International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0094
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 585-590

Deep Fake Video Detection Using Transfer Learning Resnet50

S. Praveena [1], R.Kaviya [2] , K.Sheerin Farhana [3] S.Bhuvanasri [4]
[1]Professor, Department of AIML (Artificial intelligence and machine learning), Manakula Vinayagar Institute of Technology, Puducherry, India.
[2,3,4] Under Graduate student, Manakkula vinayagar institute of technology, Puducherry, India.
Email ID: *praveenacse11@gmail.com[1], kavirkaviyad12344@gmail.com[2], bhuvisri6124@gmail.com[3], sheerinfarhana38@gamil.com[4]*

Abstract
The rapid development of deep learning technologies has enabled the creation of highly realistic deepfake videos, raising concerns in areas such as media integrity, privacy, and security. Detecting these deepfakes has become a significant challenge, as conventional methods struggle to keep pace with increasingly sophisticated techniques. This journal explores the application of transfer learning using ResNet50, a pre-trained convolutional neural network, for deepfake video detection. We present an overview of deepfake creation, the role of ResNet50 in transfer learning, the implementation process, and the results of using this approach to detect deepfakes in video content.

Keywords: *Deepfake Detection, Transfer Learning, ResNet50, Convolutional Neural Network, Media Integrity.*

1. Introduction
In recent years, advancements in artificial intelligence and machine learning have enabled the creation of highly sophisticated fake media, particularly deepfake videos. Deepfakes utilize deep learning algorithms, specifically Generative Adversarial Networks (GANs) and Autoencoders, to generate hyper-realistic video content by swapping faces, manipulating facial expressions, or altering speech to create convincing yet fabricated scenarios. This technology has raised significant ethical concerns, as it can be exploited for malicious purposes, such as spreading misinformation, creating defamatory content, or manipulating public opinion. The implications of deepfake technology are profound, impacting areas such as journalism, politics, and social media. For instance, the potential to create realistic yet fabricated videos of public figures can undermine trust in media and influence public perceptions. In the realm of personal privacy, deepfakes can lead to identity theft or defamation, as individuals can be portrayed in compromising situations without their consent. The challenge of identifying these manipulated videos is becoming increasingly pressing as the techniques for creating deepfakes evolve, leading to an urgent need for effective detection methods. Traditional detection techniques often rely on analyzing digital artifacts or inconsistencies in video frames. However, as deepfake generation techniques become more advanced, these methods are frequently insufficient. There is a critical demand for robust detection systems that can identify deepfakes even when they appear nearly indistinguishable from authentic videos. Deep Learning, particularly through the use of Convolutional Neural Networks (CNNs), has emerged as a promising approach to tackle the challenges posed by deepfake detection. CNNs excel in visual tasks by automatically extracting features from images, which is essential for recognizing the subtle differences between real and manipulated video frames. Among various deep learning architectures, ResNet50 has gained prominence due to its ability to train very deep networks while maintaining high performance, thanks to its innovative residual connections. In this study, we leverage transfer learning to adapt the pre-trained ResNet50 model for deepfake detection. Transfer learning involves utilizing a model that has been

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0094
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 585-590

previously trained on a large dataset (such as ImageNet) and fine-tuning it on a smaller, task-specific dataset. This approach allows for efficient training and better performance, particularly when data is limited. By employing transfer learning with ResNet50, we aim to harness its ability to extract intricate features and patterns that characterize deepfake content, ultimately leading to improved detection accuracy. The objective of this journal is to explore the methodology of using ResNet50 for detecting deepfake videos, analyse the results of our experiments, and discuss the effectiveness and limitations of this approach. Through this research, we aim to contribute to the ongoing efforts to combat the proliferation of deepfake technology and safeguard the integrity of digital media. [1-3]

2. Methodology

The methodology for detecting deepfake videos using transfer learning with the ResNet50 model involves a systematic approach combining data preprocessing, model selection, training, and evaluation. First, a diverse dataset containing both real and deepfake videos is gathered from reliable sources such as the Deepfake Detection Challenge (DFDC), Face Forensics++, or Celeb-DF. These datasets typically contain thousands of videos that are either genuine or synthetically generated using advanced deepfake techniques. The videos are then pre-processed by extracting individual frames, as deepfake detection typically focuses on identifying subtle visual inconsistencies within each frame. Each frame is resized and normalized to meet the input requirements of ResNet50, which typically accepts images of size 224x224 pixels. Additional preprocessing steps may include face detection, cropping to focus on facial regions, and data augmentation techniques such as rotation, flipping, and zooming to enhance model generalization. Next, the pre-trained ResNet50 model is used for transfer learning. ResNet50, having been trained on the extensive ImageNet dataset, already possesses robust feature extraction capabilities from images. In this approach, the convolutional layers of ResNet50 are retained to leverage its feature extraction power, while the fully connected layers are either fine-tuned or replaced with new dense layers that are specifically tailored for deepfake classification. Fine-tuning involves adjusting the weights of the network to adapt to the deepfake detection task, improving the model's sensitivity to the subtle artifacts often found in manipulated videos. Once the model architecture is established, the training process begins. The model is trained on the pre-processed frames extracted from both real and deepfake videos, allowing it to learn to detect the unique patterns, distortions, and inconsistencies associated with deepfakes. Since the ResNet50 architecture is deep, comprising 50 layers with residual connections, it can effectively capture hierarchical features at different levels of abstraction, from basic edges to complex textures, which are crucial in identifying deepfakes. After training, the model is evaluated using a test set comprising unseen real and deepfake videos. The performance of the model is assessed through various metrics such as accuracy, precision, recall, F1-score, and the confusion matrix, which provides insight into the number of true positives, false positives, true negatives, and false negatives. To ensure robustness, cross-validation techniques may be used, where the dataset is split into multiple subsets to test the model's consistency across different data partitions. Once the model achieves satisfactory performance, it is deployed for real-world deepfake detection. Given a new input video, the trained model processes the video frame by frame, predicting whether each frame is authentic or a deepfake. The final decision for the entire video is based on an aggregation of frame-level predictions, with majority voting or probability-based techniques employed to classify the video as real or fake. Additionally, post-processing techniques such as temporal consistency checks across frames can further enhance the reliability of predictions by addressing any potential frame-wise inconsistencies. This deepfake detection system, combining transfer learning with ResNet50, enables efficient and accurate detection of manipulated videos, making it a valuable tool for combating the spread of misinformation and protecting the integrity of digital media. With further fine-tuning, the system can continue to improve its performance on evolving deepfake generation techniques, ensuring it remains effective against future deepfake advancement[4-

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0094
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 585-590

6](Figure 1 Deepfake Detection System)

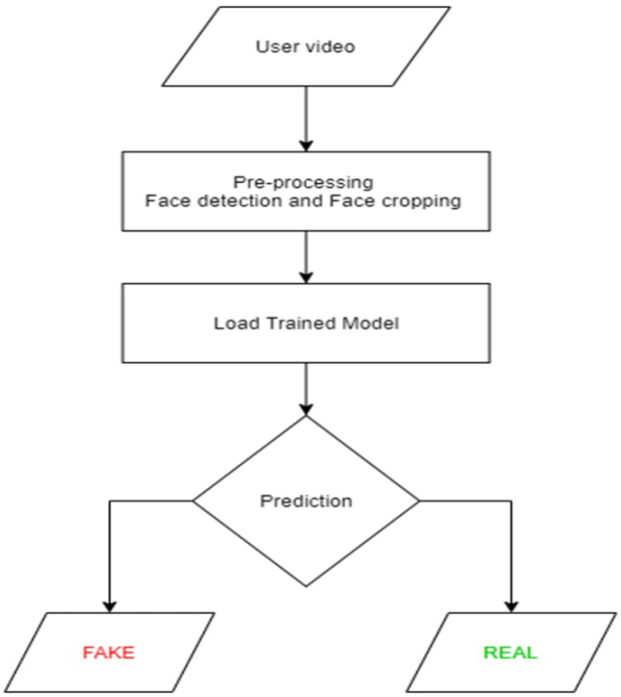

Figure 1 Deepfake Detection System

3. Literature Survey

The detection of deepfake videos has emerged as a critical area of research, driven by the rapid advancements in artificial intelligence and deep learning technologies. Deepfakes, which manipulate video and audio content using sophisticated algorithms, have posed significant threats to the authenticity of digital media. Various researchers have explored different methodologies, datasets, and models for detecting these manipulations, with particular focus on transfer learning and convolutional neural networks (CNNs) such as ResNet50. In early deepfake detection methods, researchers relied on handcrafted features and traditional machine learning techniques. Li et al. (2018) explored the use of visual artifacts, such as unnatural blinking patterns, to detect deepfakes. These methods, however, faced limitations due to the increasing sophistication of deepfake generation algorithms, which reduced the presence of easily detectable visual cues. As deepfakes became more advanced, researchers turned to deep learning-based methods, which automatically learn hierarchical features from video data. One significant contribution to deepfake detection is the use of convolutional neural networks (CNNs), which have shown remarkable success in image classification tasks. A study by Afchar et al. (2018) introduced a CNN-based approach called MesoNet, which was specifically designed for deepfake detection. While MesoNet demonstrated success in detecting manipulated faces, it lacked the generalization capability required for more complex, high-quality deepfakes. To address this, recent studies have leveraged more advanced architectures like ResNet, which can capture finer details in video frames. Transfer learning has emerged as a powerful technique in deepfake detection, allowing models pre-trained on large datasets, such as ImageNet, to be fine-tuned for deepfake classification. In 2020, Nataraj et al. demonstrated the use of ResNet-based architectures for detecting deepfakes, where the pre-trained ResNet50 model was fine-tuned on deepfake datasets such as FaceForensics++. This approach significantly improved detection accuracy compared to traditional methods, as the model could learn subtle features specific to deepfake manipulations, such as pixel-level inconsistencies and texture distortions. Several datasets have been crucial in advancing deepfake detection research. The FaceForensics++ dataset, introduced by Rössler et al. (2019), contains a diverse collection of manipulated and real videos and has become a standard benchmark for evaluating deepfake detection models. Additionally, the Deepfake Detection Challenge (DFDC) dataset, released by Facebook in collaboration with several academic and industry partners, provided an even larger and more varied dataset, allowing researchers to test their models on a wide range of deepfake techniques. These datasets have been instrumental in training deep learning models, improving their generalization across different types of manipulations and compression artifacts. While ResNet50 has proven effective for deepfake detection, other models, such as EfficientNet and Xception, have also been explored for this task. A study by Nguyen et al. (2020) compared several CNN architectures, including ResNet50, Xception, and EfficientNet, for their

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0094
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 585-590

performance in deepfake detection. The results indicated that while ResNet50 performed well, Xception provided slightly better accuracy due to its ability to capture even finer details in manipulated video frames. Nonetheless, ResNet50 remains a popular choice due to its balance between performance and computational efficiency. Another notable research trend is the application of recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks, to capture temporal information across video frames. Sabir et al. (2019) proposed a method that combines CNNs with LSTMs to leverage both spatial and temporal features in video sequences. This approach helps in detecting temporal inconsistencies in deepfake videos, which often arise due to poor frame transitions in generated content. The research also highlights the importance of generalization in deepfake detection models. A common challenge is ensuring that models trained on one dataset can generalize well to unseen data. A study by Tolosana et al. (2020) addressed this issue by evaluating the cross-dataset performance of different deepfake detection models, emphasizing the need for diverse and robust training datasets to ensure that detection systems remain effective across different types of deepfakes.[7-10]

4. Purposed System

The proposed system for deepfake video detection utilizes an advanced approach centered around transfer learning with the ResNet50 architecture to accurately identify manipulated videos. The system begins with the acquisition of a large dataset of both real and deepfake videos, sourced from well-established datasets such as the Deepfake Detection Challenge (DFDC), FaceForensics++, and Celeb-DF. The preprocessing phase involves extracting frames from each video, as deepfake detection largely relies on the analysis of individual frames to detect subtle visual inconsistencies. These frames are resized to the input size required by the ResNet50 model, typically 224x224 pixels, and normalized to standardize pixel values. The preprocessing pipeline may also include the detection and cropping of facial regions within the frames, as these areas are often the primary focus of deepfake manipulations. Data augmentation techniques like flipping, rotation, and zooming are applied to diversify the training data and prevent overfitting, which helps the model generalize better to unseen videos.[11-14] (Figure 2 Deepfake Manipulations)

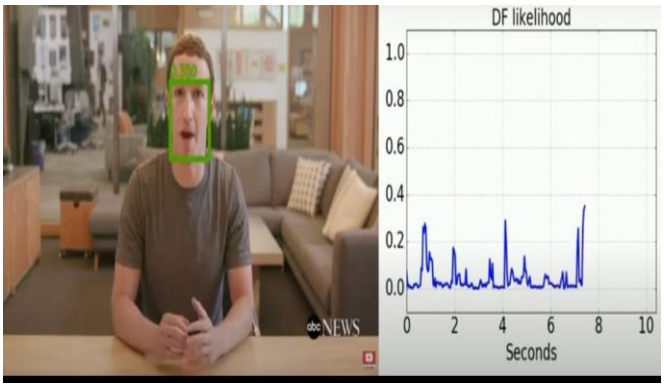

Figure 2 Deepfake Manipulations

The core of the system is the ResNet50 model, which is known for its deep residual network architecture. ResNet50 has already been pre-trained on the ImageNet dataset, which provides the model with a robust foundation for recognizing patterns and features in images. In this system, the convolutional layers of ResNet50 are retained to exploit their strong feature extraction capabilities, while the fully connected layers are modified to suit the deepfake detection task. The transfer learning approach allows the model to adapt to this specific task by fine-tuning the weights in a way that focuses on detecting unique deepfake artifacts, such as pixel-level distortions, unnatural facial expressions, and inconsistencies in lighting or shading that are often overlooked by human observers. This fine-tuning process involves retraining the model on the deepfake dataset, where the loss function—typically binary cross-entropy optimizes the model's ability to differentiate between real and manipulated frames. Once the model is trained, it is capable of processing new videos frame by frame, classifying each frame as either real or fake. A majority voting system is used to aggregate these frame-level predictions, where the most frequent class (real or fake) across the video frames determines the final classification of the entire video. To enhance the accuracy and robustness of the system, an optional component involving a Long Short-Term Memory (LSTM) network can be

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0094
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 585-590

integrated. This addition allows the system to not only analyze individual frames but also capture the temporal relationships between consecutive frames. By doing so, the model can identify inconsistencies in the motion or transitions between frames—another key characteristic of deepfake videos. This hybrid CNN-LSTM approach ensures that both spatial features (from ResNet50) and temporal dynamics (from LSTM) are taken into account, improving the system's performance on more complex and high-quality deepfakes. The evaluation phase of the system involves testing it on a separate dataset of real and deepfake videos that were not seen during training. The performance is measured using standard metrics such as accuracy, precision, recall, and F1-score, which provide a comprehensive view of the model's strengths and weaknesses in detecting deepfakes. Additionally, a confusion matrix is used to further analyze the model's classification performance by showing the number of true positives, false positives, true negatives, and false negatives. Cross-validation techniques can also be employed to assess the model's generalization across different datasets and deepfake generation techniques. This ensures that the system is not overfitted to a specific dataset and remains effective against evolving deepfake methods. The system is designed to be adaptable for real-time deepfake detection applications, where videos can be processed as they are streamed. To enable this, optimization techniques such as model quantization or pruning can be applied to reduce the computational load without sacrificing accuracy. This makes the system scalable and suitable for integration into platforms that require real-time deepfake monitoring, such as social media platforms, video-sharing websites, or media outlets that aim to combat misinformation. In summary, the proposed deepfake detection system is a highly efficient, accurate, and adaptable solution that leverages the strengths of ResNet50 and transfer learning, with optional LSTM integration to provide robust detection across a wide range of deepfake video manipulations. The system's ability to generalize across datasets and adapt to future advancements in deepfake technology makes it a valuable tool in the ongoing fight against synthetic media and its potential misuse[15-16]

Conclusion

The proposed system for deepfake detection, utilizing transfer learning with ResNet50, offers a highly effective solution for identifying manipulated videos. By leveraging the pre-trained model's powerful feature extraction capabilities and combining it with techniques like fine-tuning and optional LSTM integration, the system is able to detect subtle inconsistencies characteristic of deepfakes. Through rigorous preprocessing, data augmentation, and advanced classification techniques, the model ensures high accuracy and robustness across various datasets and evolving deepfake generation methods. The system's adaptability for real-time detection further enhances its practical applicability, making it suitable for deployment in platforms requiring efficient deepfake monitoring. Overall, this approach provides a scalable, reliable, and efficient method to combat the growing challenge posed by deepfake videos, offering a valuable tool for maintaining media integrity in the digital age.

Reference

[1]. Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(6), 1959-1971. https://doi.org/10.1109/TPAMI.2020.2970919

[2]. Roessler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1-11. https://doi.org/10.1109/ICCV.2019.00010

[3]. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 1251-1258. https://doi.org/10.1109/CVPR.2017.195

[4]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0094
e ISSN: 2584-2854
Volume: 03
Issue:03 March 2025
Page No: 585-590

Recognition (CVPR), 770-778. https://doi.org/10.1109/CVPR.2016.90

[5]. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Use of a Capsule Network to Detect Fake Images and Videos. arXiv preprint arXiv:1910.12467. https://arxiv.org/abs/1910.12467

[6]. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A Compact Facial Video Forgery Detection Network. IEEE International Workshop on Information Forensics and Security (WIFS), 1-7. https://doi.org/10.1109/WIFS.2018.8630761

[7]. Korshunov, P., & Marcel, S. (2019). DeepFakes: A New Threat to Face Recognition? Assessment and Detection. arXiv preprint arXiv:1812.08685. https://arxiv.org/abs/1812.08685

[8]. Wang, S., Wang, H., Zhang, L., & Zhang, J. (2020). Exploiting Temporal and Spatial Constraints in Deepfake Video Detection. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 16(1), 1-19. https://doi.org/10.1145/3394178

[9]. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2020). FaceForensics++: Benchmarking Deepfake Detection Tools. IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM), 3(4), 487-497. https://doi.org/10.1109/TBIOM.2020.297171

[10]. Tariq, S., Lee, J., Shahbaz, M. S., Huh, J. H., & Park, K. R. (2021). DeepFake Video Detection: A Survey. IEEE Access, 9, 154583-154614. https://doi.org/10.1109/ACCESS.2021.3129393

[11]. Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K. (2020). On the Detection of Digital Face Manipulation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5781-5790. https://doi.org/10.1109/CVPR42600.2020.00583

[12]. Li, Y., Chang, M.-C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 1-9. https://doi.org/10.1109/ICCVW.2018.00010

[13]. Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-Attention Generative Adversarial Networks. arXiv preprint arXiv:1805.08318. https://arxiv.org/abs/1805.08318

[14]. Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2020). The Deepfake Detection Challenge Dataset. arXiv preprint arXiv:2006.07397. https://arxiv.org/abs/2006.07397

[15]. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or Treat? Business Horizons, 63(2), 135-146. https://doi.org/10.1016/j.bushor.2019.11.0066. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980. https://arxiv.org/abs/1412.6980

[16]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative Adversarial Networks. arXivpreprintarXiv:1406.2661. https://arxiv.org/abs/1406.2661