DETECTION OF SYNTHETIC MEDIA USING YOLO-POWERED OBJECT RECOGNITION

[1]R.Arthi , [2] G.Vishvanath Sundharam , [3]S Udhayashankar , [4]R Mathi , [5]R Shanmathi

[1,2]Assistant Professor, Department of CSE(Cyber Security), Muthayammal Engineering College.

[3,4,5]Student, Department of CSE(Cyber Security), Muthayammal Engineering College.

[1]arthiramcse@gmail.com, [2]vishvanathsundharam@gmail.com, [3]udhayaudhaya5345@gmail.com, [4]mathinila04@gmail.com, [5]shanmathishan712@gmail.com

ABSTRACT - To address the growing challenge of Deepfake and synthetic video manipulation, this project proposes a hybrid Convolutional Neural Network (CNN) architecture with ResNet as the backbone for effective feature extraction. The proposed approach leverages deep feature learning to automatically capture complex spatial and temporal patterns that distinguish real videos from forged ones. Inverted residual blocks and linear bottlenecks are incorporated to preserve spatial information while optimizing memory usage and computational efficiency. Advanced training techniques are employed to enhance detection accuracy and reduce inference time. By combining intensive learning phases with CNN-based feature classification, the system achieves high accuracy and efficiency in identifying forged videos, making it a robust solution for ensuring digital media integrity in real-world applications. The proposed model is suitable for real-time deployment in digital forensics, cybersecurity monitoring, and media verification systems.

KEYWORDS - Synthetic Media Detection, Deepfake Detection, YOLO, Object Recognition, Convolutional Neural Network (CNN), ResNet, Deep Learning, Digital Media Integrity, Cyber Security, Video Forensics.

1. INTRODUCTION

The rapid advancement of artificial intelligence has led to the widespread creation of synthetic media, especially Deepfake videos, which threaten digital media authenticity and cybersecurity. These videos use deep learning techniques to manipulate facial features, making manual detection difficult. To address this issue, the proposed system uses a hybrid deep learning approach with YOLO-powered object recognition and a ResNet-based CNN to extract spatial and temporal features. Architectural optimizations improve efficiency and enable accurate detection of real and forged videos.The proposed approach achieves high detection accuracy with reduced inference time, making it suitable for real-time applications in digital forensics, cybersecurity monitoring, and media verification systems.

1.1 SCOPE OF THIS PROJECT

This project aims to:

- To accurately detect synthetic and Deepfake videos using deep learning techniques.
- To utilize YOLO-powered object recognition for identifying and analyzing relevant facial and object regions in video frames.
- To extract spatial and temporal features using a ResNet-based CNN architecture for reliable classification.
- To reduce false positives and improve detection accuracy compared to traditional methods.
- To optimize memory usage and computational efficiency using inverted residual blocks and linear bottlenecks.
- To support near real-time video analysis with reduced inference time.
- To provide a scalable framework that can adapt to evolving Deepfake generation techniques.

To assist digital forensics, cybersecurity monitoring, and media verification applications by ensuring digital media integrity.

2. PROPOSED WORK

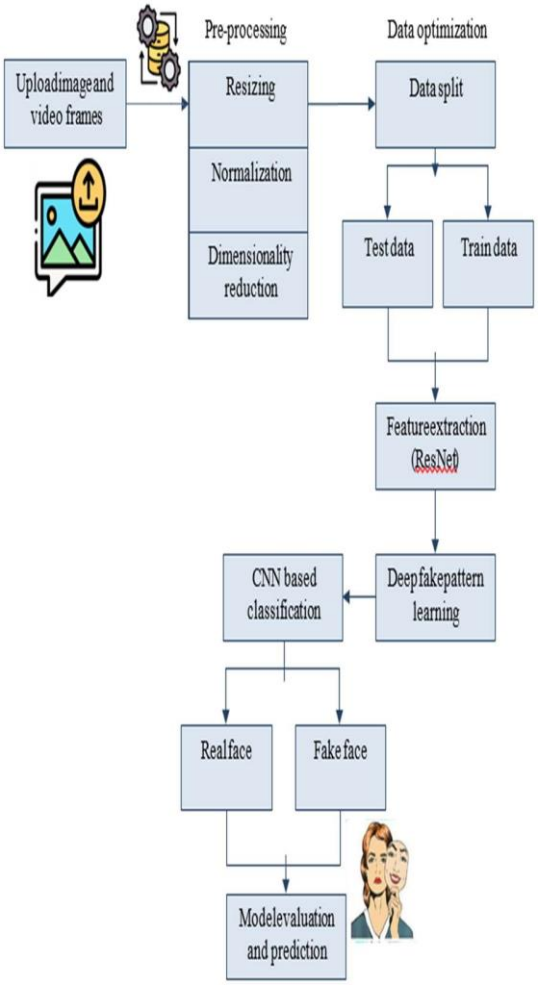
The proposed system detects synthetic media using a hybrid deep learning approach that combines YOLO-powered object recognition with a CNN architecture. YOLO is used to identify important facial and object regions in video frames, improving detection accuracy.By leveraging AI-driven security models and real-time data analysis, the system ensures robust protection against malicious attacks.

Video frames are extracted and preprocessed before feature extraction. A ResNet-based CNN captures spatial and temporal patterns to distinguish real videos from forged ones.

The extracted features are classified using CNN layers, enabling accurate Deepfake detection. The optimized architecture ensures fast inference and suitability for real-time media verification.

The system improves detection performance using inverted residual blocks and linear bottlenecks to reduce computational complexity while preserving key features. Advanced training techniques enhance accuracy and generalization. This enables efficient and real-time Deepfake detection for digital forensics and cybersecurity applications.

2.1 System Architecture



3.1 Data Processing & Preprocessing

The data processing and preprocessing stage prepares video input for accurate synthetic media detection. It includes the following steps:

- **Video Frame Extraction :** Converts input videos into individual frames while preserving temporal order.
- **Frame Resizing:** Adjusts frame dimensions to match CNN model input requirements.
- **Normalization:** Normalizes pixel values to improve model stability and accuracy.
- **YOLO-Based Region Detection:** Identifies facial and relevant object regions for focused analysis.
- **Noise Reduction:** Removes unwanted artifacts to enhance visual quality.
- **Data Batching:** Organizes processed frames into batches for efficient training and inference.

3.2 Attack Detection Techniques

3.2.1 YOLO – Based Region Detectio

YOLO is used to detect facial and important regions in video frames, enabling focused analysis of manipulated areas.

3.2.2 CNN – Based Classification

A ResNet-based CNN extracts deep features from detected regions and classifies videos as real or forged accurately.

3.3 TOOLS AND LIBRARIES

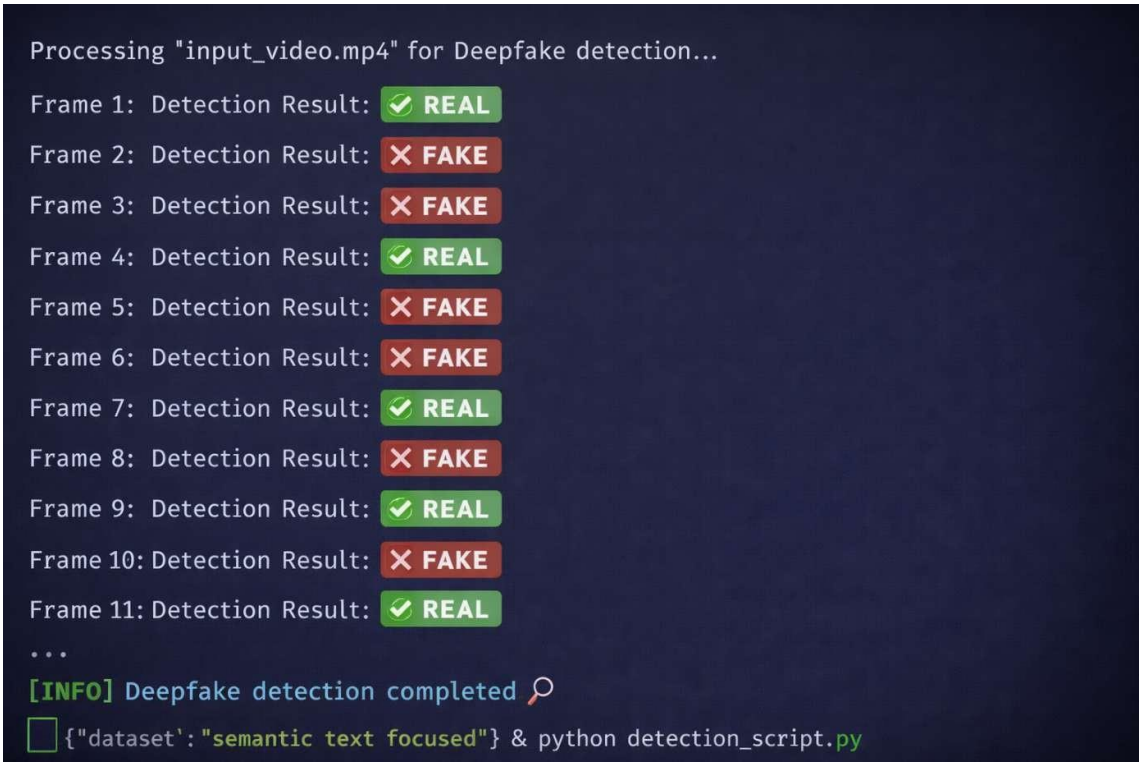
The proposed system utilizes the following tools and libraries to enable video analysis, YOLO-based region detection, deep learning model development, and efficient classification of real and forged videos.

- **Python:** Primary programming language for implementing the detection framework.
- **YOLO:** Used for object and facial region detection in video frames.
- **OpenCV:** Handles video processing, frame extraction and operations.
- **TensorFlow / Keras :** Deep learning framework for building and training CNN models.
- **ResNet:** Backbone network for deep feature extraction.
- **NumPy:** Supports numerical computations and array operations.

4. PROGRAM

```python
import cv2
import numpy as np
from keras.models import load_model
from yolov5 import YOLO

# Load YOLO model and ResNet-based CNN
yolo_model = YOLO('yolov5s-face.pt')
cnn_model = load_model('resnet_deepfake.h5')

# Process video frames
cap = cv2.VideoCapture('input_video.mp4')
while True:
    ret, frame = cap.read()
    if not ret:
        break

    detections = yolo_model(frame)
    faces = [det for det in detections if det['class'] == 'face']
    features = cnn_model.predict(np.array(faces))
    result = 'Fake' if np.argmax(features) == 1 else 'Real'
    print("Detection Result:", result)

cap.release()
```

5. RESULTS AND OUTPUT



Results and Output : The system effectively detects Deepfake videos with high accuracy and efficiency.

- **Detection Results :** The system classifies video frames as Real or Fake with good accuracy.
- **Performance Evaluation :** The system achieves high detection accuracy with reduced inference time, making it suitable for real-time digital media verification.

6. CONCLUSION

This project presents an efficient approach for detecting synthetic media using YOLO-powered object recognition combined with a ResNet-based Convolutional Neural Network. The system identifies Deepfake videos by analyzing spatial and temporal features from video frames. Optimized architectural components improve accuracy while reducing computational complexity and inference time. Overall, the proposed solution supports real-time synthetic media detection for applications in digital forensics, cybersecurity, and media verification.

REFERENCES

[1] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for Deepfake Forensics," *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 3207–3216, 2020.

[2] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection: A Survey," *Computer Vision and Image Understanding*, vol. 223, pp. 1–25, 2022.

[3] Y. Nirkin, Y. Keller, and T. Hassner, "FSGANv2: Improved Subject-Agnostic Face Swapping and Reenactment," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 560–575, 2022.

[4] R. Wu, G. Zhang, S. Lu, and T. Chen, "Cascade EF-GAN: Progressive Facial Expression Editing with Local Focuses," *Proc. IEEE/CVF CVPR*, pp. 5021–5030, 2020.

[5] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the Latent Space of GANs for Semantic Face Editing," *Proc. IEEE/CVF CVPR*, pp. 9243–9252, 2020.

[6] A. Rossler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 1–11, 2019.A. Seaborne and E.

[7] Z. Wang, Y. Guo, and W. Zuo, "Deepfake Forensics via an Adversarial Game," *IEEE Trans. Image Processing*, vol. 31, pp. 3541–3552, 2022.

[8] Y. Huang, F. Juefei-Xu, Q. Guo, Y. Liu, and G. Pu, "FakeLocator: Robust Localization of GAN-Based Face Manipulations," *IEEE Trans. Information Forensics and Security*, vol. 17, pp. 2657–2672, 2022.

[9] I. Goodfellow et al., "Generative Adversarial Nets," *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2672–2680, 2014.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE CVPR*, pp. 770–778, 2016.

[11] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[12] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *Proc. IEEE CVPR*, pp. 779–788, 2016.

[13] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," *Proc. IEEE WIFS*, pp. 1–7, 2018.