Facial Emotion Recognition from Video Data Using Deep Learning and Computer Vision Techniques

University of HUDDERSFIELD
Inspiring global professionals

Author: Dirdh Prafullkumar Patel
Supervisor: Dr.Tianhua Chen
MSc. Artificial Intelligence(2023-24) ,University of Huddersfield

Motivation

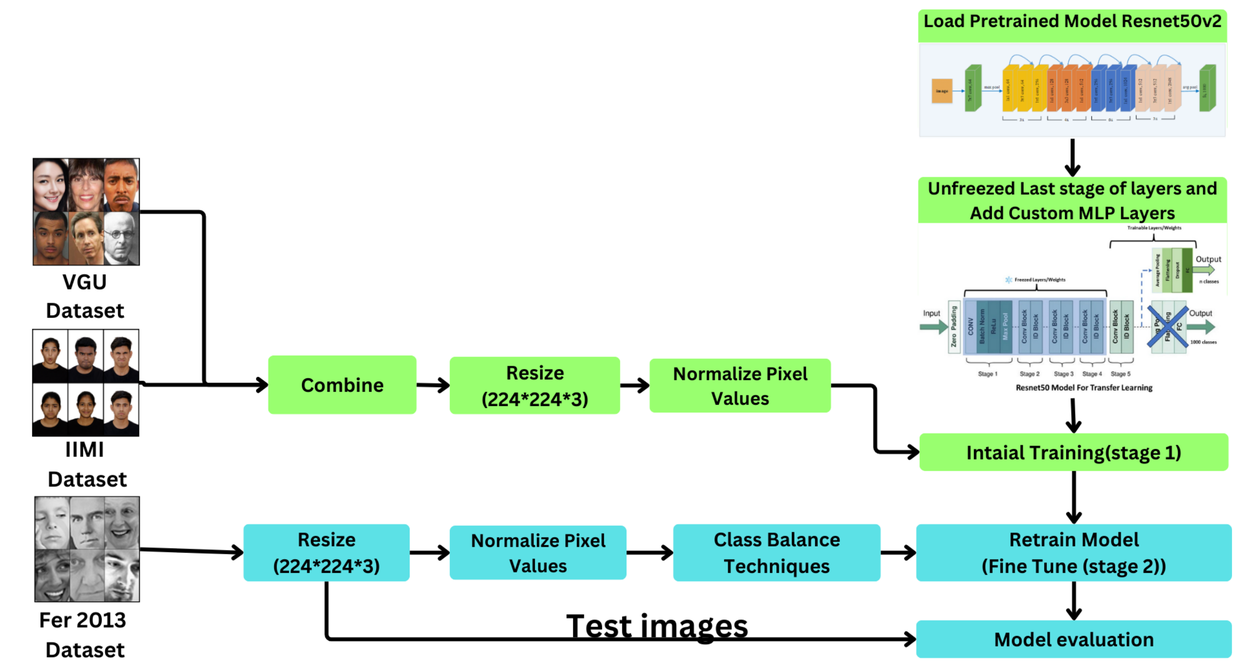
- Growing reliance on virtual communication highlights the need for accurate emotion recognition systems.
- **Key applications:** Telehealth therapy , Virtual meetings, and Human-computer interaction.
- **Current challenges:** Real-time processing, handling complex emotions, and generalization across facial expressions.

Objective

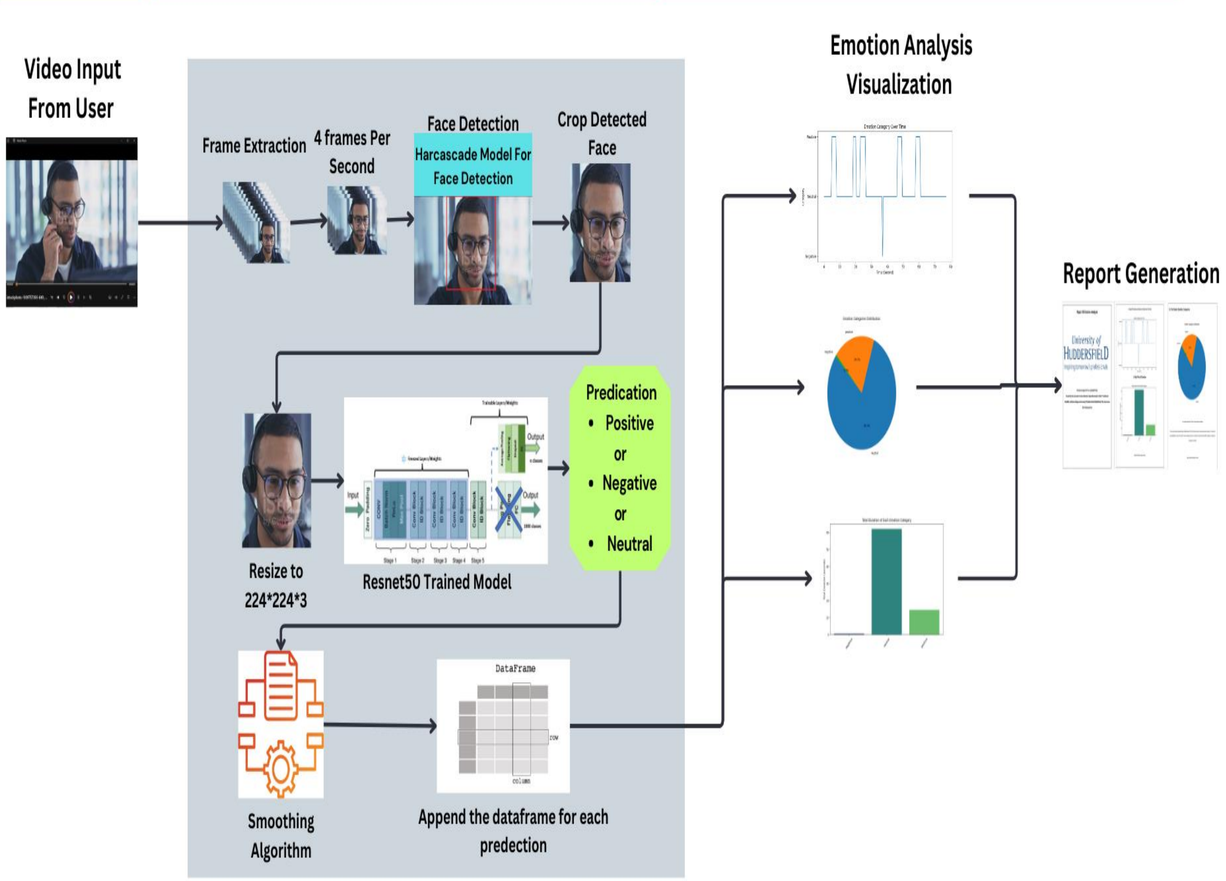
- Develop a deep learning-based system using a CNN-based architecture for emotion detection from video data with minimal computing resource expense.
- Ensure robust performance in detecting both static and dynamic facial expressions.
- Provide emotional trend visualizations and generate detailed reports.

Methodology

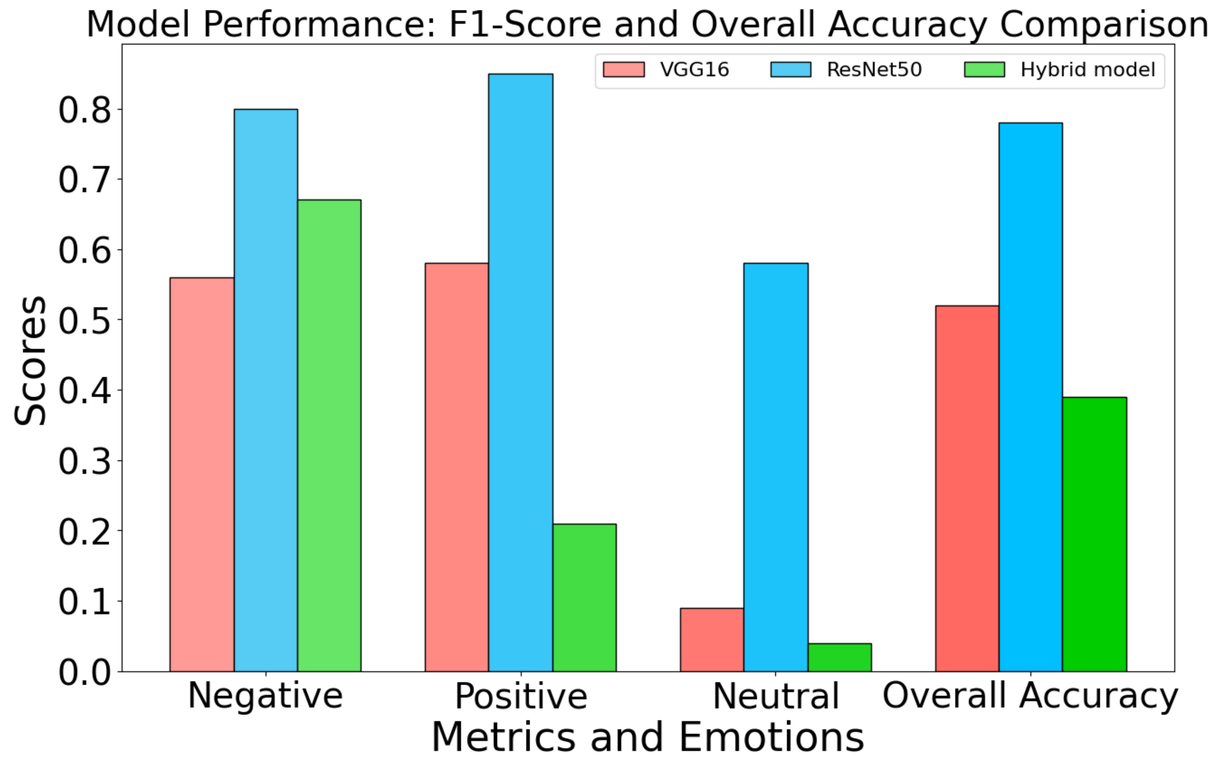
1. **Data Collection:**
   - Datasets: VGU, IIMI, FER2013.
   - Preprocessing: Resizing, normalization, and class-weight balancing to handle emotion imbalances.
2. **Model Architecture:**
   - ResNet-50 with transfer learning for emotion detection.
   - Initial training on VGU and IIMI, fine-tuned on FER2013 for improved accuracy.



3. System Architecture:

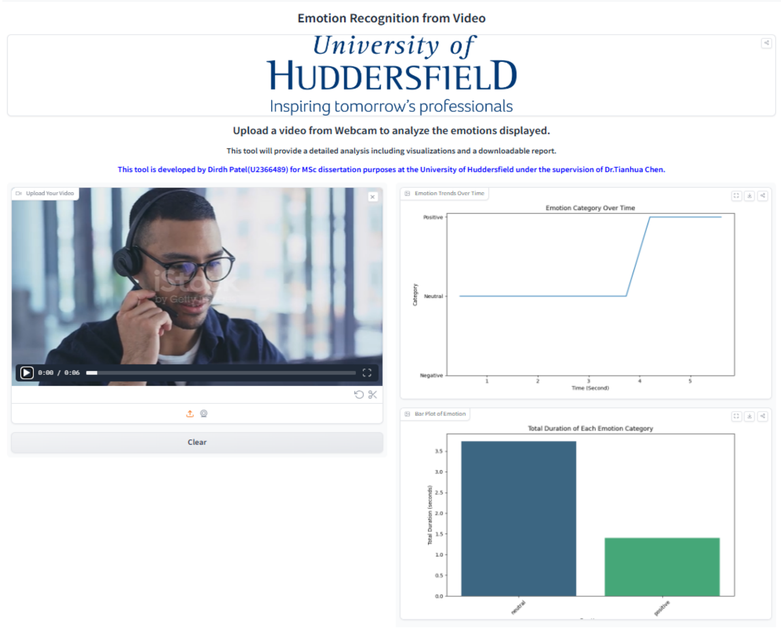


Results



- ResNet-50 performed best, achieving 78% accuracy and F1-scores of 0.80 (Negative), 0.85 (Positive), and 0.58 (Neutral).
- **Manual testing on 100 videos yielded 95% accuracy after applying a smoothing algorithm.**
- The system handles mixed-emotion scenarios effectively, providing accurate real-time results.

User Interface



- Real-time emotion detection with interactive visualizations (line graphs, bar charts, pie charts).
- Generates CSV and PDF reports for detailed analysis.
- Simple, user-friendly interface for video upload and instant emotional trend tracking.

Future Work

- Expand datasets for improved cultural and demographic diversity.
- Develop a multi-task model to recognize both face and facial emotions, generating individualized reports.
- Enhance the smoothing algorithm for detecting subtle emotional transitions in real-time.
- Explore real-time emotion detection for live video feeds.

Conclusion

The system utilizes a CNN-based architecture optimized through transfer learning with ResNet-50, leveraging pre-trained weights to reduce training time and computational costs. The training strategy involves fine-tuning with smaller datasets like VGU and IIMI, followed by retraining on FER2013 to improve accuracy. To ensure minimal computing resource expense, the model is designed for efficient real-time performance. The system is deployed on Hugging Face, offering easy access and scalability. For the user interface, Gradio is integrated, allowing users to upload or record videos, view real-time emotion analysis, and export detailed reports in CSV and PDF formats. This setup provides a user-friendly, accessible platform with minimal resource overhead.

Acknowledgments

I would like to express my heartfelt gratitude to my supervisor, Dr.Tianhua Chen, for his invaluable support and guidance throughout this project. Additionally, I extend my thanks to **the University of Huddersfield** for providing the resources and opportunities that made this research possible.