Extending Lightweight Driver FER: A Video-Based Approach

Brandenburg University of Technology Cottbus - Senftenberg

Aditya Chaudhary, Saadi Ibtissam, Douglas W. Cunningham

Department of Graphical Systems, Univ. BTU Cottbus-Senftenberg, Cottbus, Germany

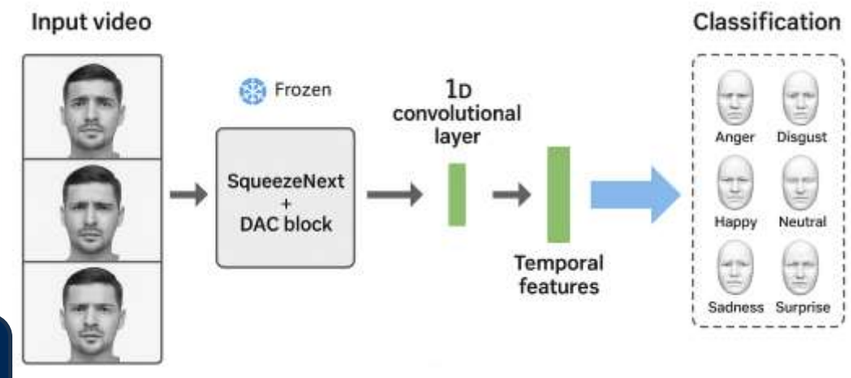
1. INTRODUCTION

Uddin (2025) [1] presents the DALDL model, optimized for efficient facial emotion recognition (FER) in real-time ADAS. This research extends DALDL by integrating temporal aggregation for video-based FER, enhancing dynamic emotion detection.

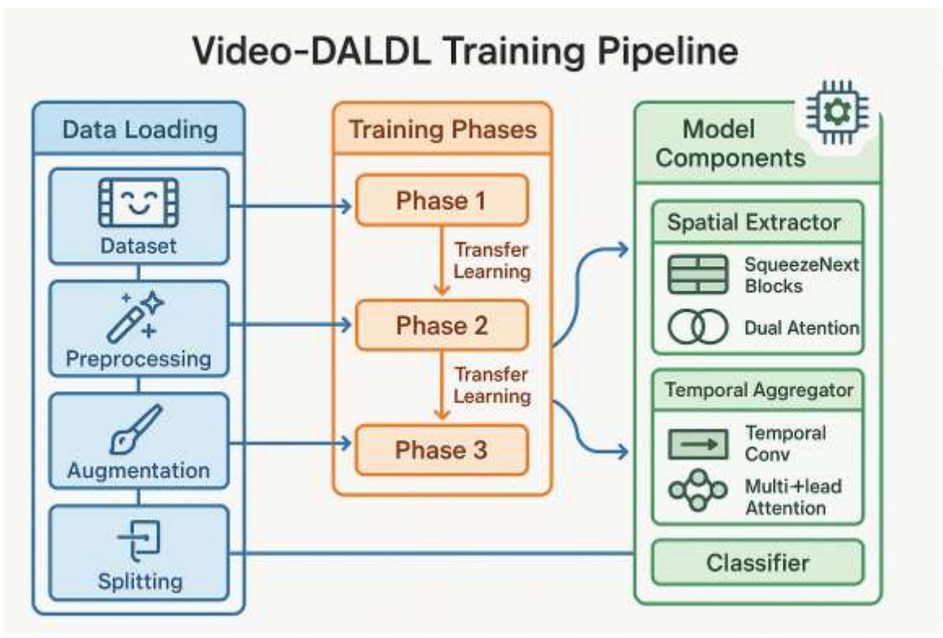
- Leverages SqueezeNext with DAC block.
- Compact: 0.75M params, 3.9ms inference.
- Handles 3-5 frame video sequences.

2. RESEARCH QUESTION

DALDL processes static images, missing temporal dynamics of driver emotions. Can video-based FER with temporal aggregation enhance ADAS reliability by capturing expression transitions in real-world driving scenarios?



3. METHODOLOGY



- **Objective:** Process 3-5 frame videos with DALDL.
- **Approach:** DALDL for spatial features, 1D Conv for temporal aggregation, classify 7 emotions
- **Why DALDL?** Lightweight, accurate, ADAS-fit.
- **Dataset:** KMU-FED (drivers' FEs, NIR camera).
- **Data Prep:** Group 3-5 frames into sequences; preprocess with resizing, equalization, grayscale conversion, and normalization.
- **Training:** 80-20 split, 100 epochs, Adam.

4. Results

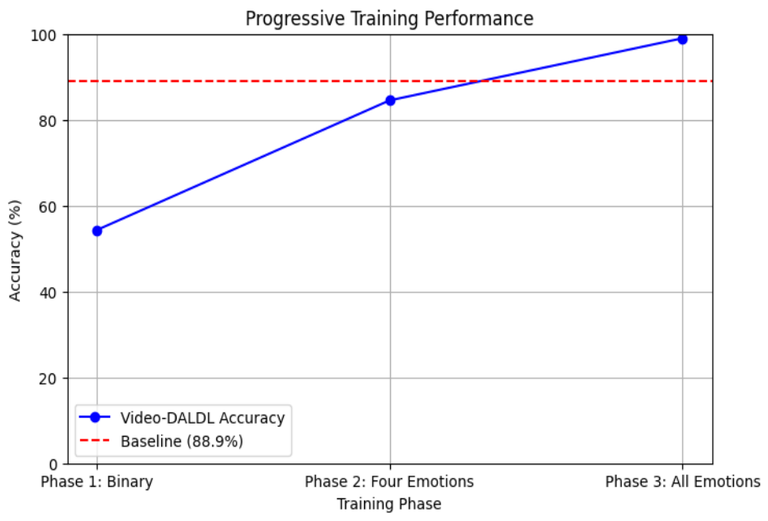
Progressive Training Results

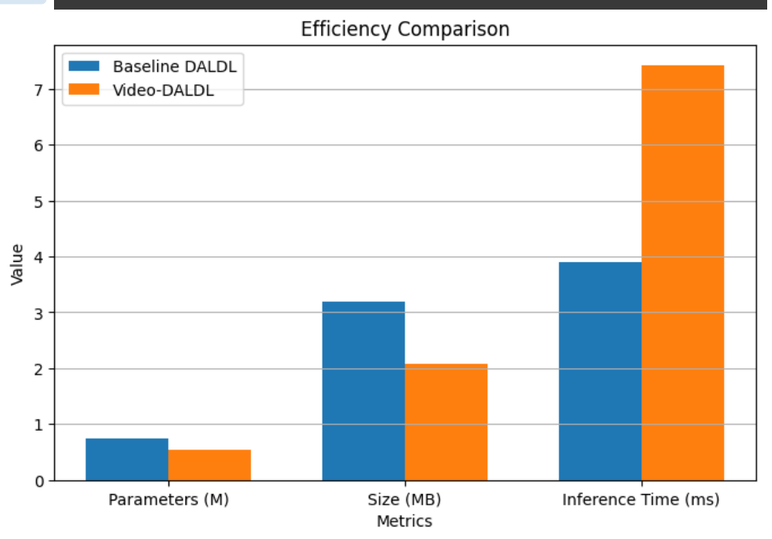
Phase 1 (Binary): Achieved 54.29% accuracy.

Phase 2 (Four emotions): Achieved 84.51% accuracy.

Phase 3 (All emotions): Achieved 98.99% accuracy (10.09% above 88.9% baseline).

```
GPU Memory: 0.02 GB
Early stopping at epoch 42
Phase completed. Best validation accuracy: 98.99%

============================================================
FINAL RESULTS - Memory-Optimized Video-DALDL Extension
============================================================
Phase 1: 54.29% ✅
Phase 2: 84.51% ✅
Phase 3: 98.99% ✅

Final Model Statistics:
Total Parameters: 543,749
Model Size: 2.07 MB
Final GPU Memory Usage: 0.02 GB

Inference Speed Test:
Average Inference Time: 7.42 ± 1.09 ms

============================================================
TARGET ASSESSMENT vs Uddin (2025) Baseline
============================================================
Final Accuracy: 98.99% (Target: >88.9%)
Model Size: 2.07 MB (Target: <3.3 MB)
Inference Time: 7.42 ms (Target: <5 ms)
```





5. Conclusion and Future Work

- **Conclusion:** This demonstrates the effectiveness of combining video-based temporal processing with lightweight deep learning architectures for robust performance in dynamic driving environments.
- **Future Work:** Optimize the model's inference time through techniques like quantization and pruning to meet real-time ADAS requirements under 5ms.

Contact :
chaudadi@b-tu.de

References : [1] Uddin, J. A Novel Lightweight Deep Learning Approach for Drivers' Facial Expression Detection. *Designs* **2025**, *9*, 45. https://doi.org/10.3390/designs9020045