Real-Time Deepfake Detection

Arya Goyal
New York University
New York, USA
ag9961@nyu.edu

Manvi Pandya
New York University
New York, USA
mp6813@nyu.edu

Sanam Palsule
New York University
New York, USA
sp7940@nyu.edu

***Abstract*—The proposed project aims to develop a real-time AI-based tool for detecting and flagging deepfake content on social media platforms, focusing on images and videos. Utilizing OpenCV and deep learning models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers, the tool will identify visual and temporal inconsistencies commonly found in deepfakes. To enhance performance and scalability, GPU acceleration with CUDA and CuDNN will be employed for efficient processing and faster detection.**

I. Introduction

Deepfakes, or manipulated media content, pose a growing threat to online platform integrity. This project proposes an AI-based tool capable of real-time deepfake detection to support content moderators. By leveraging computer vision techniques and deep learning models, the tool will analyze both spatial and temporal data to flag potentially misleading content. Optimization through GPU acceleration aims to ensure rapid processing and real-time performance, making the tool effective for high-volume platforms.

II. Related Work

Recent advances in deep learning have significantly improved deepfake detection, shifting from traditional forgery detection methods to sophisticated approaches like CNNs, RNNs, and GAN-based models. Aparna et al. (2024) demonstrated that CNNs and LSTMs can achieve up to 97% accuracy in detecting face-swapped videos. However, challenges remain in generalization across datasets and robustness against adversarial attacks. Efforts to enhance detection resilience against adversarial attacks and improve generalization include contrastive learning frameworks (Li et al., 2021) and data augmentation techniques (Zhao et al., 2021). Self-supervised learning (Zhang et al., 2021) and few-shot learning (Cao et al., 2022) have addressed the scarcity of labeled data, allowing models to generalize better across unseen deepfake manipulations. Additionally, lightweight models such as those based on MobileNetV3 (Chandrasekaran et al., 2022) have enabled real-time detection on resource-constrained devices.

III. Methods

The proposed tool will employ the following methods:

- **Data Collection & Preprocessing:** Utilize OpenCV for image and video frame extraction, resizing, and normalization to prepare inputs for model training and inference.
- **Model Development:** Use CNNs for detecting spatial anomalies in individual frames, RNNs to analyze temporal inconsistencies across frames, and Transformers to capture complex patterns and relationships in the content.
- **Optimization with CUDA/CuDNN:** Accelerate the model's training and inference by leveraging GPU capabilities through CUDA and CuDNN libraries, improving processing speed and supporting real-time detection.
- **Integration and Testing:** Integration with APIs from social media platforms to analyze newly uploaded content for potential deepfakes. Test the system for accuracy, speed, and scalability on large-scale datasets.

IV. Conclusion

This project aims to create a real-time, scalable, and user-friendly AI tool for deepfake detection on social media platforms. By leveraging OpenCV, deep learning models, and GPU acceleration with CUDA/CuDNN, the system will effectively assist in moderating misleading content, ensuring quick identification and response to potential

References

- Aparna, S., Kumar, V., & Gupta, R. (2024). Face-swapped deepfake video detection using convolutional neural networks and long short-term memory networks. IEEE Transactions on Multimedia, 26, 1254-1265
- Li, J., Li, X., & Li, H. (2021). Contrastive learning for improved deepfake detection. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 4750-4759.
- Zhao, Y., Zhang, T., & Xu, G. (2021). Data augmentation for improved deepfake video detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3415-3424.