Deepfake Video Detection

Deep Learning Project

Shubh Khandelwal
CS22B1090

Deepfake Detection Overview

1 **Vision Transformer Architecture**

This project utilizes **Vision Transformer (VIT)** architecture for detecting fake videos effectively.

2 **Celeb-DF Dataset**

Training and testing are conducted using the **Celeb-DF v2 dataset**, which is crucial for model performance.

3 **Evaluation Metrics**

Key evaluation metrics include **Accuracy**, **AUC**, **Precision**, and **Equal Error Rate (EER)** for performance assessment.

4 **Deepfake Video Classification**

The main objective is to classify **deepfake videos** accurately, enhancing detection capabilities.

5 **Project Development**

The report details the comprehensive **development and evaluation** process of the proposed model.

Essential Libraries for Project

1 **torch: Core library for tensor computations**

PyTorch is essential for building and training deep learning models, providing the foundation for neural network calculations.

2 **torchvision: Image processing tools**

Torchvision offers datasets, model architectures, and image transformations that are useful for computer vision tasks.

3 **scikit-learn: Machine learning library**

Scikit-learn is crucial for implementing various machine learning algorithms, aiding in model evaluation and data preprocessing.

4 **numpy: Numerical operations**

Numpy is fundamental for numerical computations in Python, supporting large multi-dimensional arrays and matrices.

5 **opencv-python: Computer vision library**

OpenCV is widely used for image processing and computer vision applications, providing tools for real-time processing.

6 **pandas: Data manipulation and analysis**

Pandas is key for data manipulation, offering data structures and functions for efficient data analysis in Python.

Celeb-DF v2 Dataset Overview

The dataset used for this research is the Celeb-DF v2, which is pivotal for evaluating deepfake detection techniques.

Classes: Real (0) and Fake (1)

The dataset comprises around 5639 videos, providing a substantial amount of data for analysis and model training.
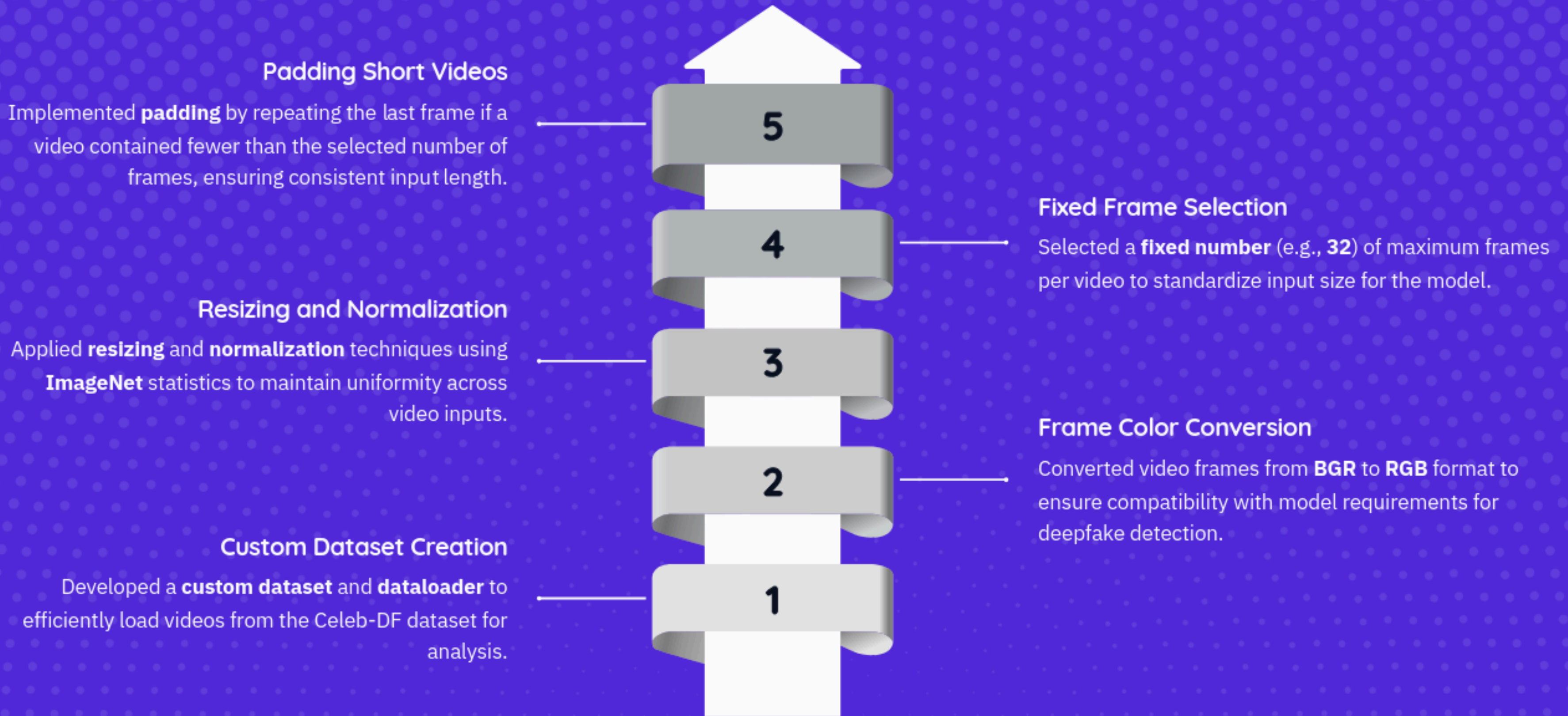
Format: MP4

Dataset: Celeb-DF v2

It categorizes videos into two classes: Real (0) and Fake (1), essential for training and validation of detection models.
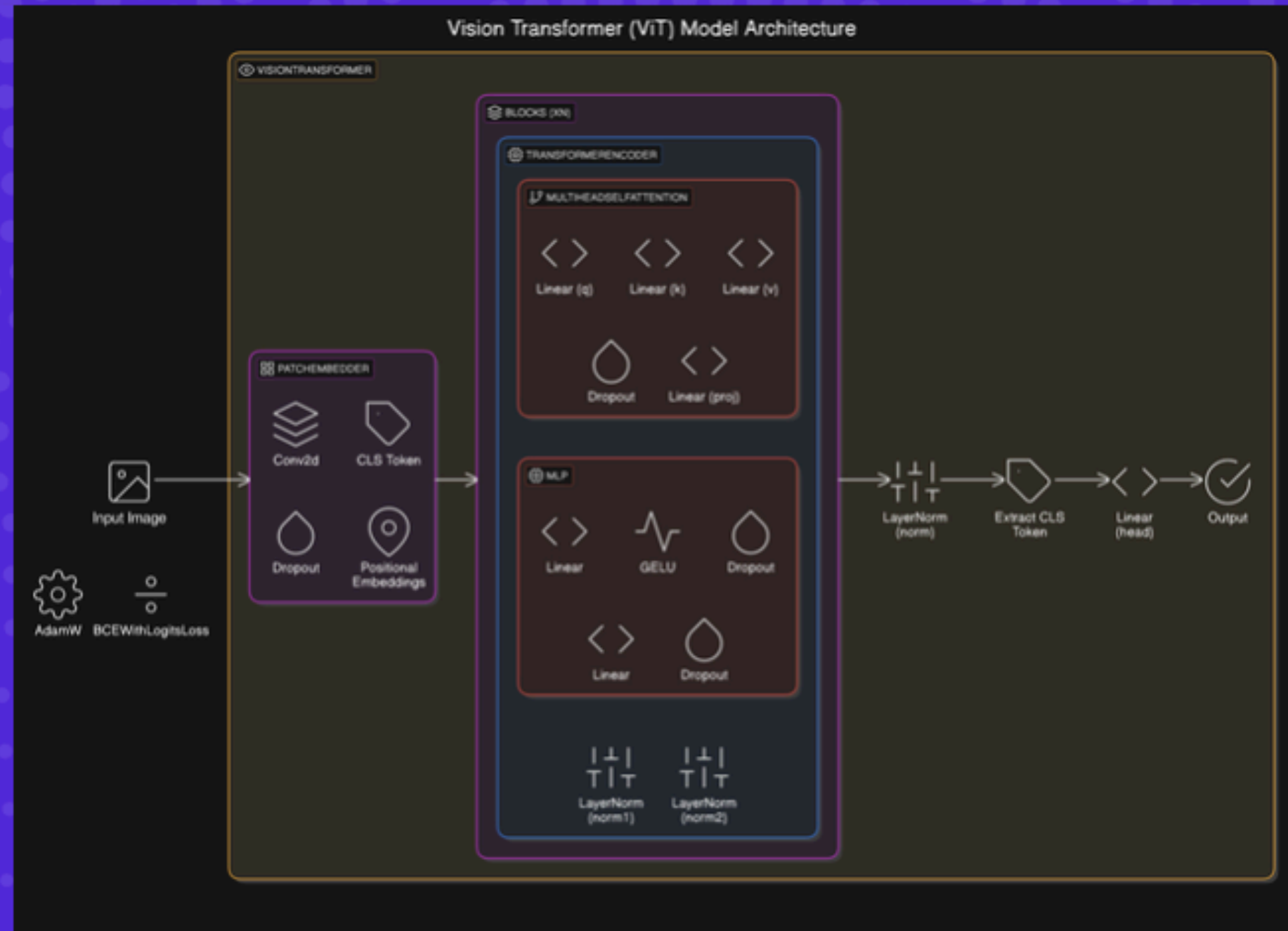
Total Videos: Approximately 5639

All videos in the dataset are in MP4 format, ensuring compatibility with most video processing tools and libraries.

Preprocessing Steps for Videos

Padding Short Videos

Implemented **padding** by repeating the last frame if a video contained fewer than the selected number of frames, ensuring consistent input length.

5

Fixed Frame Selection

Selected a **fixed number** (e.g., **32**) of maximum frames per video to standardize input size for the model.

4

Resizing and Normalization

Applied **resizing** and **normalization** techniques using **ImageNet** statistics to maintain uniformity across video inputs.

3

Frame Color Conversion

Converted video frames from **BGR** to **RGB** format to ensure compatibility with model requirements for deepfake detection.

2

Custom Dataset Creation

Developed a **custom dataset** and **dataloader** to efficiently load videos from the Celeb-DF dataset for analysis.

1

ViT Architecture

Vision Transformer Architecture Overview

This slide details the architecture of Vision Transformer (ViT), including patch embedding, multi-head self-attention, and transformer encoder blocks.

Training Parameters Overview

- **Optimizer: AdamW**

  The AdamW optimizer is utilized to improve convergence speed and performance during training. It combines the benefits of Adam with weight decay for better regularization.

- **Learning Rate: 3e-4**

  A learning rate of 3e-4 is selected to balance the trade-off between training speed and model performance, allowing for efficient updates during optimization.

- **Weight Decay: 0.05**

  Weight decay at 0.05 is applied to prevent overfitting by penalizing large weights during training, promoting a simpler model.

- **Loss Function: BCEWithLogitsLoss**

  Binary Cross Entropy with Logits Loss is used as the loss function, suitable for binary classification tasks such as deepfake detection.

- **Epochs: 10**

  The model will be trained for 10 epochs, allowing sufficient iterations through the training data to learn the patterns effectively.

- **Batch Size: 2**

  A small batch size of 2 is used to allow for more frequent updates to the model parameters, which can help in better generalization.

Evaluation of Detection Metrics

Accuracy: 88.52%

The model achieved an impressive **accuracy** of 88.52%, indicating a high level of correct detections in deepfake videos.

Area Under Curve (AUC): 0.4987

The **Area Under Curve (AUC)** score of 0.4987 suggests a moderate capability of the model in distinguishing between real and fake videos.

Precision: 88.52%

With a **precision** of 88.52%, the model demonstrates a strong ability to identify true positives among its predictions.

Equal Error Rate (EER): 0.00

An **Equal Error Rate (EER)** of 0.00 indicates that there are no false positives or false negatives, showcasing the model's effectiveness.

Deepfake Detection References

Transformers in Image Recognition

Explores the use of transformers for image classification at scale.

Survey on Face Manipulation

Covers various techniques in face manipulation and detection methods.

Celeb-DF Dataset for Forensics

Introduces a challenging dataset tailored for deepfake forensic analysis.

Attention Mechanism

Presents the foundational model of transformer architecture focused on attention.

Thank You for Your Attention