Phase 2 Detailed Report: Implementation, Optimization, and Evaluation

Project: Facial Expression Recognition (FER) System

Course: Artificial Intelligence - K. N. Toosi University of Technology

Instructor: Dr. Pishgoo

Team Members: Soroush Soleimani, Parham Kootzari, Amirhossein Babaee

----

1. Evolutionary Architectural Design

In Phase 2, we moved beyond the baseline (Phase 1) to address the "Underfitting" problem. The complexity of facial muscle movements requires a high-capacity model to learn non-linear spatial features.

- **Deep CNN Architecture:** Our final model is a deep Convolutional Neural Network with approximately **5.8 Million parameters**. This scale was necessary to map the high-dimensional input of 48x48 pixels into 7 distinct emotion manifolds.

- **Feature Extraction Layers:** We utilized sequential convolutional blocks. Each block increases in depth (64, 128, 256, 512 filters) to capture everything from simple edges to complex facial structures like the curvature of a smile or the furrow of a brow.

- **Batch Normalization & Dropout:** To stabilize the internal covariate shift and force the network to learn redundant representations (preventing overfitting), we applied **BatchNormalization** and a high **Dropout** rate of 0.5 in the fully connected layers.

عکس

2. Advanced Training Strategy: The Turning Point

The most significant achievement in Phase 2 was the strategic control of the training process. We didn't just "run" the model; we managed its convergence.

ReduceLROnPlateau (The Key to 90% Accuracy): We observed that the model often got stuck in local minima. By implementing this callback, the system automatically reduced the Learning Rate by a factor of 0.2 whenever the validation loss stopped improving. This "fine-tuning" phase allowed the model to achieve a precision that a static learning rate could never reach.

Optimization Function: We used **Adam Optimizer**, which combines the benefits of AdaGrad and RMSProp, providing an adaptive gradient descent that is well-suited for image data.

Early Stopping: To ensure our model remains "Generalizable" and not just "Memorized," we used EarlyStopping to halt training exactly when the validation loss reached its global minimum.

3. Performance Analysis & Visualization

The training results exceeded our initial expectations, reaching a peak accuracy of ~**90%.**

Convergence Analysis: As shown in our training plots, the Training and Validation curves follow each other closely. This is a "Success Indicator" in Deep Learning, showing that our Data Augmentation from Phase 1 and our Regularization (Dropout) in Phase 2 worked in perfect harmony.

Loss Minimization: The categorical cross-entropy loss dropped significantly, proving that the model's "Confidence" in its predictions increased exponentially over 50+ epochs.

عکس

4. Final Evaluation & Metric Deep-Dive

Using the **evaluation.ipynb** notebook, we conducted a rigorous test on unseen data.

4.1. Confusion Matrix & Intraclass Challenges

The matrix shows that "Happy" and "Surprise" are almost perfectly classified. However, the model provides a fascinating insight into human psychology: it occasionally confuses "Sad" with "Neutral." This is a known challenge in the FER-2013 dataset where many "Neutral" faces have a slight downward lip curvature, mimicking "Sadness."

عکس

4.2. ROC Analysis (The Gold Standard)

Our ROC Curves are the strongest evidence of the model's power. Almost all classes achieved an **AUC (Area Under Curve) of 0.98 or higher**.

What this means: An AUC of 0.98 means that if we pick a random "Happy" image and a random "Angry" image, there is a 98% chance the model will rank them correctly. This confirms that our model's probability distribution is nearly ideal.

عکس

5. Final Conclusions

The project successfully met all industrial and academic criteria:

Resolution of Underfitting: Successfully scaled from a simple baseline to a 5.8M

parameter powerhouse.

Strategic Optimization: Proved that callbacks like ReduceLROnPlateau are essential for high-performance AI.

Deployment Readiness: With 90% accuracy and stable loss, the model is ready to be integrated into real-time applications (Webcam Demo).