27th International Conference on Computer and Information Technology
December 20-22, 2024, Cox's Bazar, Bangladesh

Paper ID:1028

Advancing Deepfake Detection: Integrating Time Series Processing With Vision Transformers

Srabonti Deb, Tasmia Jannat, Mohiuddin Ahmed
Department Of Computer Science & Engineering
Rajshahi University Of Engineering & Technology, Rajshahi-6204, Bangladesh

Presented By

Srabonti Deb
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology

21st December, 2024

Table of Content

Introduction

DeepFake

Real image

Manipulated image

Synthetic media -

Manipulated images-

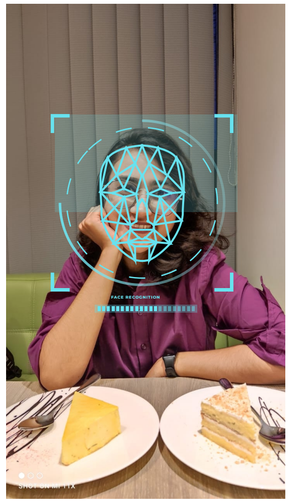
Altered videos or audio -

Generated by AI -

Introduction

DeepFake

Real Image



Mask

Fake Image

Traditional Deepfaking Approach

- DeepFake
- Face2Face
- FaceSwap
- Neural Textures
- FaceShifter

Motivation

DeepFake Detection

Misinformation Spread:
-Address the spread misinformation and false narratives.

Legal and Ethical Complications:
-Render unreliable evidence
-Compromise the fairness of legal proceedings.

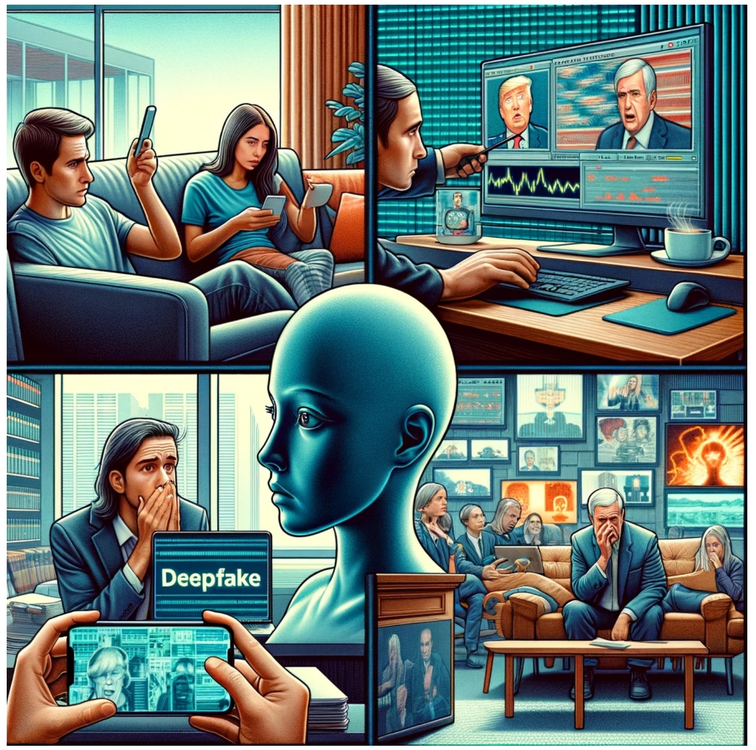


Figure-1: Deepfake problems

Cybersecurity Threats:
-Underscores the risks associated with cybercrime,
-Identity theft
-Fraud.

Social and Emotional Impact:
-Cause emotional distress and
-Societal harm.

A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer [1]

Contribution

- Introduces a convolutional layer adaptable to various forms of image manipulation.
- Reduces overfitting problem.
- Has low computational cost.

Limitations

- Encounter inconsistencies in handling time series data.
- Sensitivity to data noise.
- Struggle to detect video manipulations.

MesoNet: a Compact Facial Video Forgery Detection Network [2]

Contribution

- Implements a compact CNN architechture.
- Focuses on mesoscopic properties, such as textures.
- Optimized for low computational complexity, making it suitable for resource-constrained environments.

Limitations

- Struggles with high quality or subtle modification in manipulated videos.
- Fails to capture spatial-temporal information.

DeepFake Video Detection through Facial Sparse Optical Flow based Light CNN [3]

Contribution

- Detects manipulations using sparse optical information flow.
- Enhances CNN model's performance by focusing on motion inconsistencies.
- Develops a highly compact and low-dimensional light CNN model.

Limitations

- Unable to handle static facial manipulations, such as subtle texture changes or lighting alterations that don’t involve motion.
- Exhibits higher false positive rates, misclassifying unaltered videos.

Challenges

Large Datasets Processing

Overfitting issues

Biasness over any manipulation technique

Crossing local minima

Objectives of Our Study

To ensure the integrity and ethical standards of multimedia content.

To enhance temporal detection.

To utilize batch processing for time series data handling.

To implement a vision transformer model.

To ensure comprehensive testing.

To design an optimized model.

Working Dataset

FaceForensics++ [4]

- 1,000 Original Videos and
  1000 manipulated videos for each technique.
- Manipulated by Five Automated
  Techniques
- Total Size - 509 GB
- Four versions: Raw, High, Medium, Low

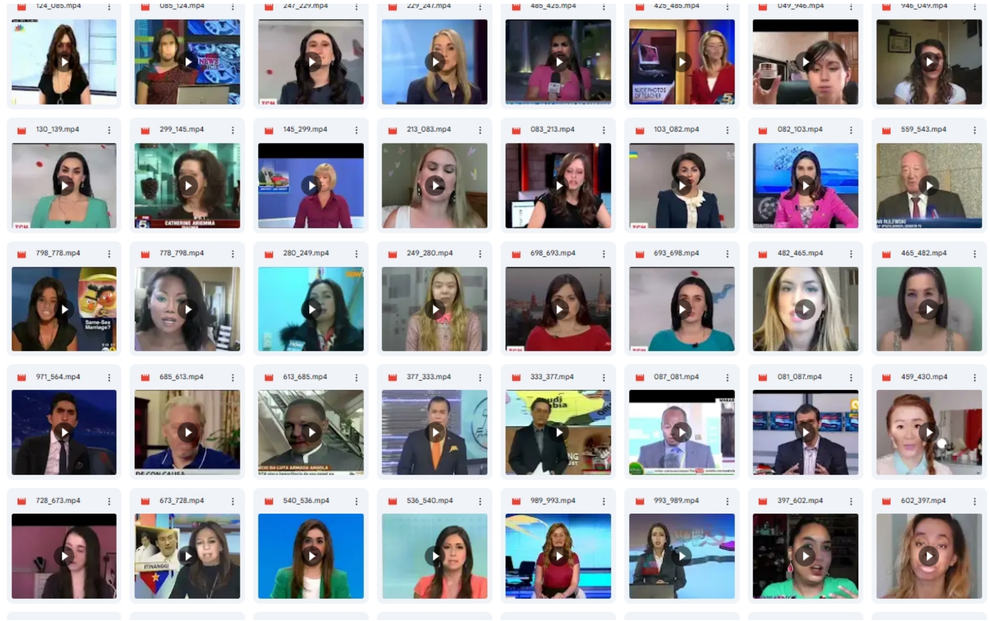
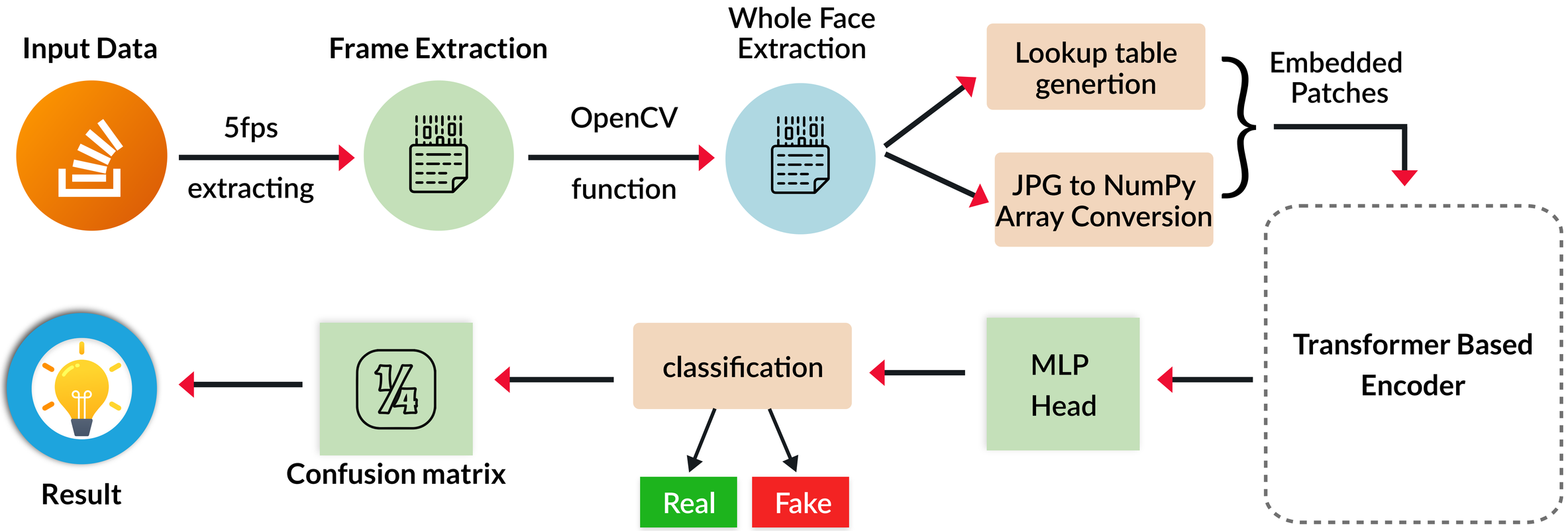


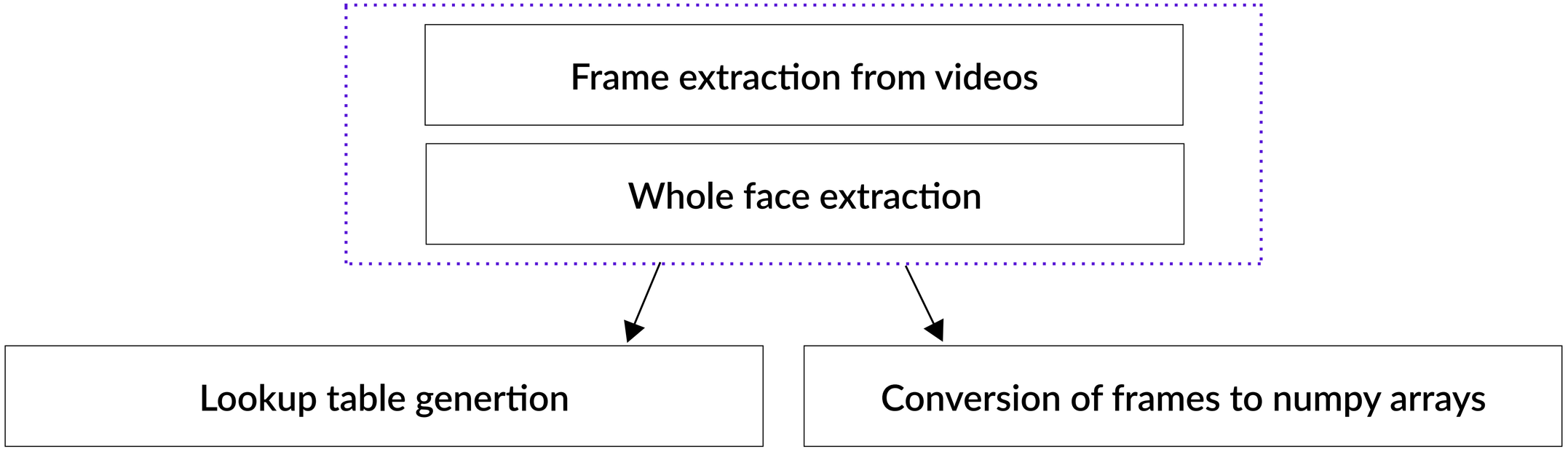
Figure-2:FaceForensics++ Dataset

System Overview



Preprocessing Overview

| Steps



Frame Extraction

| Steps

-Frame selection from videos. Frame selection frequency-5fps.

-Conversion of selected frames to grayscale.

-Facial region detection.

-Cropping and Resizing.

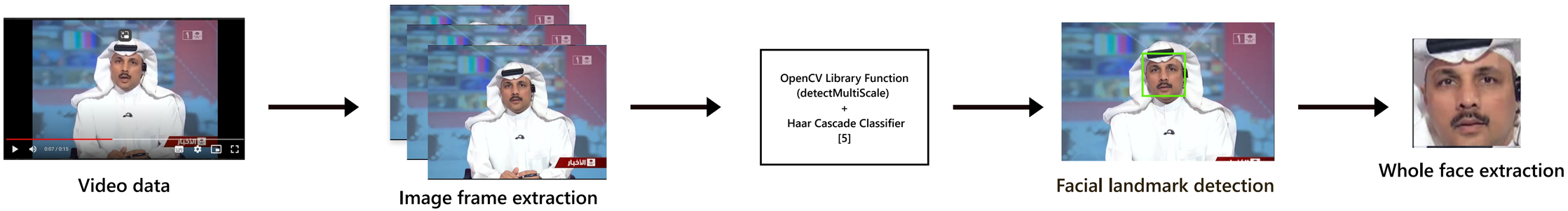


Figure 4: Frame Extraction from Video Input

Lookup Table Generation

	A	B	C	D	E	F	G
1	sample_index	video_index	video_title	video_type	type_label	part_num	part_frames
2	22496	4937	496	5	og_videos	5	20
3	4128	939	169_227	1	Face2Face	3	4
4	13514	2992	551_631	3	FaceSwap	4	20
5	8724	1943	291_874	2	FaceShifter	2	15
6	9055	2005	353_383	2	FaceShifter	2	20
7	21239	4679	238	5	og_videos	5	20
8	18741	4123	682_669	4	NeuralTextures	5	20
9	747	168	168_222	0	Deepfakes	4	20
10	7473	1663	011_805	2	FaceShifter	0	15
11	6452	1433	750_743	1	Face2Face	0	20
12	7868	1758	106_198	2	FaceShifter	2	17
13	17429	3844	403_497	4	NeuralTextures	5	18
14	5815	1299	588_556	1	Face2Face	3	20
15	18248	4019	578_636	4	NeuralTextures	0	20
16	13534	2997	556_588	3	FaceSwap	3	15
17	9979	2198	546_621	2	FaceShifter	2	20
18	4524	1028	283_313	1	Face2Face	5	20
19	13327	2952	511_472	3	FaceSwap	6	20
20	18483	4068	627_658	4	NeuralTextures	3	15

Figure 5: Sample Lookup Table

Architechture of Implemented Model

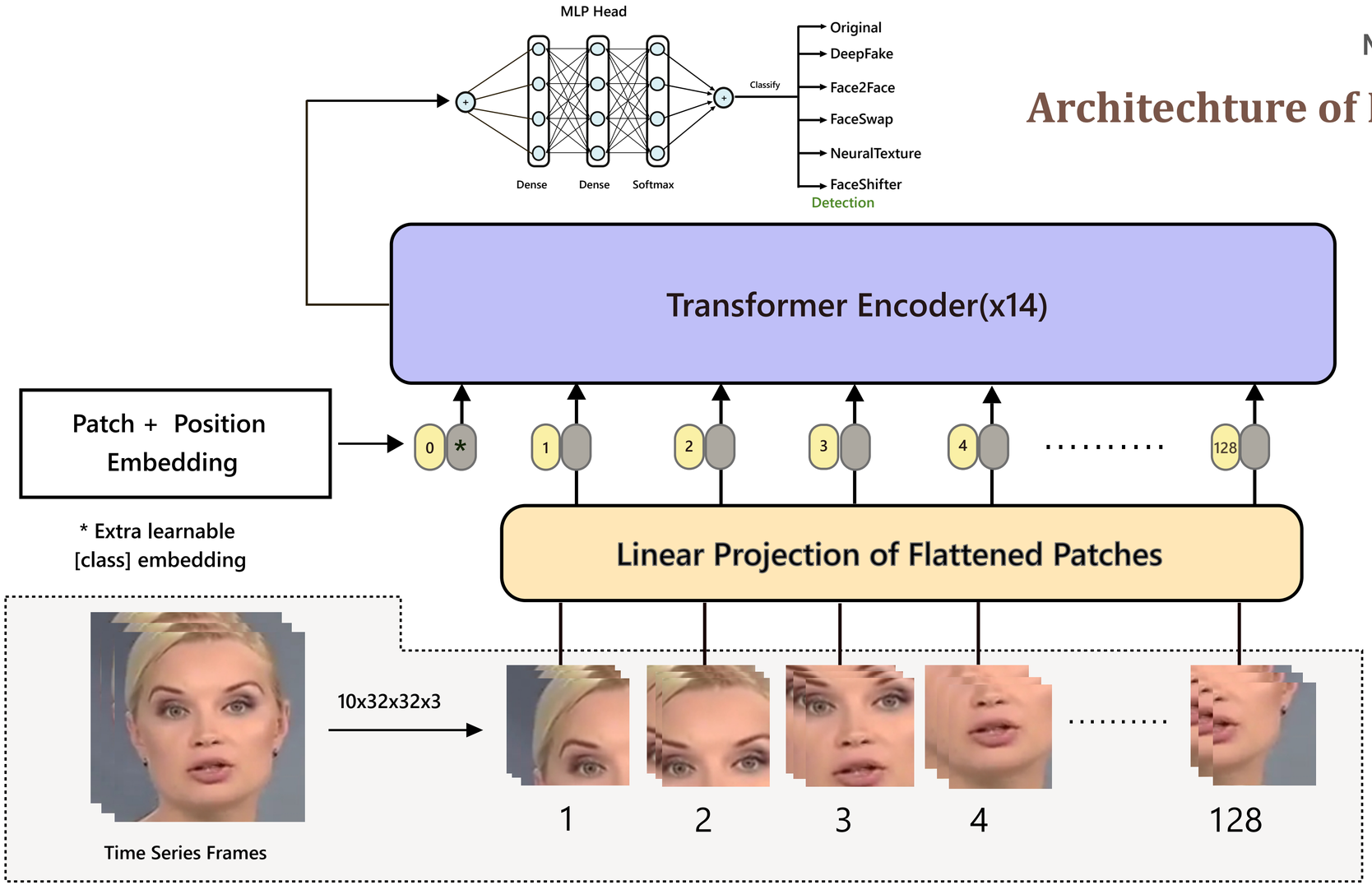


Figure-6: Block diagram of our implemented model

Encoder Block

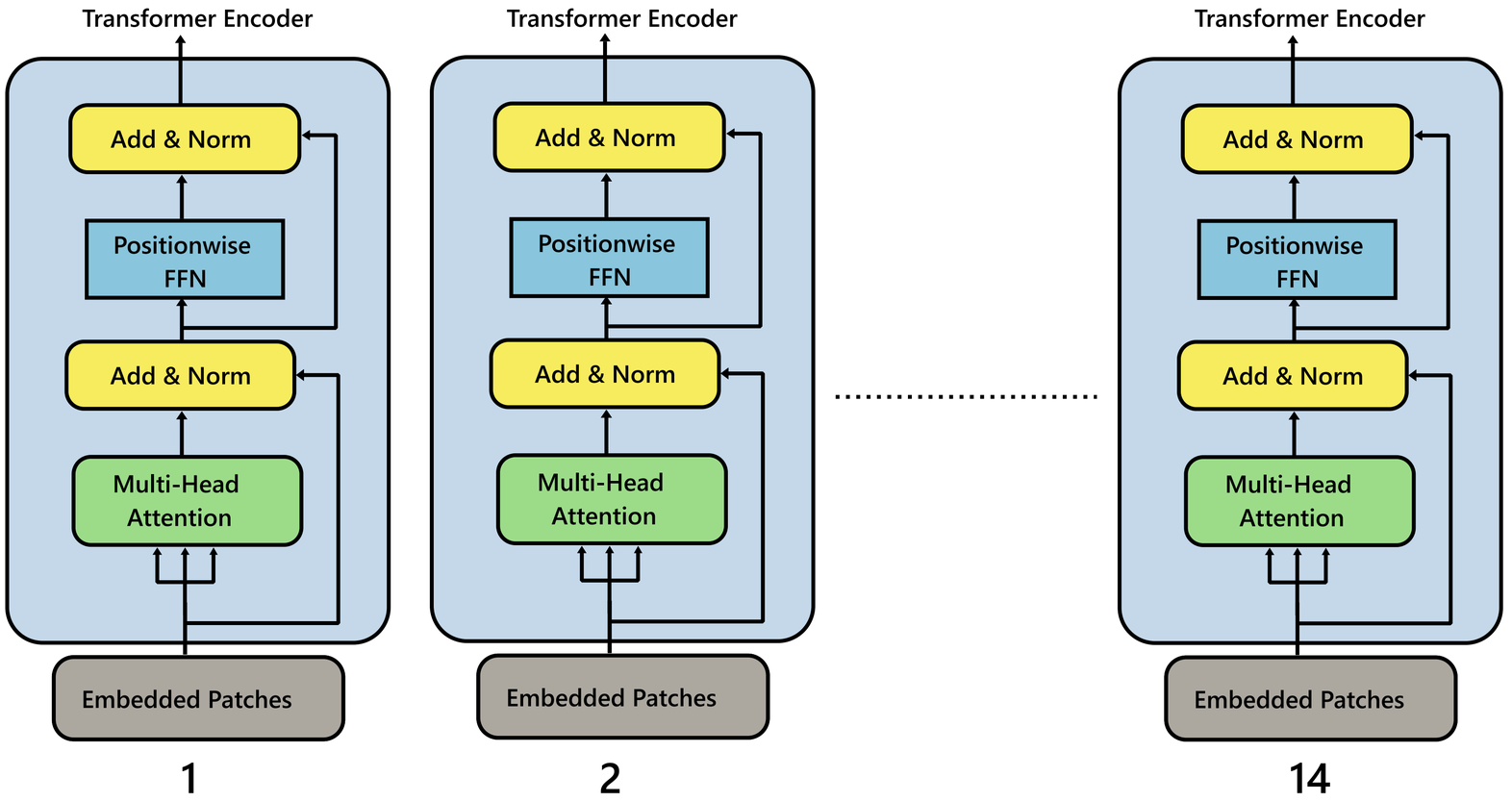


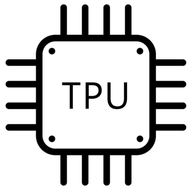
Figure-7: Architechture of Transformer Encoder [6]

Experimental Setup

Python 3.8

Tensorflow

NVDIA CUDA



TPU V2

226GB Hard Disk

335GB

Dataset

Table-1: Train and Test Split of Dataset

Class	Training Sample	Testing Sample
Deepfakes	2248	538
Face2Face	2444	686
FaceShifter	2302	501
FaceSwap	2980	697
NeuralTextures	2871	806
Original	2965	712
Total	15810	3940

Evaluations: Convolution Matrix

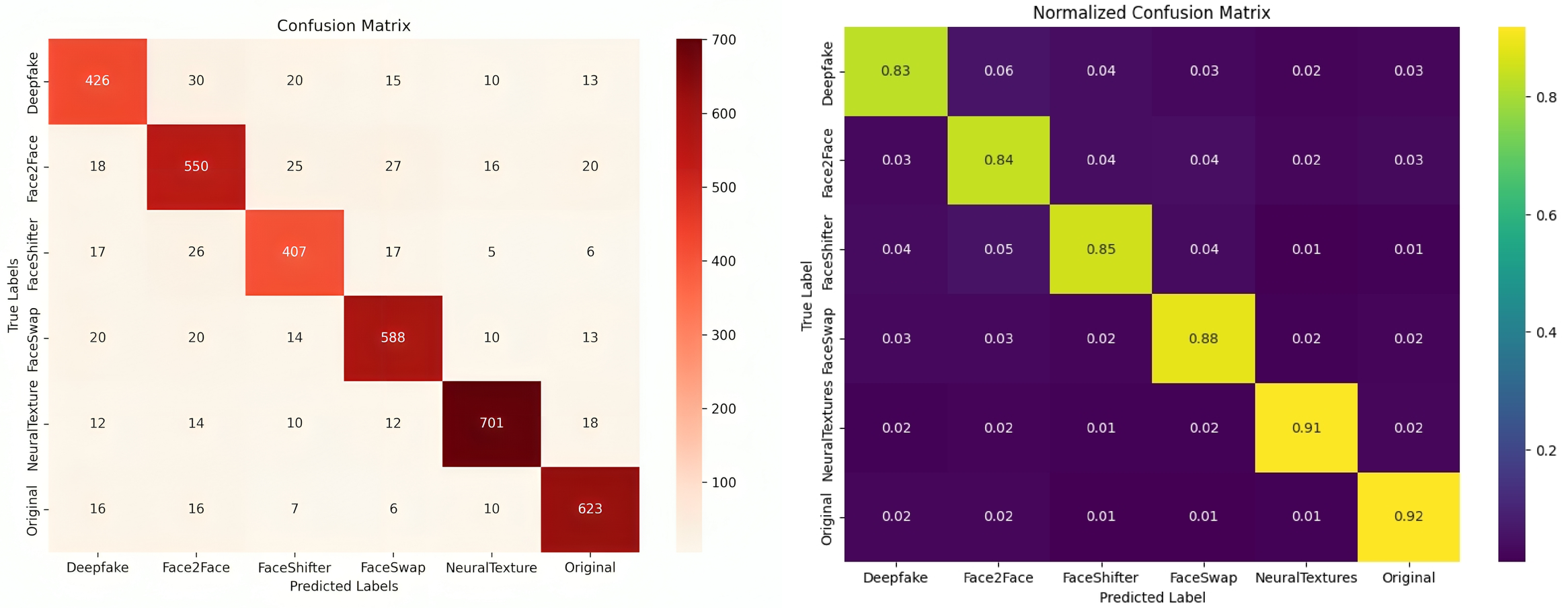


Figure-8: Confusion Matrix and Normalized Confusion Matrix

Evaluations: Precision,Recall,F1 Score

Table-2: Precision, Recall and F1 score for each class

Class	Precision	Recall	F1 Score
Deepfake	83.69%	82.88%	83.28%
Face2Face	83.84%	83.84%	83.84%
Faceshifter	84.27%	85.15%	84.70%
FaceSwap	88.42%	88.42%	88.42%
NeuralTextures	93.22%	91.40%	92.30%
Original	89.90%	91.89%	90.88%

Evaluations: Comparison with Other State of the Art

Table-3: Accuracy Comparison of Proposed Model with Other Models

Model	Accuracy
Cozzolino et al. [7]	70.97%
Bayar et al. [1]	83.10%
Rahmouni et al [8]	78.45%
Afchar et al [2]	82.97%
Fang et al. [3]	79.08%
Our Model	**83.63%**

Conclusion

Contribution

- Detects complex deepfake generation techniques by analyzing both spatial and temporal video dynamics.
- Less sensitive to data noise.
- Effectively handles inconsistencies in time series data.
- Utilization of Self Attention mechanism allowing for better context understanding and high accuracy.

Future Works

- Hyperparameter tunning
- Model optimization
- Evaluating on cross dataset

References

[1] Bayar, B., & Stamm, M. C. (2016, June). A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM workshop on information hiding and multimedia security (pp. 5-10)..

[2] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018, December). Mesonet: a compact facial video forgery detection network. In 2018 IEEE international workshop on information forensics and security (WIFS) (pp. 1-7). IEEE.

[3] Fang, S., Wang, S., & Ye, R. (2022, April). Deepfake video detection through facial sparse optical flow based light cnn. In Journal of Physics: Conference Series (Vol. 2224, No. 1, p. 012014). IOP Publishing..

[4] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1-11).

References

[5] OpenCV 3 Object Detection : Face Detection using Haar Cascade Classfiers - 2020. (n.d.). https://www.bogotobogo.com/python/OpenCV_Python/python_opencv3_Image_Object_Detection_Face_Detection_Haar_Cascade_Classifiers.php

[6] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[7] D. Cozzolino, D., Poggi, G., & Verdoliva, L. (2017, June). Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection. In Proceedings of the 5th ACM workshop on information hiding and multimedia security (pp. 159-164).

[8] Rahmouni, N., Nozick, V., Yamagishi, J., & Echizen, I. (2017, December). Distinguishing computer graphics from natural images using convolution neural networks. In 2017 IEEE workshop on information forensics and security (WIFS) (pp. 1-6). IEEE..

THE END

THANK YOU

FOR CONSIDERATE AUDIENCE

Srabonti Deb