Deep Fake Detection

SIM DD Team
15 May 2025

01

Overview

Overview

2. Problem Statement
3. Technology Stack
4. Model Architecture (Machine Learning)
5. Demo
6. Q&A

02

Problem Statement

Problem Statement

Background: "When used maliciously, deepfake can pose detrimental implications... damaging the reputation of prominent individuals, and influencing public opinions" (Seow et al., 2022)

Problem: Existing detection methods struggle to counter increasingly sophisticated Deepfake techniques (Tolosana et. al, 2020).

Objective: Leverage machine learning algorithms (eg. ResNet) to enhance accuracy and reliability in identifying manipulated media effectively (Khan et. al, 2024)

03

Technology Stack

- Open-source library for building and training machine learning models.

- Python library for audio analysis and feature extraction.

- Python library for creating static, interactive, and animated visualizations and plots.

04

Model Architecture (Machine Learning)

Multimodal Approach

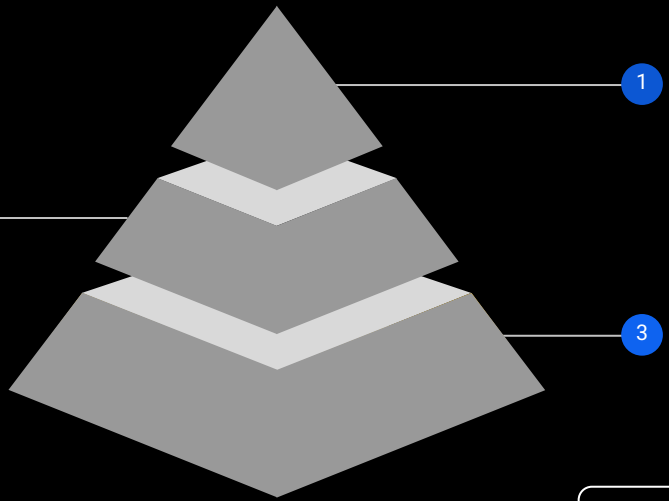
Spatial

- Image quality analysis
- Detect spatial artifacts and anomalies
- Xception and swin transformer

Audio-Visual

- Sync analysis
- Match audio tracks and lip movements
- Resnet and Wave2Vec2

Model Architecture



1 **Swin Transformer (Spatial Analysis)**

2 **Wave2Vec2 (Audio-Visual Analysis)**

3 **PyDeepFakeDet - GitHub**

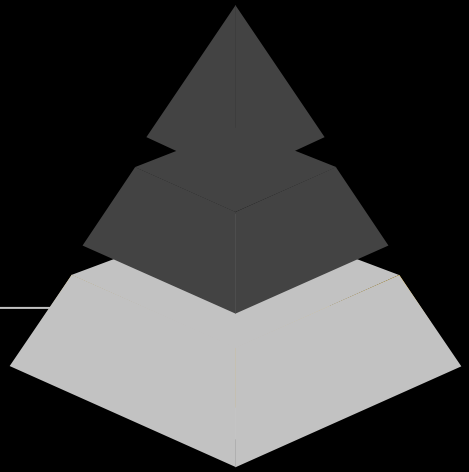
- ResNet
- Xception

Model Architecture

PyDeepFakeDet

- Developed by Fudan Vision and Learning Lab
- Well-rounded model

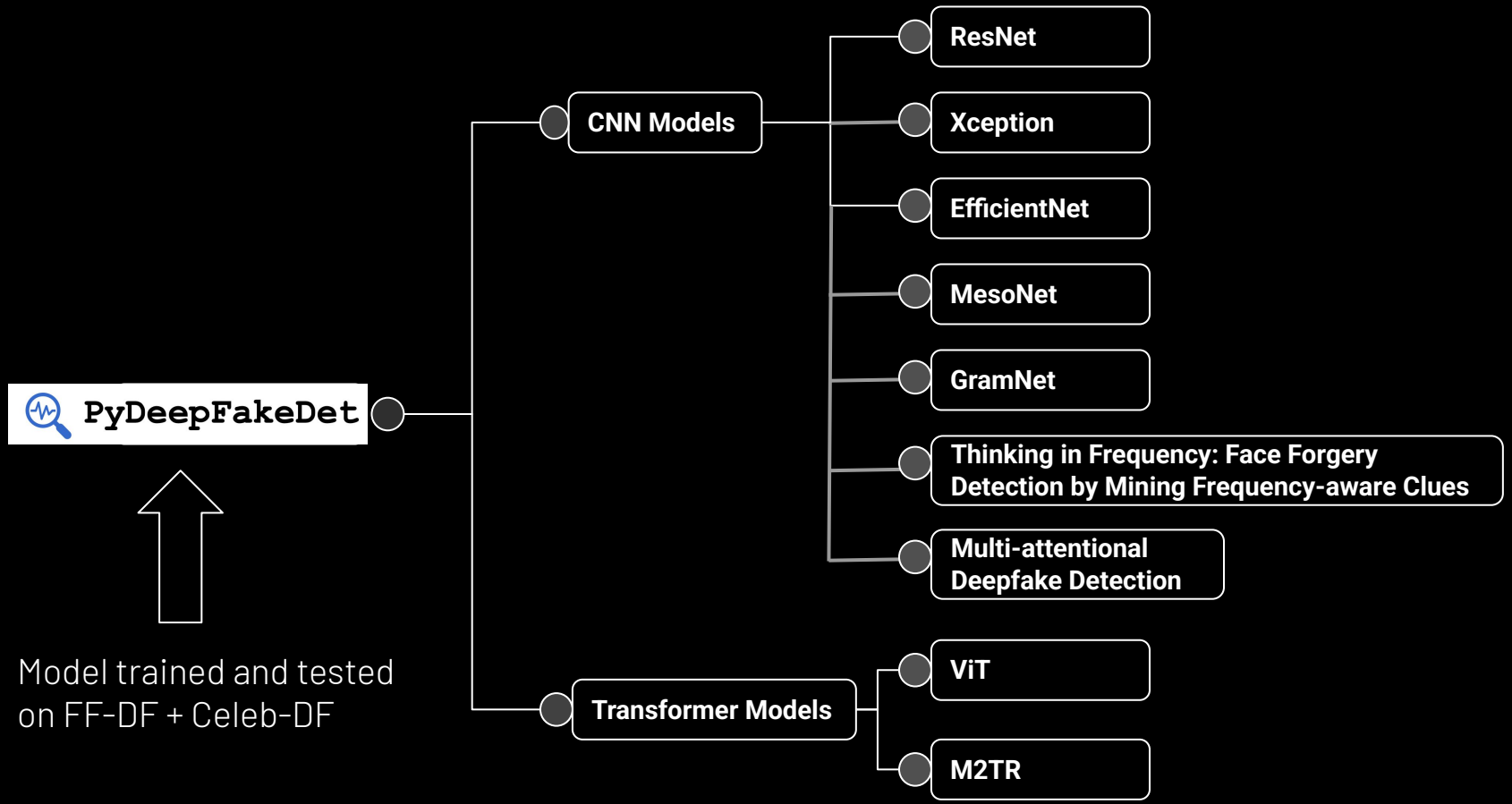
PyDeepFakeDet - GitHub 3

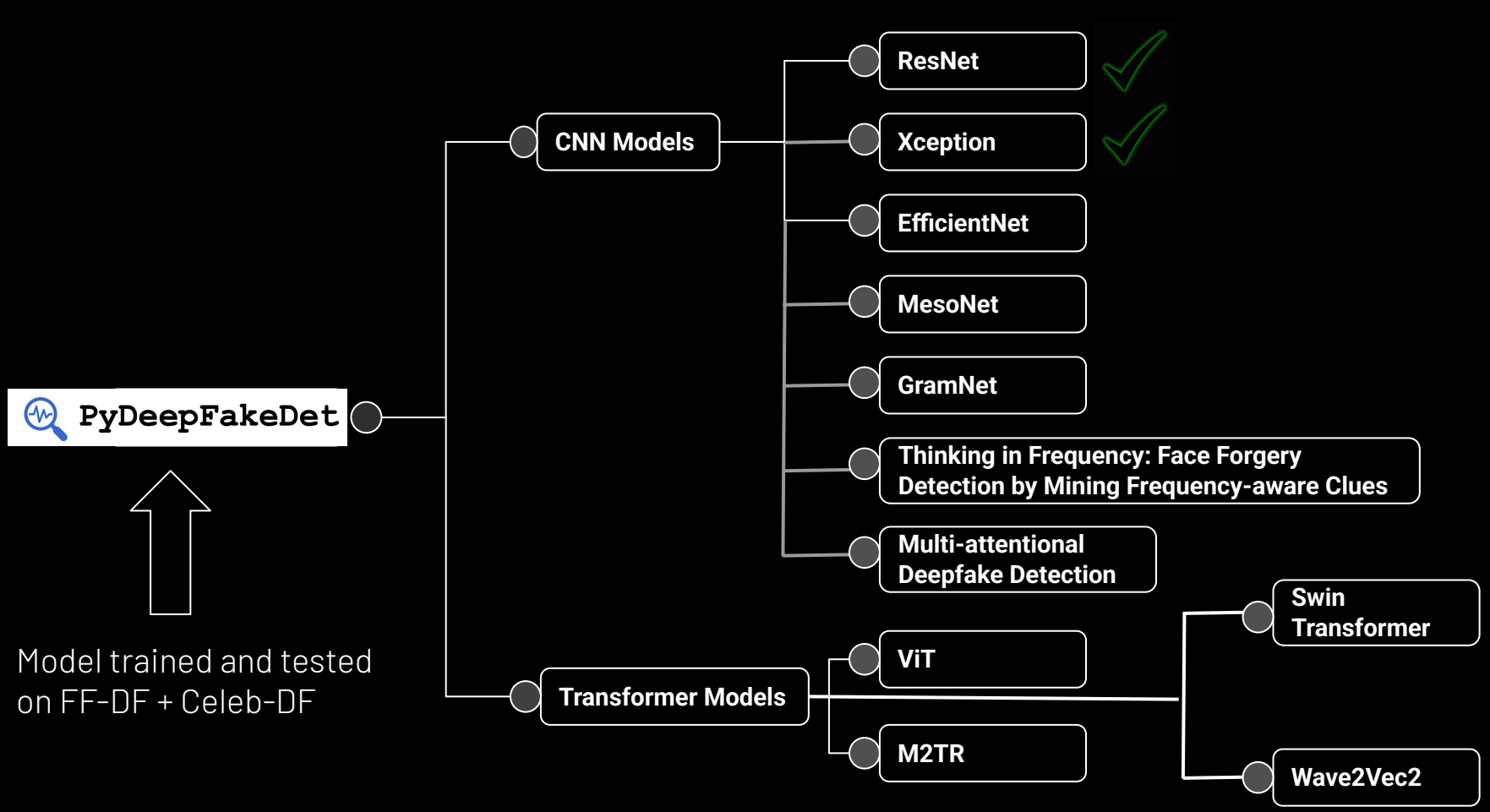


Model Architecture

PyDeepFakeDet

- Developed by Fudan Vision and Learning Lab
- Well-rounded model


PyDeepFakeDet
Model trained and tested on FF-DF + Celeb-DF
CNN Models
ResNet
Xception
EfficientNet
MesoNet
GramNet
Thinking in Frequency: Face Forgery Detection by Mining Frequency-aware Clues
Multi-attentional Deepfake Detection
Transformer Models
ViT
M2TR


PyDeepFakeDet
Model trained and tested on FF-DF + Celeb-DF
CNN Models
ResNet
Xception
EfficientNet
MesoNet
GramNet
Thinking in Frequency: Face Forgery Detection by Mining Frequency-aware Clues
Multi-attentional Deepfake Detection
Transformer Models
ViT
M2TR
Swin Transformer
Wave2Vec2

Model Architecture

Spatial Analysis: Models

- Xception - PyDeepFakeDet
  - Lightweight & fast
  - High spatial sensitivity

- Swin Transformer
  - Type of Vision Transformer (ViT)
    - Global + local attention (focuses on finer details)
    - Computationally efficient
  - Outperforms typically used CNNs like ResNet, even original ViT transformer
    - Context understanding between frames that ResNet might miss
    - Enhanced version of ViT

Model Architecture

Spatial Analysis: Xception + Swin Transformer

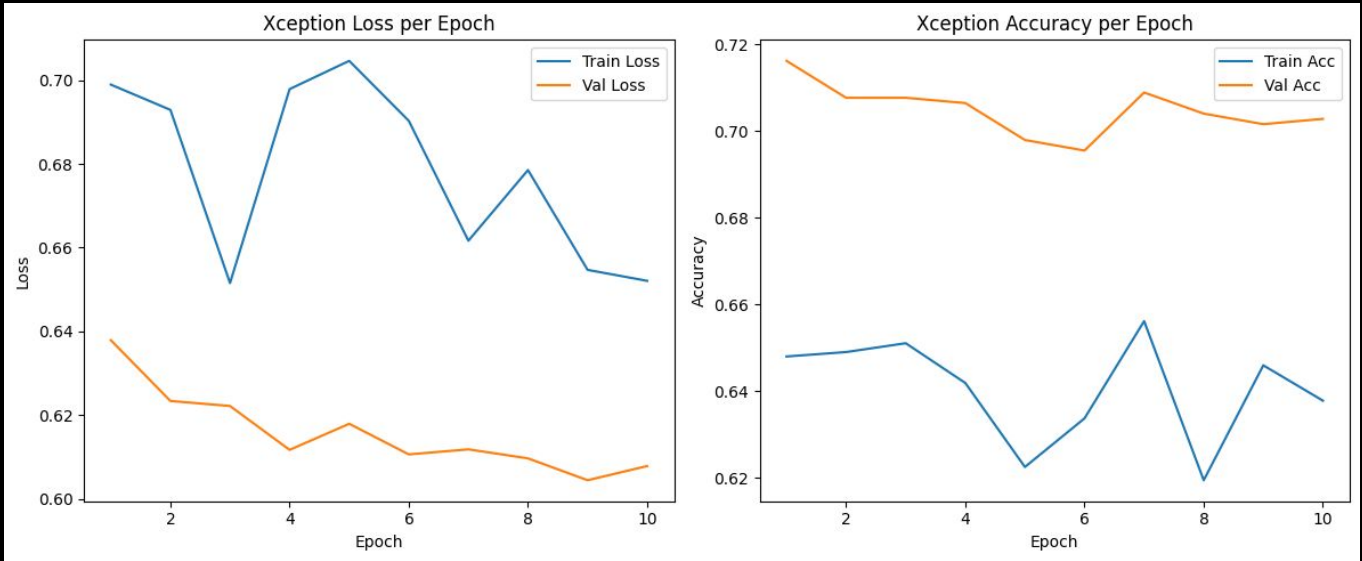
- Pipeline

Environment setup	Pre-processing	Augmentation	Model	Training	Evaluation
01	02	03	04	05	06
• Mounting gdrive to colab and base code (PyDeepFakeDet) • Install dependencies	• Frame extraction • Data split	• Resizing • Converting images to tensors • Random horizontal flips • Normalisation	• Swin Transformer • Xception - fine tuned	• CrossEntropyLoss with AdamW Optimiser • Model evaluated on val set after each epoch • Early stopping employed	• Compute accuracy • Classification report • Soft-voting ensemble

Model Architecture

Spatial Analysis: Xception

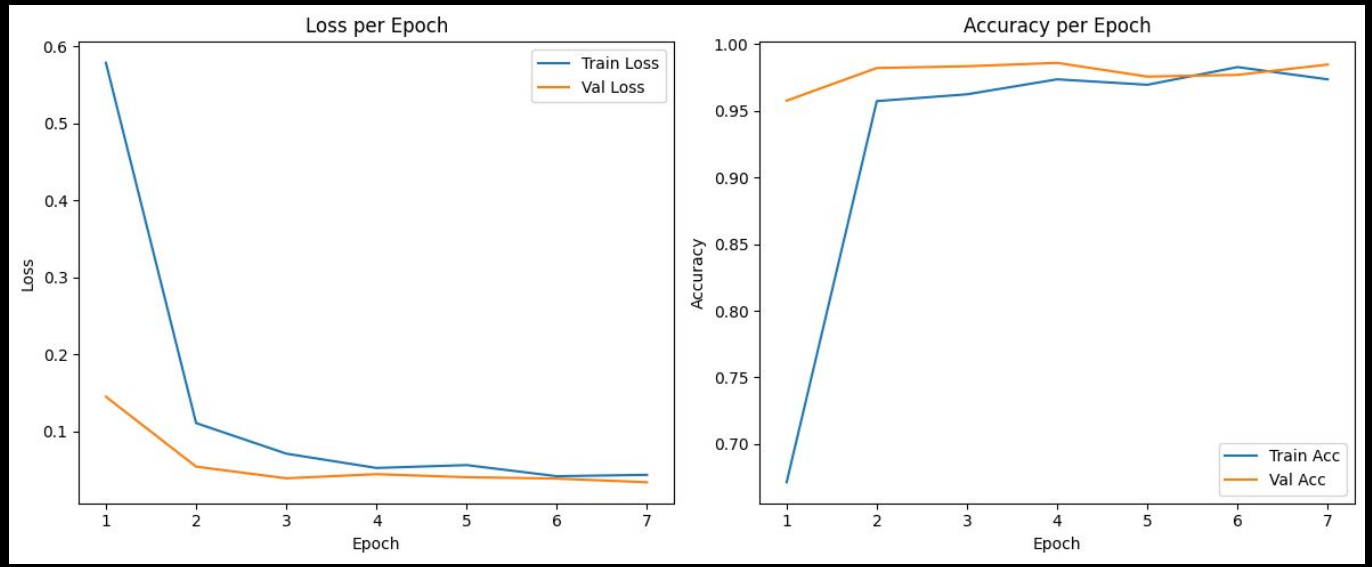
- Model Performance



Model Architecture

Spatial Analysis: Swin Transformer

- Model Performance



Model Architecture

Spatial Analysis: Xception + Swin Transformer

- Model Performance

swin Accuracy: 0.9817

	precision	recall	f1-score	support
fake	0.98	0.98	0.98	301
real	0.98	0.98	0.98	301
accuracy			0.98	602
macro avg	0.98	0.98	0.98	602
weighted avg	0.98	0.98	0.98	602

xception Accuracy: 0.7060

	precision	recall	f1-score	support
fake	0.70	0.71	0.71	301
real	0.71	0.70	0.70	301
accuracy			0.71	602
macro avg	0.71	0.71	0.71	602
weighted avg	0.71	0.71	0.71	602

Ensemble Test Accuracy: 0.9817275747508306

	precision	recall	f1-score	support
fake	0.98	0.98	0.98	301
real	0.98	0.98	0.98	301
accuracy			0.98	602
macro avg	0.98	0.98	0.98	602
weighted avg	0.98	0.98	0.98	602

Model Architecture

Audio-Visual Analysis: Models

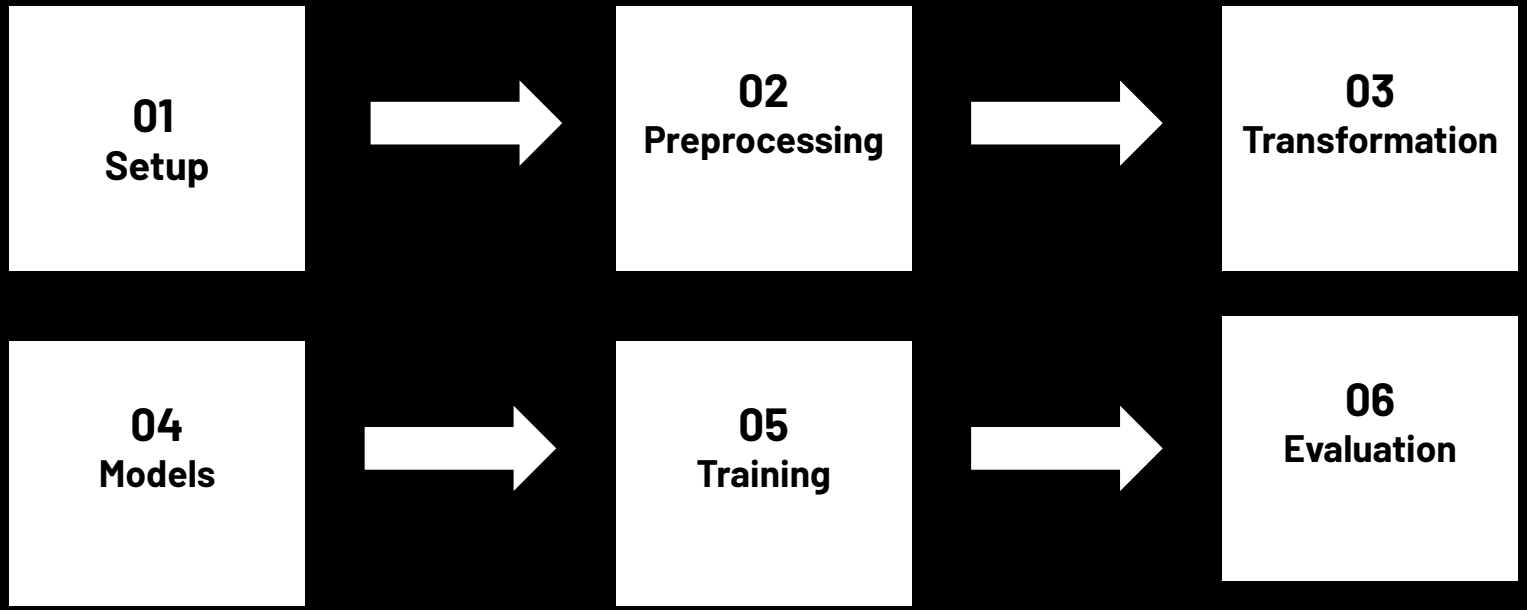
- ResNet - PyDeepFakeDet
  - Used for video frames
  - One of the most widely used CNN for image classification
  - Excellent at capturing visual inconsistencies
  - Efficient and scalable

- Wave2Vec2
  - One of the leading model for speech recognitions
  - Powerful speech feature extraction
  - Sync verification

Model Architecture

Audio-Visual Analysis:

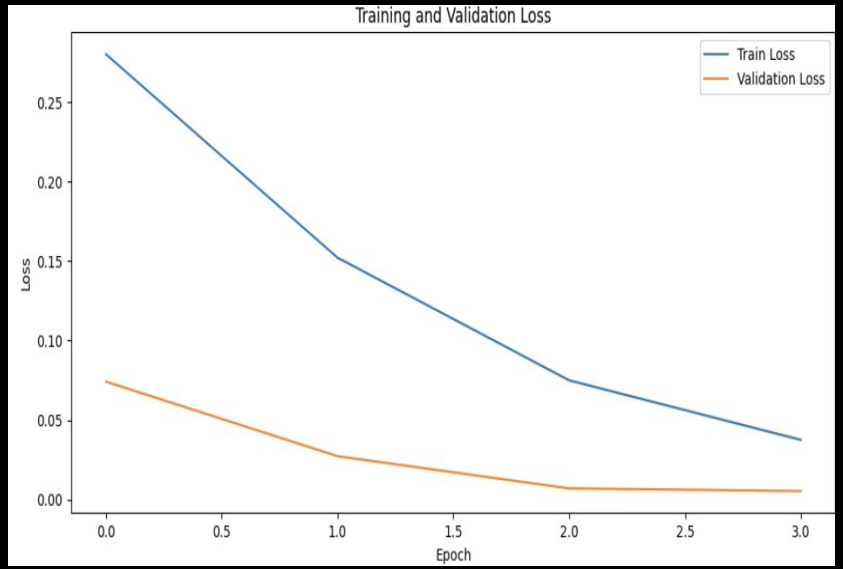
- Pipeline

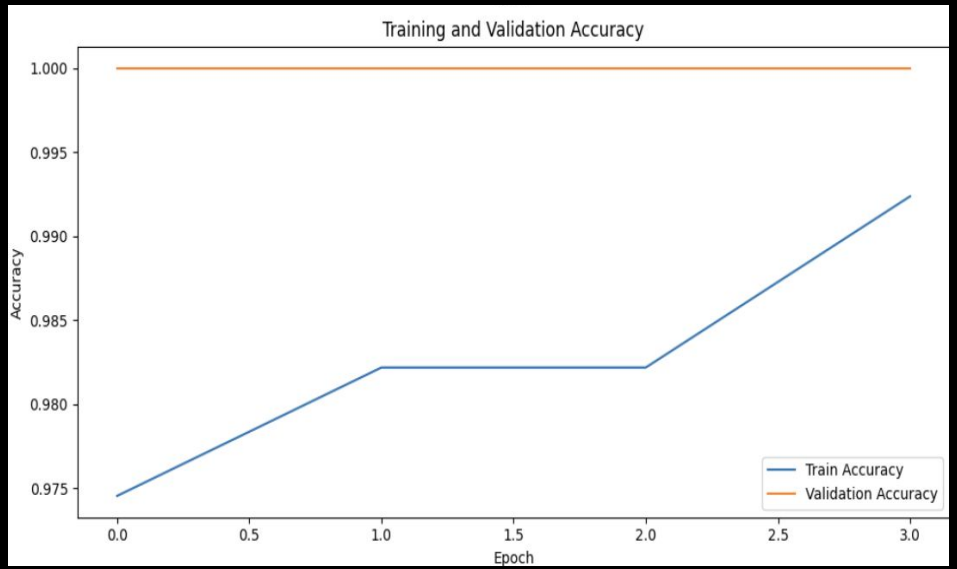


Model Architecture

Audio-Visual Analysis:

- Model Performance





05

Demo

06

Q&A

Thank You