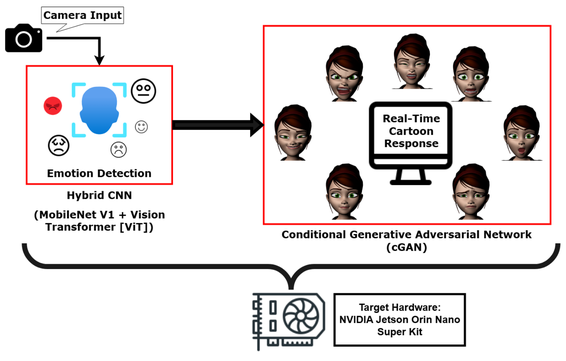
Real-Time Facial Expression Recognition & Response System for Children

Pavan Sai Guntha, Vidhaya Datta Reddy Metta

Background & Motivation

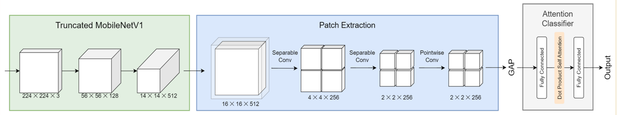
- Children often struggle to recognize and express emotions, especially at a young age or with developmental differences.
- Creating an environment that responds to their feelings in real time can greatly support emotional growth and well-being.
- By using animated characters that react naturally to children's expressions—like happiness, sadness, anger, or surprise—we can create more engaging and empathetic digital experience.
- Our project uses real-time emotion classification to power these interactions, helping children feel seen, understood, and emotionally supported.

System Architecture



- The frontend focuses on detecting the child's facial expressions using a hybrid Convolutional Neural Network (CNN) model.
- A rule-based decision system maps the classified emotion to the backend for generating a therapeutic response for the kid.
- The backend employs a Conditional Generative Adversarial Network (cGAN) to create an animated character response that corresponds to the detected emotion.
- The entire system is designed and optimized to run on the **NVIDIA Jetson Orin Nano** as the target hardware platform.

Facial Expression Recognition [FER]



Goal:

Accurately and efficiently classify the child's facial expressions captured by a camera in real-time, generating a recognized emotion label for the backend to process.

Model Architecture:

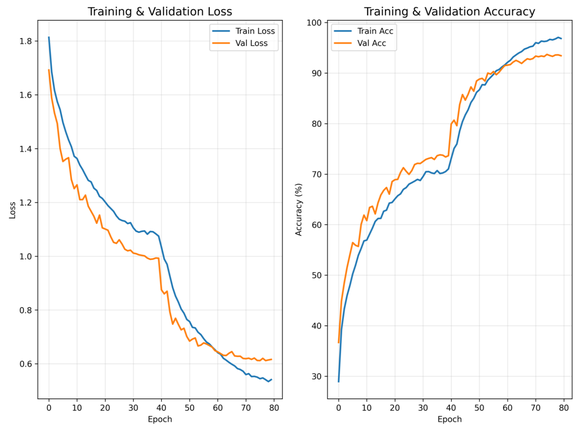
- **Truncated MobileNetV1**: Uses pre-trained MobileNetV1 as lightweight feature extractor, truncated at *14×14×512* to preserve efficiency while leveraging learned representations.
- **Patch Extraction**: Applies depthwise separable convolutions to split feature maps into *2×2×256* patches, reducing parameters while capturing local facial features.
- **Attention Classifier**: Employs multi-head self-attention mechanism instead of standard FC layers to dynamically weight patch features for improved expression classification.

Dataset Preparation:

RAF-DB, CK+, and FER+ datasets are used to train this model.

- **Face Detection Filtering**: Applied MTCNN face detection with 0.8 confidence threshold to filter out invalid samples.
- **Cross-Dataset Normalization**: Unified the three datasets by calculating individual dataset statistics and normalizing all images to shared mean/standard deviation values.

Achieved Test Accuracy of 93.71%



Rule-Based Emotion Mapping System

Goal:

Generate research backed emotional regulation sequences for the detected emotion class, following these principles:

- **Validation first**: Mirror child's emotion to show understanding.
- **Gradual regulation**: Slowly guide toward calmer states.
- **Self-soothing modeling**: Demonstrate healthy coping.
- **Positive reinforcement**: Support emotional growth.

Example therapeutic response:

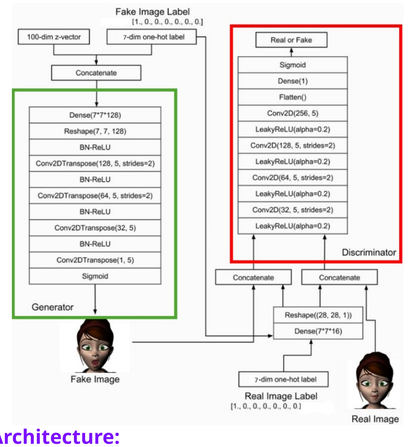
Anger: anger(20) → neutral(25) → happy(30) → neutral(15)

Fear : neutral(20) → fear(15) → neutral(25) → happy(25)

Animated Character Response Generation

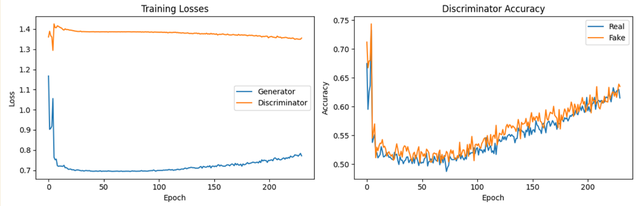
Goal:

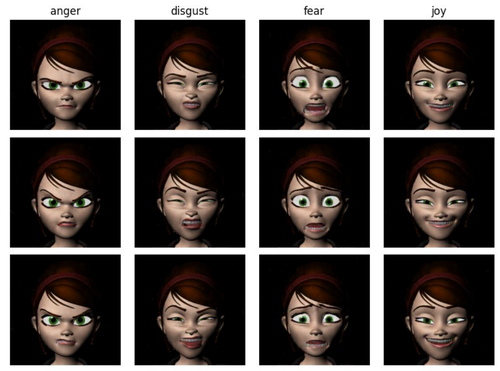
Accurately and efficiently generate a real-time animated character response as per the instructions from rule-based emotion mapping system.



Model Architecture:

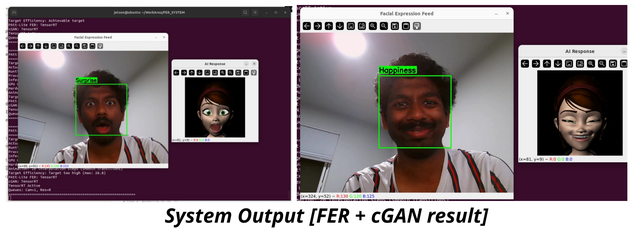
- **Conditional GAN Design:** 256×256 facial expression generator with 6-stage transposed convolution upsampling (*4×4→256×256*) and corresponding discriminator with label embedding conditioning for emotion-specific generation across 7 emotion classes.
- **Training Strategy:** Dual generator updates per discriminator step with diversity loss (*λ=0.15*) to prevent mode collapse and encourage intra-class variation and label smoothing.





cGAN generated samples

System Results



System Output [FER + cGAN result]

Conclusion

Developed an end-to-end engaging system for children, achieving **93.71%** test accuracy for FER and **~30 FPS** real-time animated character response using TensorRT acceleration on Nvidia Jetson Orin Nano.

References

[1] J. L. Ngwe, K. M. Lim, C. P. Lee, T. S. Ong, and A. Alqahtani, "PAtt-Lite: Lightweight Patch and Attention MobileNet for Challenging Facial Expression Recognition," *IEEE Access*, vol. 12, pp. 79327–79341, 2024, doi: 10.1109/access.2024.3407108

[2] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." *arXiv preprint arXiv:1411.1784* (2014).