Enhancing Face Anti-Spoofing Using Transformer Networks

Varsha B J.A. Varsha Dr.R.Kanchana (Guide)

Department of CSE, Sri Sivasubramaniya Nadar college of Engineering

Final Year Project, May 2024

Highlights of Proposed Model

To develop a robust face anti-spoofing system that

- Utilizes a pretrained vision transformer model for enhanced accuracy.
- Implements the Selective Patch Attention Network (SPAN) for capturing subtle cues.
- Integrates the Multi-Scale Weighted Fusion (MSWF) for improved depth estimation.
- Demonstrates exceptional accuracy across diverse scenarios.
- Transformed into a real-time web application for testing face anti-spoofing measures.

Challenges in the proposed Face Anti-spoofing Method:

- Integration of the proposed modules into the pre-trained vision transformer.
- Addressing variability in real-world environments, including diverse lighting conditions and background clutter.

Performance Metrics for the Proposed System

- Accuracy
- Attack Presentation Classification Error Rate (APCER)
- Bonafide Presentation Classification Error Rate (BPCER)
- Average Classification Error Rate (ACER)

Model Prediction Results

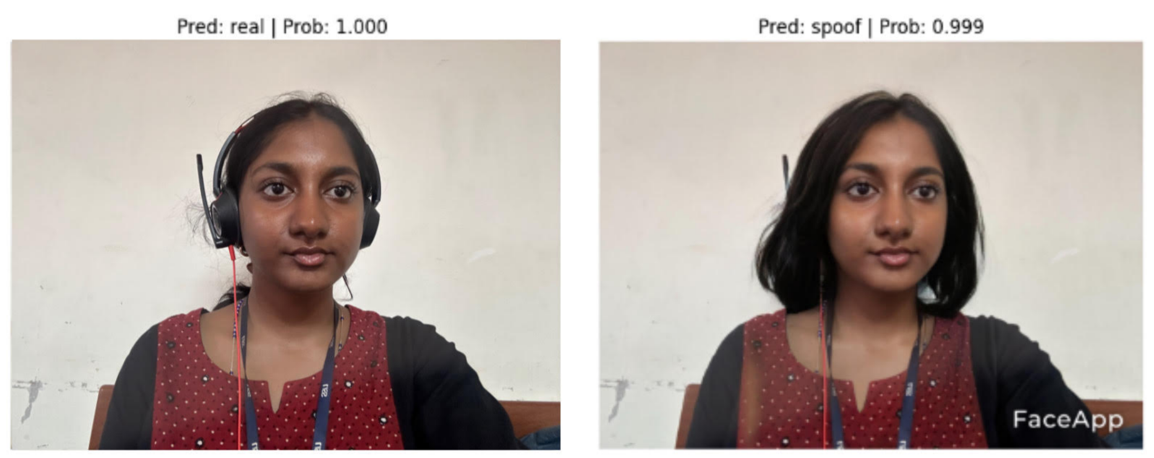

Figure: Real Image Figure: Spoofed Image

Figure 1. Model Prediction Vit+SPAN+MSWF

Functional Modules and Dataset Description

- Data Pre-processing
  - Data augmentation for real images.
  - Calculation of image statistics for data analysis.
  - Splitting dataset into training and validation sets.
- Transformer Modules
  - Multi-Head Self-Attention (MSA)
  - Selective Patch Attention Network (SPAN)
  - Multi-Scale Weighted Fusion (MSWF)
  - Face Spoof Detector

- Combined two publicly available datasets: LCC_FASD and SiW.
- LCC_FASD dataset: 1302 real images, 7444 spoof images.
- SiW dataset: 5410 real images, 866 spoof images.
- Combining datasets ensured a balanced and comprehensive dataset for training and evaluation.

Functional pipeline of proposed method

Data Preprocessing:

- Cleaning and normalization.
- Augmentation for dataset diversity.
- Splitting into training and validation sets.

Vision Transformer:

- Input embedding: Patch and positional encoding.
- Transformer encoder: Multi-head self-attention and feed-forward neural network.
- Layer normalization and residual connection.

SPAN:

- Patch selection based on attention scores.
- Cross-relation aware attention mechanism.
- Importance score computation for each patch.

MSWF:

- Normalization and non-linear transformations.
- Softmax for attention weight calculation and weight balncing
- Fusion of low and high-level features.

Web Application Integration with Face Anti-spoofing model

- **Tech Stack:**
  - Backend: Flask
  - Frontend: HTML, CSS, JavaScript, govern client-side user interactions.
  - Additional: WebRTC API
- **Workflow:**
  - WebRTC API orchestrates transmission of image data from webcam to Flask server.
  - Flask efficiently stores the received image in server directory for processing.
  - User interaction triggers complex image analysis pipeline.
  - Flask loads pre-trained model into memory for sophisticated analysis.
  - Pre-trained model executes intricate algorithms to extract features and predict outcomes.
  - Flask delivers prediction results to JavaScript for seamless display on webpage.

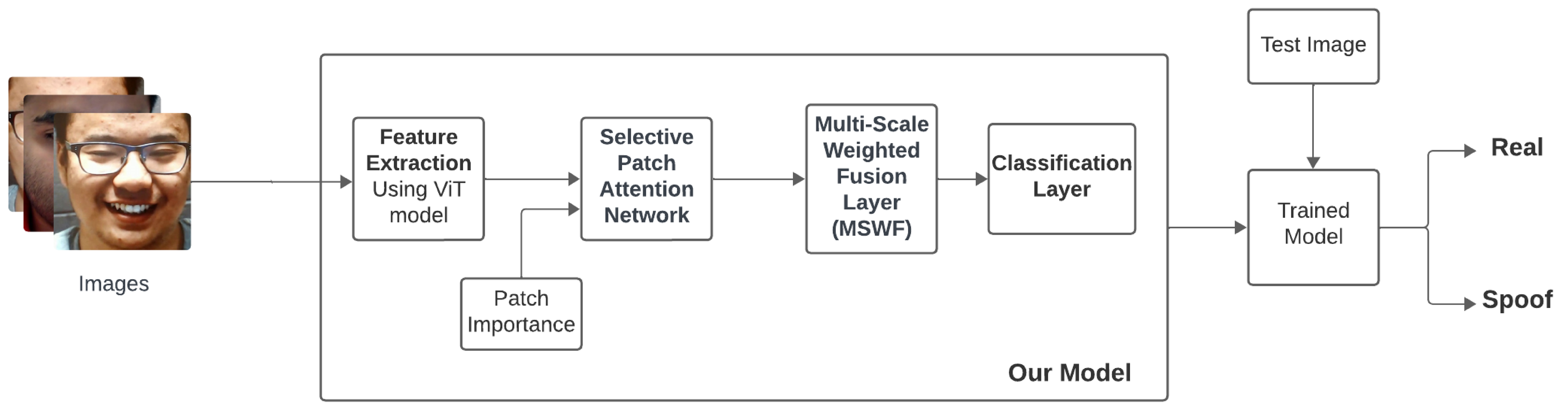

Figure 2. Proposed System Architecture

Performance Analysis

Table 1. Training and Validation results

Dataset	train_loss	train_accuracy	val_loss	val_accuracy
LCC_FASD with CRA + HFF	1.1634	0.8277	0.3981	0.8617
LCC_FASD with SPAN + HFF	0.43424	0.8498	0.6051	0.8617
LCC_FASD - SPAN + MSWF	0.7006	0.5091	0.6626	0.8607
3000 LCC_FASD+SiW Images- SPAN + MSWF	0.6947	0.4938	0.6823	0.8617
SiW Image Dataset SPAN + MSWF	**0.0000**	**1.0000**	**0.3438**	**0.9566**

Table 2. Performance Metrics of the model on the datasets

Dataset	APCER	BPCER	ACER
SiW - Development dataset	0.0000	0.0918	0.0459
SiW - Evaluation dataset	0.0000	0.0240	0.0120
LCC_FASD - Development dataset	0.8774	0.0220	0.4497
LCC_FASD -Evaluation dataset	0.9443	0.0100	0.4771

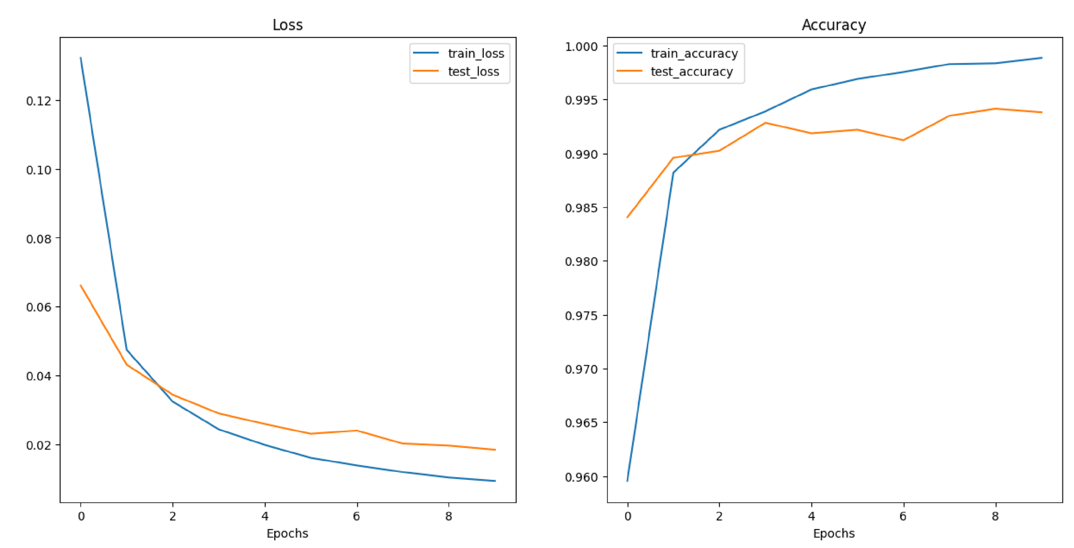

Figure 3. Loss curves of the proposed system

Web Application prediction

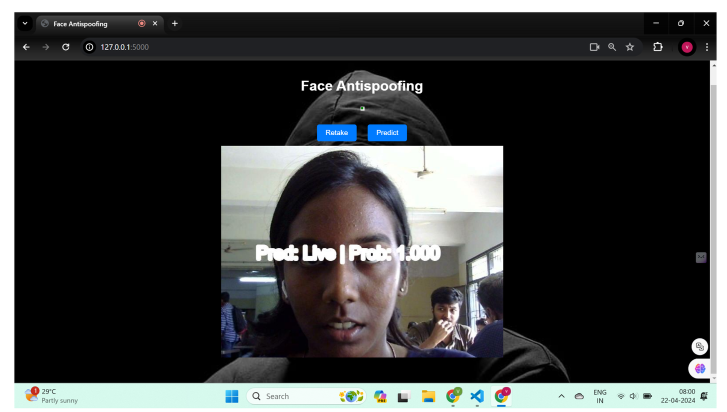

Figure 4. Web App Prediction

Inferences

- Effectiveness of SPAN and MSWF in identifying and differentiating between real and spoofed images
- Robustness against a variety of spoofing attacks, including sophisticated AI-generated spoofing attempts
- Benefits of using attention mechanisms in Image Analysis (SPAN), particularly in extracting and prioritzing important features
- Integrating multi scale features correlates with enhancing model's performance