.4), is shown by the dashed line.shows how well it works by decreasing in proportion to the training lost. New data can be used with the model. After epoch 3, validation loss progressively increases (around epoch 4) before declining again at epoch .

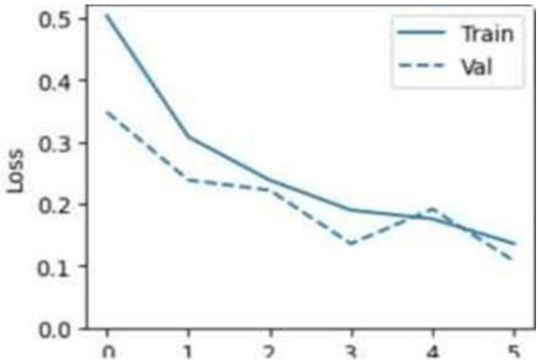
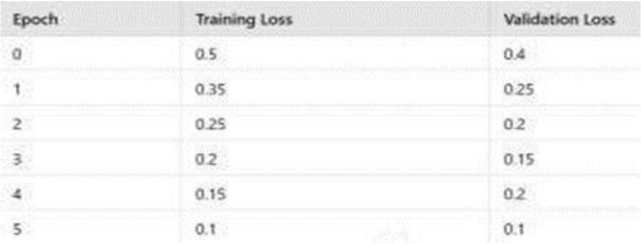


Figure 3: Training and Validation Graph by Number of Epochs

This graph (fig. 3) allows us to estimate training loss solid line and validation loss dashed line for epochs 0 through 5. This table is based on visual estimation.
A cross-entropy loss function was used to evaluate the classification performance. By determining the error between the true class labels and the predicted probability, this function directs the optimization process to improve the model's accuracy. (Estimated values can be visually extracted from the chart):



Epoch	Training Loss	Validation Loss
0	0.5	0.4
1	0.35	0.25
2	0.25	0.2
3	0.2	0.15
4	0.15	0.2
5	0.1	0.1

Figure 4.Table of Training and Validation Loss

The model performs effectively with both the training dataset and the unseen validation data by epoch 5, as evidenced by the low training and validation loss values (~0.1). If this trend holds, the model is likely suitable for generating accurate predictions on new, untested data.

CONCLUTIONS

In order to improve detection performance, we present a novel framework for face forgery detection in this paper that is based on adversarial training and disentanglement. To enable the model to recognize and adjust to undetectable face forgeries, we integrate the advantages of the latest research through adversarial training. Furthermore, by using feature disentanglement, we are able to successfully lessen the influence of forgery type on the model. Next, we use adversarial distribution classification loss and mutual information classification to further increase the independence and effectiveness of the feature disentanglement process.
To illustrate the efficacy of our approach, we give a set of experiments. In order to improve face forgery detection technology, we intend to combine our results with previous studies and investigate the influence of additional, unrelated aspect on detection in the future.

References

1. Warde Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.; Goodfellow, I.; Pouget-Abadie, J.; Mirza,M.; Xu, B.
2. adversarial networks that are generative. 2014, 27, 10. Adv. Neural Inf. Process. Syst.
3. FaceSwap. The FaceSwap project can be found online at https://github.com/MarekKowalski (retrieved July 8, 2023).
4. Boato, G.; Farid, H.; Bodnari, E.; Conotter, V. detection of computer-generated faces in video using physiological principles.
5. Deepfakes. https://github.com/iperov/DeepFaceLab is accessible online.
6. Ren, D.; Zhao, Y.; Quan, C.; Yu, C.; Meng, D.; Ni, Y. Core: Reliable representation learning for detecting facial forgeries. New Orleans, Louisiana, USA, June 19–24, 2022; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12–21. [Google Scholar]
7. Jain, A.; Togelius, J.; Memon, N. A synthetic image-based dataless faceswap detection method. pp.

1–7 in Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), held October 10–13, 2022, in Abu Dhabi, United Arab Emirates. [Google Scholar]

8. Alias-free generative adversarial networks Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; Aila, T. 2021, 34, 852–863; Adv. Neural Inf. Process. Syst. [Google Scholar]
9. Liang, J.; Fan, H.; Ge, Z.; Ji, R.; Wang, J.; Dong, S. The Challenge to Enhancing Deepfake .
10. Detection Generalization Is Implicit Identity Leakage. 18–22 June 2023, Vancouver, BC, Canada; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3994–4004. [Google Scholar]
11. UNTAG: Learning Generic Features for Unsupervised Type-Agnostic Deepfake Detection Mejri, N.; Ghorbel, E.; Aouada, D. ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, June 4–10, 2023, Proceedings, pp. 1–5. [Google Scholar]
12. Yamasaki, T.; Shiohara, K. using self-blended photos to identify deepfakes. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 19–24, 2022, New Orleans, Louisiana, USA, pp. 18720–18729.
13. Learning Pairwise Interaction for Generalizable DeepFake Detection by Xu, Y., Raja, K., and Verdoliva, L. 3–7 January 2023, Waikoloa, Hawaii, USA; Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 672– 682. [Google Scholar]
14. Liu, L.; Song, Y.; Zhang, Y.; Wang, J.; Chen, L. Towards sound generalizations for deepfake detection using self-supervised learning of an adversarial example. 19–24 June 2022, New Orleans, LA, USA; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18710–18719. [Google Scholar]
15. Yuan, H.; Miao, C.; Liu, B.; Yu, N.; Zhuang, W.; Chu, Q. Towards intrinsic common discriminative features learning for adversarial learning-based face forgery detection. 18–22 July 2022, Taipei, Taiwan, China, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. [Google Scholar]