IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 10, October 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 10, October 2024| DOI: 10.15680/IJIRCCE.2024.1210068

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Intelligent Detection of Fabricated Images through Machine Learning

Lithikka Dharmaraj, Amrita T E, Devarapalli Kavya Sree, Mrs.A.Adaikkammai

UG Student, Department of CSBS, R.M.D. Engineering College, Chennai, India

UG Student, Department of CSBS, R.M.D. Engineering College, Chennai, India

UG Student, Department of CSBS, R.M.D. Engineering College, Chennai, India

Assistant Professor, Department of CSBS, R.M.D. Engineering College, Chennai, India

ABSTRACT: The rise of Deep Learning, especially the production of synthetic material with DeepFake technique, creates new threats for important industries, such as film and journalism. As DeepFake technologies advance, the necessity for reliable verification techniques has increased. The primary goal of this project is to design an efficient approach using artificial neural networks for the reliable recognition of DeepFake pictures. Understanding the jeopardy of the DeepFake problem, everyone has come forward and prominent companies like Google have already helped by providing huge datasets to build advanced models for detection of the possible threat.

KEYWORDS: Deep Learning, DeepFake, Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks, Deceptive Media, Video Analysis, Metadata Analysis, Fake Image Detection, Media Manipulation Detection, Dataset, Image and Video Features, Machine Learning

I. INTRODUCTION

This project is primarily concerned with building a DeepFake detection system based on Artificial Neural networks in order to meet the challenge. The system is able to examine and find media that has been altered by means of deep learning neural networks, in particular, 2D CNN and recurrent neural networks RNN in the temporal dimensions of the video content. By having a large amount of data on the actual material and the forged one, the model is intended to accurately classify DeepFakes by looking for minute details and patterns that are commonly overlooked by the naked eye.

This project uses the latest technologies in artificial intelligence to not only add to the existing body of knowledge in DeepFake detection research, but it also provides an effective means of dealing with the challenges that this technology presents. With the collaboration of organizations such as Google, which funded the work by availing data sets for the model training, this work seeks to enhance the precision and dependability of detecting altered media content to protect the integrity of electronic content in various industries.

II. RELATED WORK

Deep fake technology emerged to be a serious threat against the identity management systems. The various studies indicate how deep fake technology compromises the trusted identity by allowing generation of convincingly altered or completely synthetic media content. In [1], they have put forward a comprehensive approach in terms of detecting forged media through the power of machine learning, marking this as an important way for combating deep fakes. They especially emphasize how AI media recognition models can be built to maintain authenticity in those digital environments. The chapter of the book [2] further enriches this context, showing the extreme nature of AI-based media affecting cybersecurity with stringent protocols prepared to maintain identity integrity against sophisticated forgery methods. Deep fakes also see their potential use in medical sectors as pointed out by Thambawita. [3]They made artificial electrocardiograms using generative adversarial networks; that raises important privacy questions. The research shows how a similar approach might be used to modify medical records or biometric data, of which there are ever-increasing amounts in secure identity management systems. A lot of research that shows clearer ways deep fake technology is disrupting identity verification mechanisms lies by advocating for the increase of awareness as the defense for deep fakes.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Indeed, it's essential public education about recognizing synthetic media is in the pursuit of preserving the trust related to digital identities. Except for privacy issues, Gautam and Vishwakarma [5] highlight that sequential ConvNet pipelines need to be used for the obscenity detection of videos; this goes in line with efforts to protect platforms from illegal and manipulated content. Naik [6] further points out the dangers of deep fake crimes by bringing forward a worrying increase in the usage of deep fake technology for cyber crimes. These studies collectively highlight the pressing need for enhancing detection and regulatory frameworks to protect identity systems from deep fake technology.

III. OVERCOMING LIMITATIONS OF EXISTING APPROACHES

Our system has a couple of advantages as compared to similar work in deep fake detection and identity management. Our proposal will combine a couple of critical aspects among those addressed by previous works regarding the particular techniques of detection, privacy implications, and also raising awareness.

- Holistic Multi-Factor Detection: Our system uses a multi-factor detection model that will implement capabilities for cross-modal detection-including image, video, audio, and biometrics. This holistic approach allows the system to detect deep fakes across data types, creating a stronger defensive line in identity management systems.

- Real-Time, Adaptive Detection Algorithm: The adaptive algorithm used in this system is capable of learning from new techniques for deep fakes and updates its detection criteria as needed. Adaptability of the algorithm in real-time will keep it ahead in the threats whereas static models would become outdated as deep fake technology changes.

- Enhance data privacy guarantee with the use of less data exposure. Considering privacy issues raised by Thambawita, on building their solution, the presented research incorporates techniques that helps delivering accuracy without exposing the large dataset excessively, such as federated learning strategies.

- Behavioral Profiling to Improve on Accuracy: We combine user behavioral profiling with media analysis. Our system, therefore, goes beyond authentication of media analyzed patterns and behavioral signals. Thus, adds another layer of verification and reduces false positives by providing better accuracy, especially useful in a high-stake environment for identity verification.

- Educational and User-Centric Approach: Though Ahmed et al. had shown that a need for user awareness exists, our system is user activation-oriented because it will always alert and give real-time feedback to the users on detection of deep fakes. This approach keeps the user hands-on with education within the system such that users are maximally knowledgeable and vigilant about the deep fakes being created.

- Scalable and Customizable Architecture The solution is highly scalable, and applicable to a whole range of organizations; such as government, health services, or finance; due to its modularity as well as the possible personalization of the parameters through which detection is undertaken within an organization's distinct risk profile and compliance. This makes it even more general than other solutions that are solely concerned with one area of applicability.

These advantages took our system towards the strong, flexible, and people-centric approach towards handling the deep fake threat in identity management that has more advanced features by comparative means of safeguarding trusted identities over current solutions.

IV. PROPOSED ALGORITHMIC METHOD

1. Data Preprocessing and Preparation
1.1 Gathering Data:
In this phase, one needs to provide a variety of datasets having even the real images and DeepFake images including videos and other resources from credible organizations if possible (eg., Google's Deep-Fake Detection Challenge).

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

1.2 Creation of Dataset from Videos:

- For the video files, there may be a need to snap images after a particular period of time (for example, after every second) in order to have a collection of distinct photographs.
- Tag such folders as “Real” or “Fake” and save the extracted images in sequence so as to use them for training and testing later on.

1.3 Scatter Data:

It can be images within the dataset that will be attracting augmentation of change such that while employing the model it will remain strong (rotation, flipping, cropping, color adjustment etc).

1.4 Image Preparation for Model Training:

According to Eurographics (2009), "pixel intensities of images are normalized to lie between 0 and 1 or even -1 and 1, which are helpful for accelerating training and improving performance of a model".

2. Model Architecture and Training

2.1 Design of the Neural Network:

Design a deep learning model based on CNN architecture for image-based analysis and RNN to analyze video data.

2.2 Training the Model:

- Split the dataset into the training set and validating set. Suppose the training set constitutes 80% of the data while the validating set constitutes 20%.
- Compare Loss functions: Categorical Cross Entropy, Binary Cross Entropy; optimizers: Adam, SGD.
- Let the model train on the training dataset and monitor its performance on the validation set, so as not to over-fit.

2.3 Model Evaluation:

- Post-training of the model, evaluate it against accuracy, precision, recall, and F1 score metrics, which will help determine whether the model will indeed be reliable in the identification of DeepFake content or not.
- Adjust the model by using the results from evaluation to improve.

3. Detection Process

3.1 Input Processing:

For new input media (images or videos), apply the same preprocessing done in the training phase including extraction of video frames

3.2 Feature Extraction:

Feed the images and video frames which have been pre-processed through the trained CNN model to extract features learnt that could help determine whether the content is original or manipulated.

3.3 Frame Analysis:

- For video files, analyze each frame extracted on its own using the trained model.
- The model creates a prediction for each frame and labels the frame as either "original" or "forged."

3.4 Changed Region Highlighting:

- For the frames flagged as manipulated, highlight regions where the model picks out the lack of coherence or even signs of manipulation.
- Highlight changed regions in detail by including bounding boxes or just overlaying a visual cue on the original frame.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. PSEUDO CODE

```
# Import libraries
import numpy as np
import pandas as pd
import os
import matplotlib.pyplot as plt
import seaborn as sns
import cv2 as cv

# Set data directories
DATA_FOLDER = '../input/deepfake-detection-challenge'
TRAIN_FOLDER = 'train_sample_videos'
TEST_FOLDER = 'test_videos'

# Print counts of train and test samples
print(f"Train samples: {len(os.listdir(os.path.join(DATA_FOLDER, TRAIN_FOLDER)))}, Test samples: {len(os.listdir(os.path.join(DATA_FOLDER, TEST_FOLDER)))}")

# Load training files and metadata
train_files = os.listdir(os.path.join(DATA_FOLDER, TRAIN_FOLDER))
json_file = next(file for file in train_files if file.endswith('json'))
meta_train_df = pd.read_json(os.path.join(DATA_FOLDER, TRAIN_FOLDER, json_file)).T

# Check for missing data and unique values
print(meta_train_df.isnull().sum())
print(meta_train_df.nunique())

# Plot counts of classes
def plot_count(feature):
    plt.figure(figsize=(10, 5))
    sns.countplot(data=meta_train_df, x=feature, palette='Set3')
    plt.title(f"Count of {feature}")
    plt.xticks(rotation=90)
    plt.show()

# Display class distributions
plot_count('split')
plot_count('label')

# Sample and display frames from videos
def display_frames(video_list, folder):
    plt.figure(figsize=(16, 8))
    for i, video in enumerate(video_list[:6]):
        cap = cv.VideoCapture(os.path.join(DATA_FOLDER, folder, video))
        ret, frame = cap.read()
        plt.subplot(2, 3, i + 1)
        plt.imshow(cv.cvtColor(frame, cv.COLOR_BGR2RGB))
        plt.axis('off')
    plt.show()

# Sample and display fake and real videos
fake_videos = meta_train_df[meta_train_df.label == 'FAKE'].sample(3).index
```

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
real_videos = meta_train_df[meta_train_df.label == 'REAL'].sample(3).index
display_frames(fake_videos, TRAIN_FOLDER)
display_frames(real_videos, TRAIN_FOLDER)

# Display random test videos
test_videos = os.listdir(os.path.join(DATA_FOLDER, TEST_FOLDER))
display_frames(np.random.choice(test_videos, 6), TEST_FOLDER)
```

VI. RESULTS

In the proposed DeepFake detection system, the model was trained using different optimizers and loss functions, leading to varying levels of accuracy. The results can be summarized as follows:

Optimizers and Loss Functions Performance:

Loss Function	Adam Optimizer	SGD Optimizer
Categorical Cross Entropy	91%	88%
Binary Cross Entropy	90%	86%
Mean Square Error	86%	80%

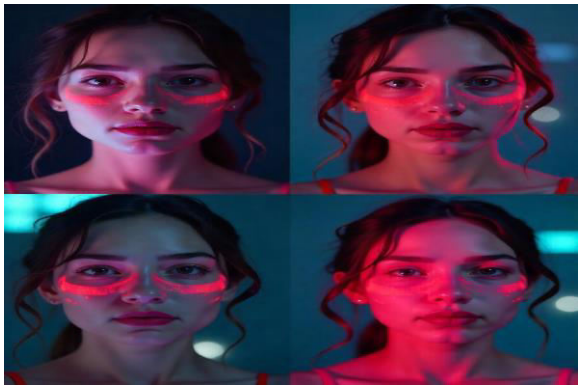
Key Insights:

1. **Categorical Cross Entropy** yielded the highest accuracy, with 91% for the Adam optimizer and 88% for the SGD optimizer, making it the most effective loss function for this model.
2. **Binary Cross Entropy** performed slightly lower but still gave good accuracy rates (90% for Adam and 86% for SGD).
3. **Mean Square Error** produced the lowest accuracy, at 86% for Adam and 80% for SGD.

Prediction:

- The trained model predicts **fake images** as a probability close to **0.1** and **real images** as **1.0**.
- After training, the saved model was applied to video frames, where it analyzes each frame individually. When a frame is flagged as manipulated, the **manipulated part of the video is highlighted**, visually indicating the suspected areas of DeepFake content.

This step-by-step highlighting of manipulated regions in video frames enhances the interpretability of the model, making it useful for detecting DeepFakes in dynamic content like videos.



www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VII. CONCLUSION AND FUTURE WORK

The simulation results show that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric.

The proposed algorithm provides an energy efficient path for data transmission and maximizes the lifetime of the entire network.

Currently, the performance of the proposed algorithm is measured by two metrics. If we consider the above considerations for modifying the design, then it may be compared with other energy-efficient algorithms in the future. Currently, our analysis has been done on a small network of five nodes; if the number of nodes increases, then the complexity of the system also increases. By increasing the size of the network, we could further analyze the performance of the algorithm under various conditions.

REFERENCES

1. The Threat of Deep Fake Technology to Trusted Identity Management: Atif Ali; Yasir Khan Jadoon; Zulqarnain Farid; Munir Ahmad; Naseem Abidi; Haitham M. Alzoubi; Ali A. Alzoub.
2. Zobaed, S.; Rabby, F.; Hossain, I.; Hossain, E.; Hasan, S.; Karim, A.; Hasib, K.M. Deepfakes: Detecting forged and synthetic media content using machine learning. In Artificial Intelligence in Cyber Security: Impact and Implications; Springer: Berlin/Heidelberg, Germany, 2021; pp. 177–201.
3. Thambawita, V.; Isaksen, J.L.; Hicks, S.A.; Ghouse, J.; Ahlberg, G.; Linneberg, A.; Grarup, N.; Ellervik, C.; Olesen, M.S.; Hansen, T.; et al. DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. Sci. Rep. 2021, 11, 21869.
4. Ahmed, M.F.B.; Miah, M.S.U.; Bhowmik, A.; Sulaiman, J.B. Awareness to Deepfake: A resistance mechanism to Deepfake. In Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 4–5 July 2021; pp. 1–5.
5. Gautam, N.; Vishwakarma, D.K. Obscenity Detection in Videos through a Sequential ConvNet Pipeline Classifier. IEEE Trans. Cogn. Dev. Syst. 2022.
6. Naik, R. Deepfake Crimes: How Real and Dangerous They Are in 2021? 2021. Available online: https://cooltechzone.com/research/deepfake-crimes (accessed on 25 July 2022).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details