Facial Emotion Recognition Using CNN

Abstract:

Facial Emotion Recognition (FER) is a significant challenge in computer vision that involves identifying human emotions from facial expressions. This task is complex due to diverse facial features, environmental conditions, and varying emotional intensities across individuals. Accurate FER is critical in applications such as mental health monitoring, human-computer interaction, and surveillance systems. This study addresses the problem by leveraging the RAF-DB (Real-world Affective Faces Database), which includes over 15,000 labeled facial images across seven emotion categories: Happy, Sad, Angry, Fear, Disgust, Surprise, and Neutral. We designed and implemented a Convolutional Neural Network (CNN) model that automatically extracts features and classifies emotions from raw image data. The approach includes comprehensive preprocessing steps such as data normalization, class balancing using augmentation, and one-hot encoding to prepare the dataset for robust model training. The trained model achieved **84.31% training accuracy**, **83.84% validation accuracy**, and **83% test accuracy**, demonstrating high generalization capability. The evaluation through classification reports and a confusion matrix confirmed that the model performed consistently across multiple emotion classes, showing promising potential for real-world FER systems

Introduction:

Facial Emotion Recognition (FER) is a key area in affective computing and human-computer interaction. It has vital applications in healthcare, customer service, surveillance, and social robotics. Automating FER can help in identifying emotional cues effectively and enhance real-time decision-making. Deep learning methods, especially Convolutional Neural Networks (CNNs), have shown state-of-the-art performance in visual tasks, including FER.

This project used the RAF-DB dataset—a collection of real-world facial images annotated with seven emotions—to train and evaluate a CNN model capable of classifying emotions accurately.

Background or Literature Review:

Traditional approaches to FER involved handcrafted features and shallow classifiers such as SVMs and KNN. However, these methods lacked generalization capabilities and struggled with noisy, real-world data. Deep learning models, particularly CNNs, outperformed classical methods due to their ability to automatically learn complex features directly from data. Previous studies using RAF-DB have highlighted challenges related to class imbalance and variability in facial expressions due to age, ethnicity, lighting, and occlusions.

Methods and Materials:

- **Dataset:** RAF-DB (Real-world Affective Faces Database), annotated with 7 emotions: happy, angry, sad, fear, disgust, surprise, and neutral.
- **Preprocessing:**
    - **Merge:** Training and testing sets were combined for uniform processing.
    - **Resize:** Most images were already (100x100x3); resizing was not required.
    - **Shuffle:** Ensured randomness in training to prevent learning bias.
    - **Normalization:** Pixel values were scaled to [0, 1].
    - **Label Encoding:** One-hot encoding was applied after subtracting 1 for proper label indexing.
- **Class Balancing:**
    - **Oversampling and Data Augmentation:**
        - Overrepresented class ("happy") was downsampled.
        - Underrepresented classes were augmented with transformations (rotation, flip, zoom).
- **Model Architecture:**
    - Convolutional layers with ReLU activation and max pooling
    - Fully connected layers with dropout regularization
    - Softmax output layer for multiclass classification
- **Training:**
    - Optimizer: Adam
    - Loss: Categorical Crossentropy

Layer (type)	Output Shape	Param #
conv2d_8 (Conv2D)	(None, 98, 98, 32)	896
max_pooling2d_8 (MaxPooling2D)	(None, 49, 49, 32)	0
conv2d_9 (Conv2D)	(None, 47, 47, 64)	18,496
max_pooling2d_9 (MaxPooling2D)	(None, 23, 23, 64)	0
conv2d_10 (Conv2D)	(None, 21, 21, 128)	73,856
max_pooling2d_10 (MaxPooling2D)	(None, 10, 10, 128)	0
conv2d_11 (Conv2D)	(None, 8, 8, 512)	590,336
max_pooling2d_11 (MaxPooling2D)	(None, 4, 4, 512)	0
flatten_2 (Flatten)	(None, 8192)	0
dense_4 (Dense)	(None, 512)	4,194,816
dropout_2 (Dropout)	(None, 512)	0
dense_5 (Dense)	(None, 7)	3,591

Figure 1: CNN Model Architecture

Data and Results:

The proposed CNN model was trained for 60 with a batch size of 64. Training converged smoothly, as shown in Fig 2. The final training accuracy reached 84.31%, validation accuracy 83.84%, and test accuracy 83%.

Fig. 3 presents the precision, recall, and F1-scores for each emotion class on the test set. The model performed particularly well on each emotion.

Fig. 4 shows the confusion matrix, highlighting the class-wise distribution of errors. Data augmentation significantly helped improve minority class performance (Sad, Fear, Disgust).

A sample of correct and incorrect predictions is shown in Fig. 5, including the associated confidence scores. Misclassifications often occurred in low-light or partially occluded images.

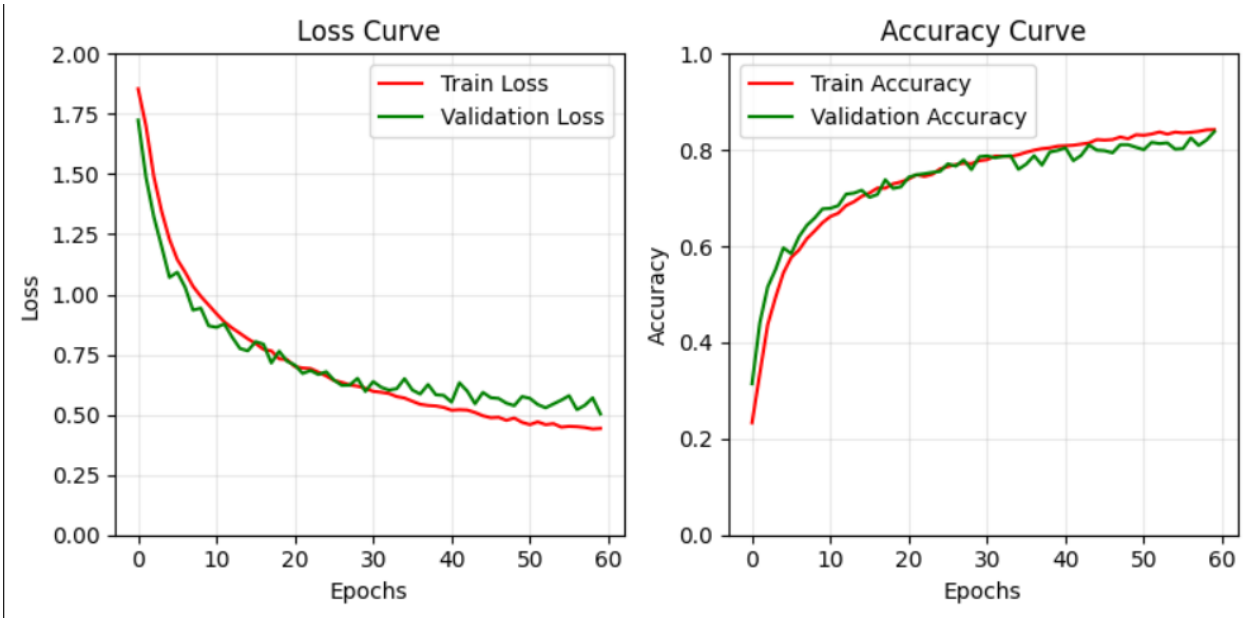

Figure 2: Loss and Accuracy Curve

Classification Report:

	precision	recall	f1-score	support
surprise	0.89	0.87	0.88	870
fear	0.91	0.95	0.93	840
disgust	0.74	0.83	0.78	910
happy	0.89	0.84	0.86	887
sad	0.78	0.74	0.76	865
angry	0.89	0.90	0.90	859
neutral	0.73	0.69	0.71	894
accuracy			0.83	6125
macro avg	0.83	0.83	0.83	6125
weighted avg	0.83	0.83	0.83	6125

Figure 3: Classification Report on Test Data

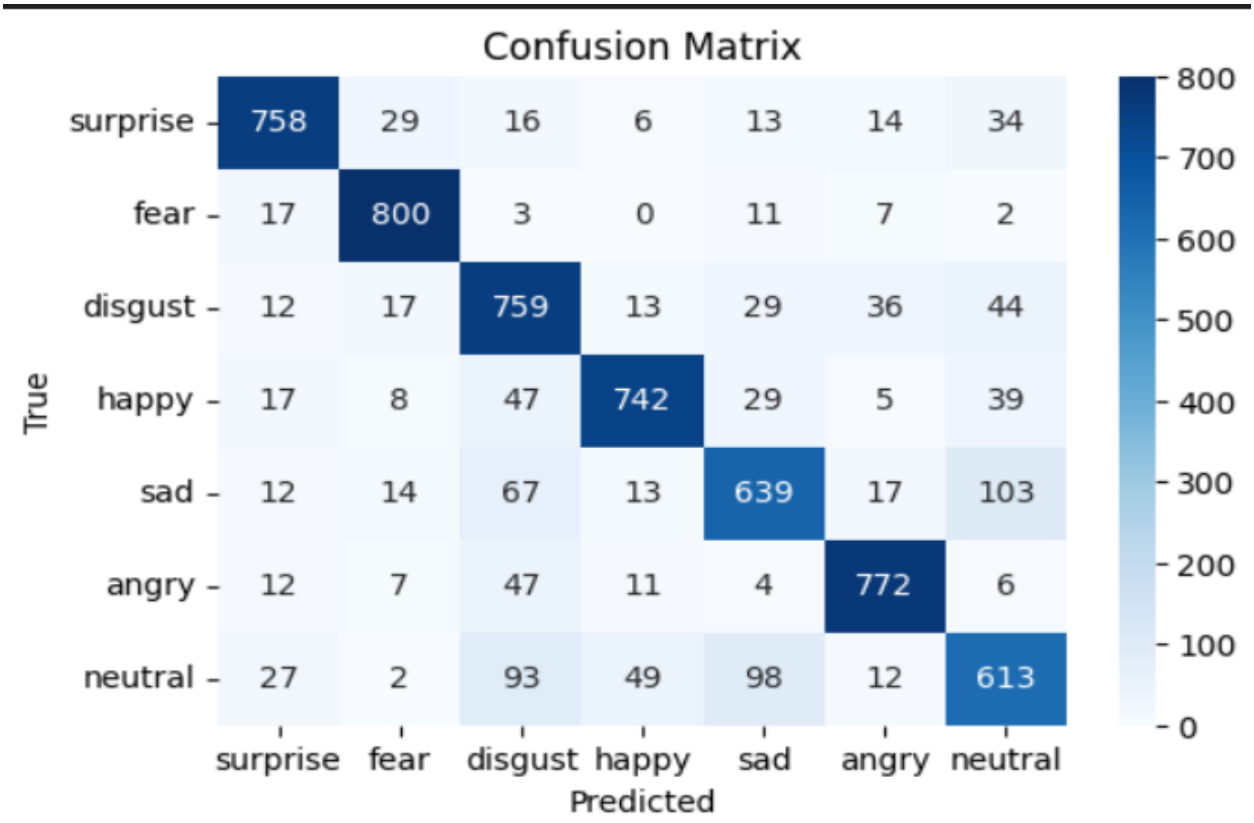

Figure 4: Confusion Matrix

Figure 5: Sample Predictions

These results demonstrated the model's high performance even on minority classes, validating the effectiveness of preprocessing and oversampling strategies.

Conclusion:

1: **Summary of the Findings:**

- CNN Model successfully classified facial emotions with 83% accuracy on real-world data.
- Oversampling and augmentation significantly improved recognition for minority classes.
- The model was robust and generalized well despite real-world variations in image data.

2: **Limitations of the Study:**

- The RAF-DB dataset still posed challenges due to occlusions and mixed emotions.
- Performance under varied lighting or backgrounds need further exploration.
- Transfer learning and ensemble models could potentially improve performance further.

References:

[1] Li, S., Deng, W., & Du, J. (2017). Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild. *IEEE CVPR*.

[2] Kaggle RAF-DB Dataset

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[4] Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.