Deepfake Detection for Online Media

Deep learning system for detecting AI-generated deepfake images, with a focus on Vietnamese faces, delivered through a web platform and browser extension

Background & Motivation:

Deepfakes have become a growing threat to media integrity, politics, and online security worldwide. While many detection systems exist, they are often trained on Western datasets, which limits their effectiveness for non-Western contexts. Vietnam, like many countries, faces increasing risks of misinformation and fraud through manipulated facial images. To address this, our project develops a deepfake detection system tailored to Vietnamese facial data, using deep learning methods to achieve high accuracy and real-time usability.

Key Features:

- CNN-based model with SE block and dual-pooling enhancements
- Trained on Vietnamese + global deepfake datasets (CelebDF, FF++, DFDC)
- Achieves strong benchmark accuracy on multiple datasets, with over 90% overall testing accuracy
- Dual interface: web platform and browser extension
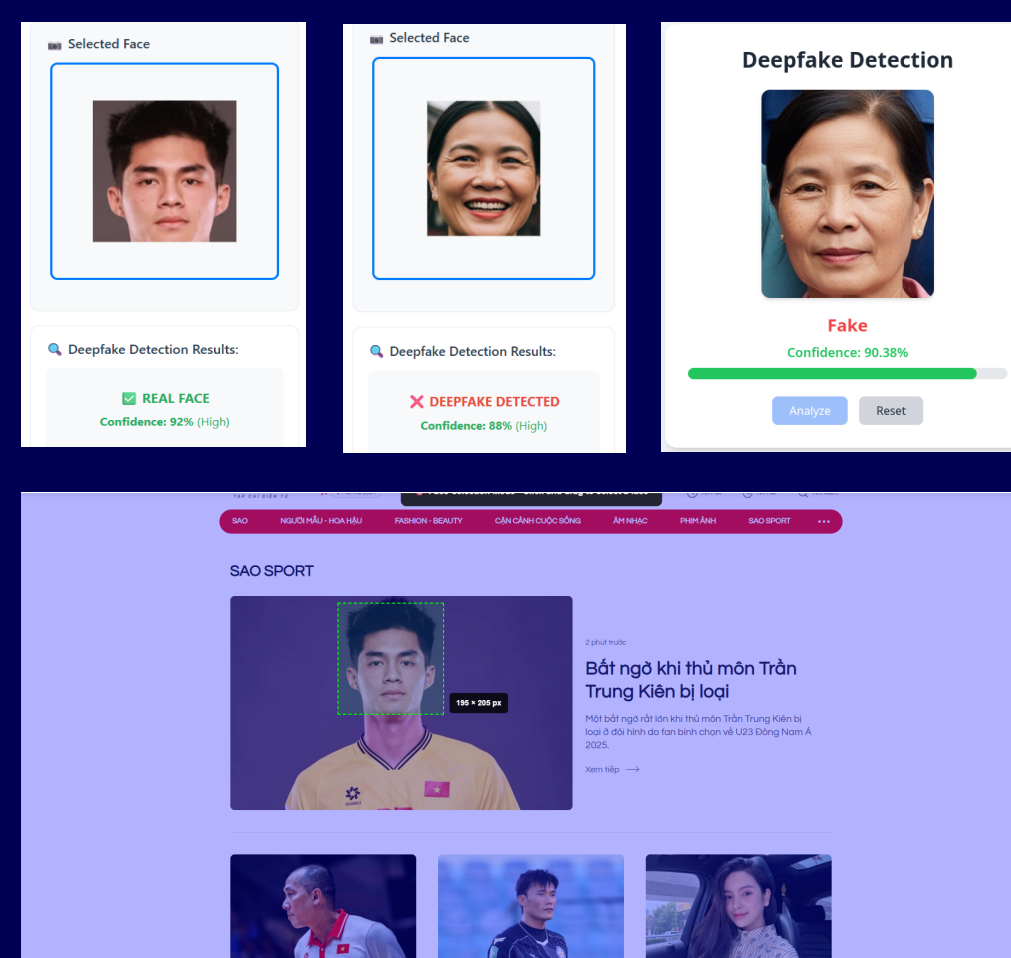- Real-time image verification with intuitive results

Model Architecture:

- Convolutional layer
- Batch Normalization
- SE Block
- Max Pooling
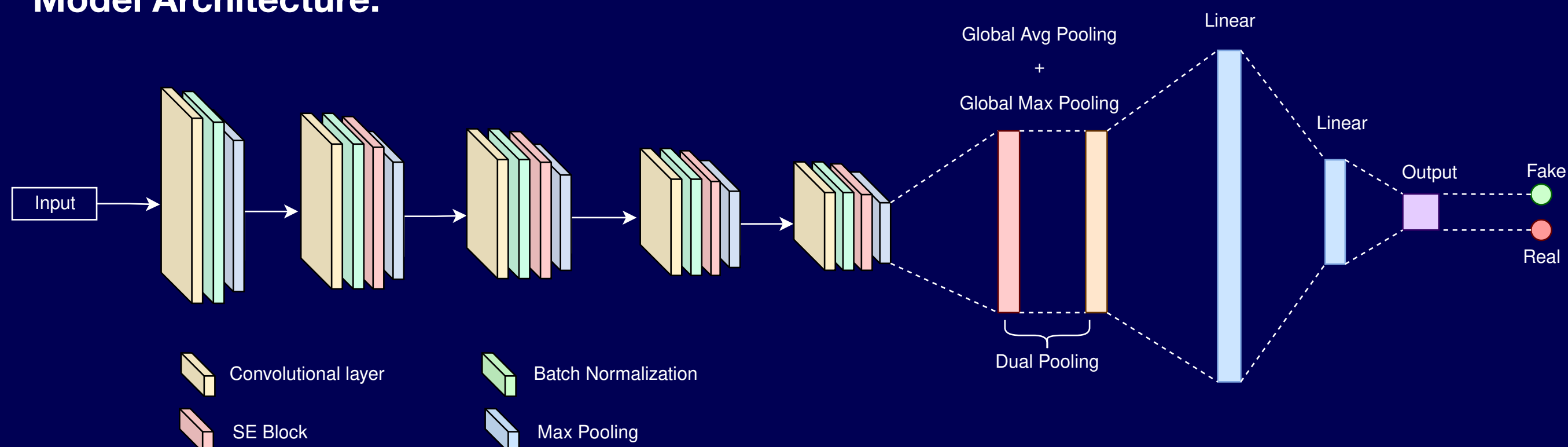
System Diagram:
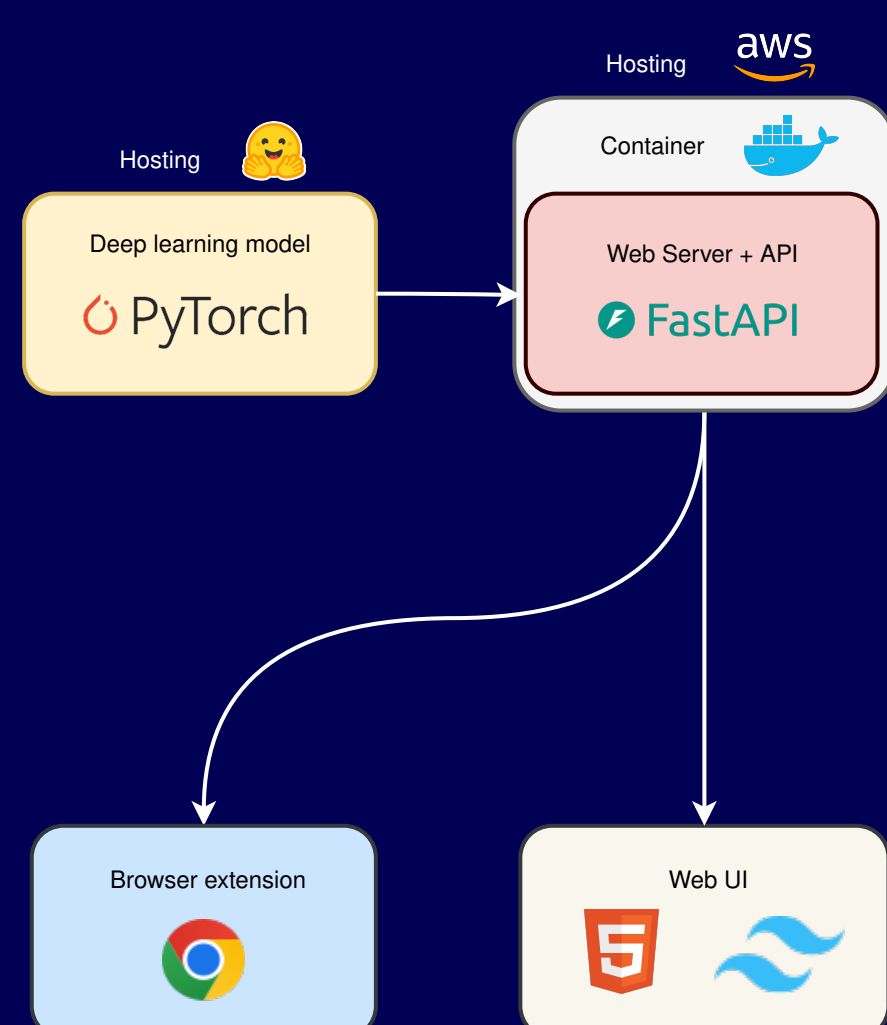
Results:

Team:
The Penguins

Nguyen Phuc Nguyen
Ngo Viet Hung
Nguyen Viet Phap

Supervisors:
Dr. Kapil Dev
Dr. Sam Goundar

RMIT UNIVERSITY — School of Science, Engineering & Technology