Facial Emotion Recognition Using ResNet-CBAM on FER2013 Dataset

Overview

We aim to build a deep learning model for facial emotion recognition (FER) using the FER2013 dataset from Kaggle. The target application is real-time emotion recognition through webcam integration, suitable for deployment on devices such as smart carts or interactive systems.

Our implemented model is based on ResNet-18 enhanced with CBAM (Convolutional Block Attention Module) to improve spatial and channel attention mechanisms. The dataset consists of 7 emotion categories: *angry, disgust, fear, happy, neutral, sad, surprise*.

Implementation Summary

Key Components:

- Dataset: FER2013 (from Kaggle), preprocessed into train/test/val folders.
- Augmentation: Applied via Albumentations for robust generalization (resize, flip, brightness/contrast, dropout, Gaussian blur).
- Model: Custom ResNetCBAM initialized from pretrained ResNet-18, with attention blocks added to convolutional layers.
- Loss Function: Weighted CrossEntropy with label smoothing (to address class imbalance).
- Training Strategy:
    1. **Stage 1:** Train the classifier head only (frozen backbone).
    2. **Stage 2:** Fine-tune the full network end-to-end.
- Evaluation: Accuracy, F1-score, and confusion matrix on test set.

Results

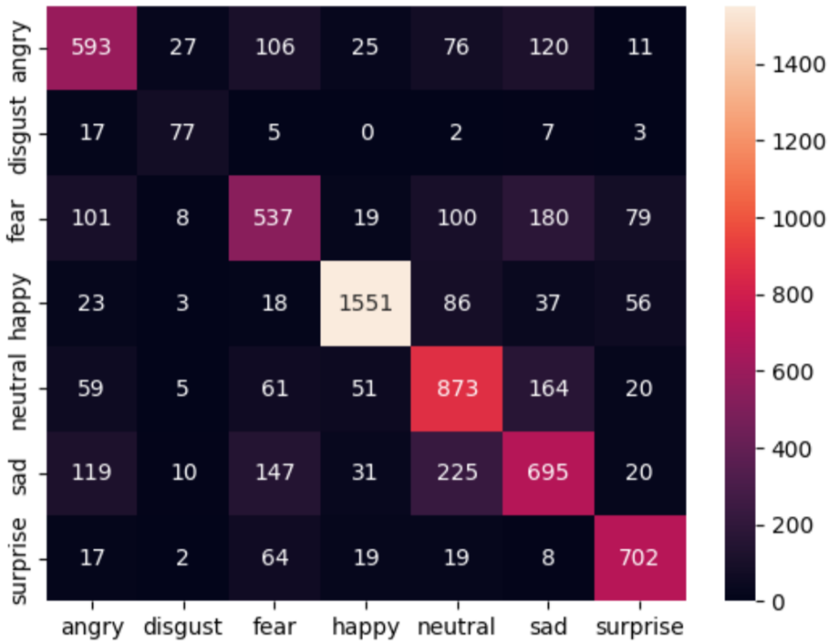
The model achieved the following performance on the test set:

- Final Accuracy: 70.0%
- Best F1-score: ~0.89 for *happy* class
- Macro F1-score (calculate the F1-score for each of the 7 emotions separately, then take the average): 0.68 across 7 emotions

Emotion	Precision	Recall	F1-score
Angry	0.64	0.62	0.63
Disgust	0.58	0.69	0.63
Fear	0.57	0.52	0.55
Happy	0.91	0.87	0.89
Neutral	0.63	0.71	0.67
Sad	0.57	0.56	0.57
Surprise	0.79	0.84	0.82

These results demonstrate significant improvement, especially in detecting positive emotions (*happy*, *surprise*), though negative emotions (*fear*, *sad*) still show room for enhancement.

The following confusion matrix visualizes the model's predictions compared to the true emotion labels on the test set:



Challenges

1. Class Imbalance: The FER2013 dataset is heavily imbalanced, with *disgust* having much fewer samples than *happy*.
   - Solution: Class weights and label smoothing were applied.
2. Low Initial Accuracy: Training started with very low accuracy (~15%), indicating difficulty in early convergence.
   - Solution: Multi-stage training (head first, then full model) helped stabilize learning.
3. Overfitting Risk: With a relatively small and imbalanced dataset, there's risk of overfitting.
   - Solution: Regularization via dropout, data augmentation, and early stopping.
4. Limited Resolution & Grayscale Images: FER2013 images are 48x48 grayscale.
   - Solution: Converted to RGB and resized to 224x224 to match pretrained ResNet input requirements.

Next Steps

1. **Real-Time Integration**
   We plan to connect the trained model to a webcam so it can recognize facial emotions live, frame by frame. To make this process faster and smoother, we may convert the model into a lighter format using TorchScript or ONNX, which helps speed up prediction during real-time use.

2. **Improving Accuracy**
   To enhance the model's performance, we aim to:

- Use more advanced architectures such as EfficientNet or Vision Transformers.
- Experiment with combining multiple models (ensembles) or adding facial landmarks to give the model more information.
- Fine-tune the model using higher-quality images or generate more training samples through data augmentation.

3. **Stabilizing Predictions**
   In real-time applications, emotion predictions may change rapidly from frame to frame. To reduce this flickering effect, we will apply temporal smoothing, such as averaging predictions over a few frames, to make the results more stable and reliable.

Conclusion

We successfully implemented and trained a CBAM-enhanced ResNet model for facial emotion recognition on the FER2013 dataset, reaching 70% test accuracy. While results are promising, real-time deployment and improving robustness, especially for negative emotions, remain key goals for the next phase of the project.