Original Article
ISSN (Online): 2350-0530
ISSN (Print): 2394-3629

International Journal of Research - GRANTHAALAYAH
April 2025 13(4), 276–280

REVIEW PAPER ON DEEP FAKE DETECTION USING DEEP LEARNING

Shikha Singh [1], Shavez Khan [1], Chandan Pandey [1], Sandhya Sahani [1], Abhishek Singh [1], Vivek Patel [2]

[1] Students of Department of Computer Science and Engineering, KIPM College of Engineering and of Technology, India
[2] Assistant Professor of Department of Computer Science and Engineering, KIPM College of Engineering and of Technology, India

Received 09 March 2025
Accepted 12 April 2025
Published 30 April 2025

DOI
10.29121/granthaalayah.v13.i4.2025.6177

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2025 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License.

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

OPEN ACCESS

ABSTRACT

In the modern digital age, the emergence of artificial intelligence has made it possible to create hyper-realistic human faces that do not exist in reality. Such AI-generated images, popularly referred to as deepfakes, pose an increasing threat as they can fool the human eye and be used for nefarious activities.

The study centers on the identification of such forged images via deep learning, that is, CNNs. Unlike the majority of current research that mainly focuses on video-based deep fake detection, our method tackles the problem at image level. This is important, as even our one single doctored image can be used to spread false information, impersonate others, or breach security mechanisms.

We built a detection model able to pick out subtle artifacts and inconsistencies-numerically, in many cases invisible to the naked eye-performed by the process of generating an image.

Our intention is not only to design and performative models but also to help strengthen the general attempt at restoring and sustaining trust within digital material through developing available and precise fake image detection.

Keywords: Deepfake Images, AI-Generated Faces, Deep Learning, Image Forensics, CNN, Fake Content Detection

1. INTRODUCTION

Over the past few years, AI has been used to manipulate or synthetically produce media content, giving rise to the biggest threat in the form of deepfakes. They are media-in particular images and videos-that are real-looking but have been modified or completely produced by AI models. Although these technologies have positive applications in the entertainment and creative sectors, they also carry significant threats, including identity theft, disinformation, political propaganda, and cybercrime.

With social media becoming a primary source of information for a majority of the users, it becomes imperative to create tools that can distinguish between real and faked images. Most of the current deepfake work involves video detection

involving temporal information or audio-visual signals. Nevertheless, there is an important requirement for systems that can operate at the image level.

In this paper, we introduce a deep learning-based method based on Convolutional Neural Networks (CNNs) for identifying deepfake images. Our method detects embedded artifacts produced during AI-based face generation even if they are invisible to the human eye.

2. RELATED WORK

Different methods have been tried for deepfake detection. In Li et al. (2018), authors employed eye-blinking frequency for detecting fake videos, taking advantage of the abnormal blinking pattern in AI-generated videos. In Afchar et al. (2018), Maysonet was introduced as a CNN-based light network for face forgery detection.

Other authors such as Afshar et al. Nguyen et al. (2019) employed a CNN trained on face landmarks, whereas Nguyen et al. Agarwal and Farid (2020) employed capsule networks to identify tampered images. The challenge lies in constructing systems that are both efficient and accurate for various types of manipulated data.

Most of these techniques are cantered on video frames and involve time-series analysis. Our method, however, examines static images, allowing for faster and more scalable detection—critical for social media sites where images are shared broadly.

3. METHODOLOGY AND ALGORITHM

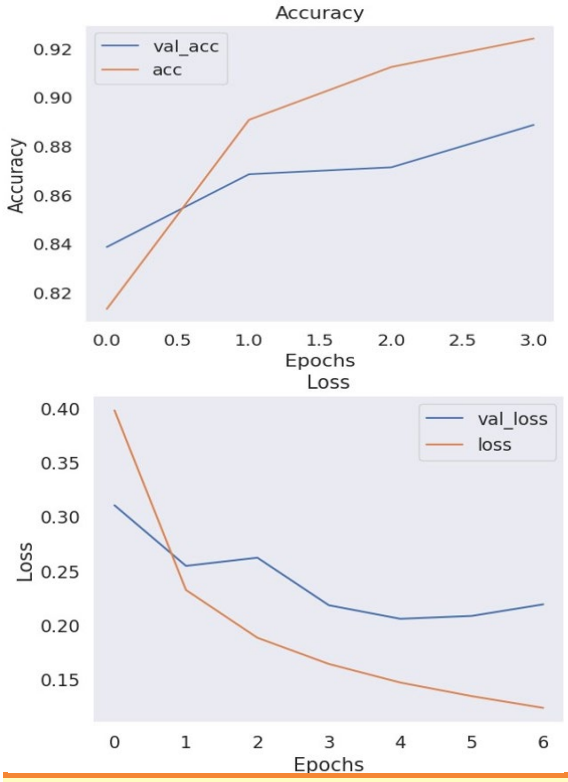
Our model adopts a CNN-based architecture that is trained to identify images as "real" or "fake." The pipeline consists of:

1) **Dataset Preparation:** We have taken organized data set from Kaggle.
2) **Preprocessing:** The images were resized to 64x64 and 128x128 and normalized. Rotation, flipping, etc., were some of the data augmentation techniques used to enhance generalization.
3) **Model Architecture:** A CNN was utilized with the following layers:
   - Convolution + ReLU
   - Max Pooling
   - Dropout
   - Fully Connected Layer
   - Sigmoid Output Layer
4) Training: The model was trained with binary cross-entropy loss and Adam optimizer for 25 epochs. Accuracy and loss were tracked on a validation set.

4. IMPLEMENTATION AND RESULTS PHASE BY PHASE

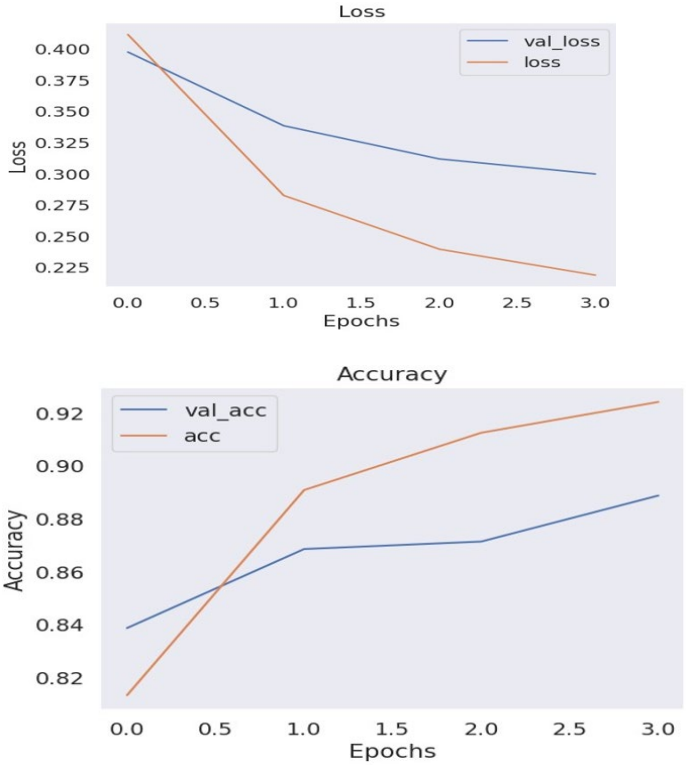
We implemented the model in Python with TensorFlow and Kera's frameworks. The training was done on Google Collab with the T4 GPU. The results are as follows:

Phase 1



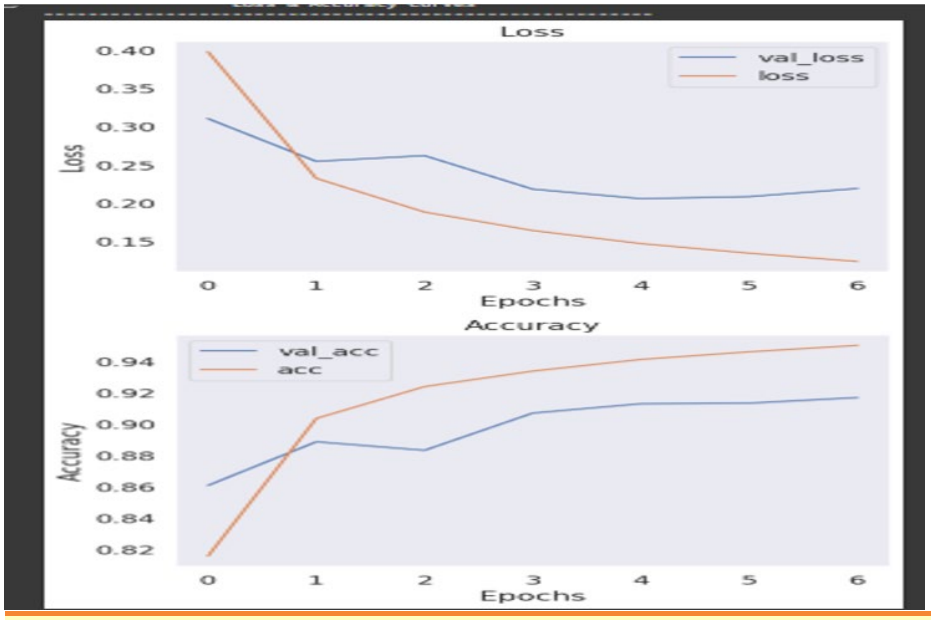
Phase 1

Phase 2

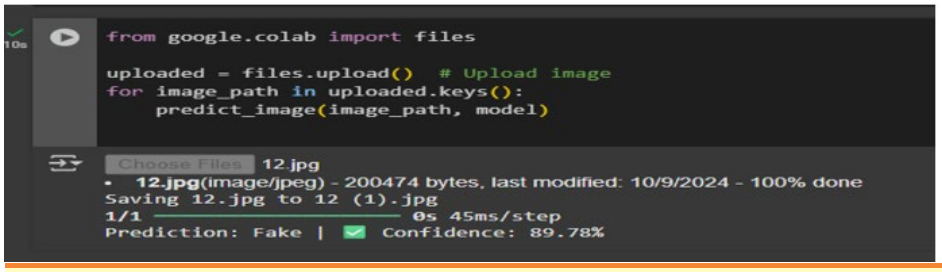


Phase 2

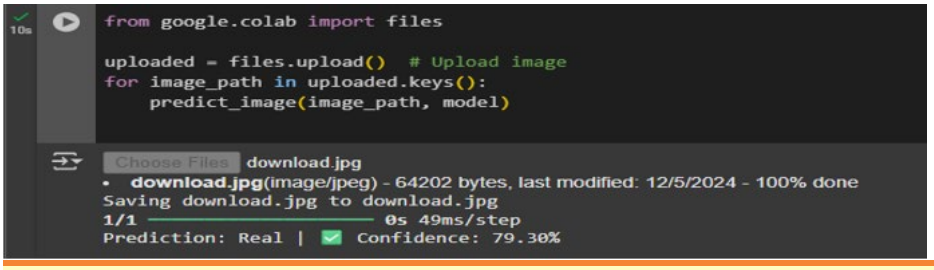
Phase 3



Phase 3 Final Improved Result



Phase 4 Training/Validation Accuracy Graphinsert Confusion Matrix or Result Image



Phase 5 Insert Sample Real Vs. Fake Detection Screenshots

These findings indicate that our model is able to effectively identify deepfake images with high accuracy and generalize to new data.

5. CONCLUSION AND FUTURE WORK

This work introduces a deep learning-based system to detect deepfake images. Our CNN model boasts more than 90% accuracy and is effective in discriminating between real and fake faces based on faint pixel-level artifacts.

In the future, we aim to:

- Expand the system to work with deepfake videos
- Test on larger and more varied datasets
- Use attention mechanisms or transformer-based architectures for better performance

As the fight against disinformation rages on, platforms such as ours help create a more reliable digital environment.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). Mesonet: A compact facial video forgery detection network. IEEE WIFS. https://doi.org/10.1109/WIFS.2018.8630761

Agarwal, N., & Farid, V. (2020). Detecting AI-generated images. arXiv preprint arXiv:2006.01567.

Chollet, F. (2017). Deep learning with Python. Manning Publications.

Flask Documentation. (2024). Flask Web Framework. Retrieved from https://flask.palletsprojects.com/

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization.

Li, Y., Chang, M., & Lyu, S. (2018). In ictu oculi: Exposing AI-generated fake face videos by detecting eye blinking. IEEE ICIP. https://doi.org/10.1109/WIFS.2018.8630787

Nguyen, H., Yamagishi, J., & Echizen, I. (2019). Capsule-forensics: Using capsule networks to detect forged images and videos. ICASSP. https://doi.org/10.1109/ICASSP.2019.8682602

OpenAI. (2024). Chat-based AI APIs. Retrieved from https://openai.com/

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

TensorFlow Documentation. (2024). TensorFlow Core. Retrieved from https://www.tensorflow.org/

Wolf, T., Debut, L., Sanh, V., et al. (2020). Transformers: State-of-the-art natural language processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. https://doi.org/10.18653/v1/2020.emnlp-demos.6