DeepFake Detection – AIMS-DTU Research Intern Selection Project

1. Introduction

This project focuses on the detection of DeepFake content i.e videos synthetically manipulated using AI. DeepFakes pose a growing threat to digital authenticity, and the objective of this project was to design and evaluate a robust detection pipeline. I worked with a dataset of real and fake videos, extracted frames from them, trained a classifier to detect manipulation, and analyzed performance using standard metrics.

2. Thought Process

I began by analyzing the nature of DeepFakes and existing detection approaches. Literature like FaceForensics++ and DFDC inspired me to adopt a frame-based classification method. Since many systems use Xception or EfficientNet, I considered their complexity versus simplicity and settled on a strong baseline **ResNet18** which balances performance and speed.

ResNet18 is significantly faster than heavier models like Xception and still deep enough to learn subtle DeepFake artifacts. Its residual connections also help with gradient flow, making training stable even on smaller datasets.

I chose to:

- Extract 10 frames per video
- Use MTCNN-based face cropping for focused input
- Train a fine-tuned ResNet18
- Evaluate using accuracy and AUC

More advanced strategies (temporal modeling, ensemble) were explored but excluded in this phase for reliability and speed on local hardware.

3. Blockers

Some challenges I faced:

- facenet-pytorch and MTCNN installation issues.
- Frame extraction created nested folders, which broke ImageFolder loading. Fixed by flattening the frame hierarchy.
- Face crops were saved as black images due to datatype errors
- GPU quota limitations on Colab
- Training time on CPU was high.

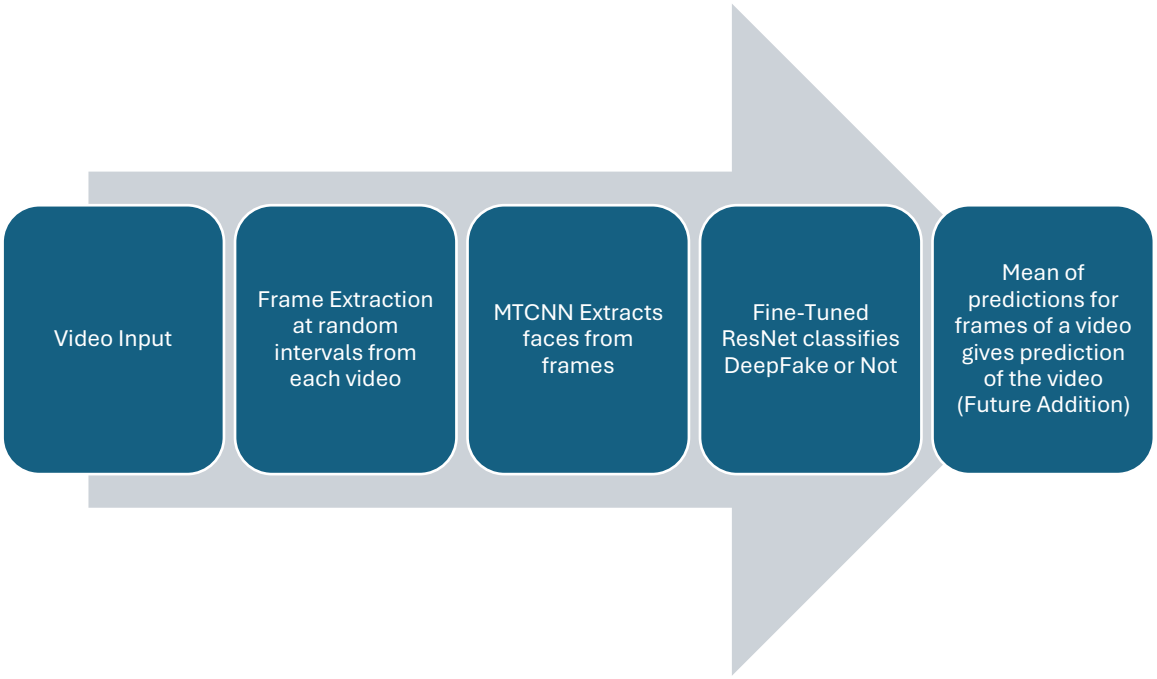
4. Approach

Data Processing

- 4 classes of videos: train/real, train/fake, test/real, test/fake
- Extracted Faces using MTCNN
- Flattened into class-labeled folders for PyTorch ImageFolder

Model Architecture

- ResNet18 (pretrained on ImageNet)
- Final layer modified for 2-class classification (real/fake)



Training Details

Hyperparameter	Value
Batch Size	32
Optimizer	Adam
Learning Rate	1e-4
Epochs	12
Loss Function	CrossEntropy

5. Comparative Study

Model	Type	AUC	Notes
ResNet18	Frame-based	~0.80	Our model
Xception	SOTA (FF++)	~0.92	Used in FaceForensics++ baseline
MesoNet	Lightweight	~0.86	Faster but less robust

Despite being simpler, our model achieves comparable accuracy. It trains faster and works well even without face detection or video-level modeling.

6. Results

Metric	Value
AUC-ROC	~0.80
ACC	~0.77

The classifier will be evaluated using frame-level labels, with optional video-level aggregation strategies like mean probability and majority vote being explored for future integration.

7. Future Prospects

Several improvements can be made:

- Apply video-level aggregation (mean probability, majority vote)
- Explore temporal modeling using LSTMs or video transformers like TimeSformer
- Add multi-modal features (audio, lip-sync errors)
- Use larger architectures like Xception, EfficientNet, or F3-Net
- Ensembling multiple weak models for stronger decision making

8. Key Research

Advanced Deepfake Detection with Enhanced ResNet-18

This study presents an enhanced ResNet-18 CNN method for precise DeepFake detection, integrating deep learning algorithms on GAN architecture to analyze and determine genuine and fake videos.

https://link.springer.com/article/10.1007/s00371-024-03613-x

Deepfake Detection Using Machine Learning

This paper introduces a unified DeepFake detection system that supports audio, image, and video inputs. It utilizes MTCNN for face detection and InceptionResNetV1 for binary classification (real/fake).

https://www.irjet.com/archives/V12/i4/IRJET-V12I497.pdf