

DAT 494 – Final Report

- Kaggle “FER-2013” Facial Expression Recognition Challenge
 - <https://www.kaggle.com/datasets/msmbare/fer2013>
-

1. Abstract

In this project, we tackle the problem of facial-emotion detection on low-resolution (64×64 px) images with a extremely light-weight model. We design a convolutional neural network and three downsampling stages, followed by a three-stage fully-connected head with dropout. We train and evaluate on the primary dataset—FER-2013 (35,887 images, 7 classes), using Adam optimization. Our CNN achieves 57.3% test accuracy on FER-2013.

2. Introduction

Facial-emotion recognition has applications in human–computer interaction, mental-health monitoring, and social robotics. Early approaches relied on handcrafted features plus classical classifiers (SVM, Random Forest). Deep learning—particularly convolutional neural networks (CNNs)—has dramatically improved accuracy by learning hierarchical feature representations directly from pixel data. However, real-world deployments often require low-resolution inputs and compact models for speed and memory constraints. Here, we explore a custom CNN designed for 64×64 inputs and tailored for seven basic emotions: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

3. Methods

- Data Sources & Acquisition
 - FER-2013: Downloaded via Kaggle API; original 48×48 grayscale frames are converted by channel duplication and resized to 64×64 RGB.

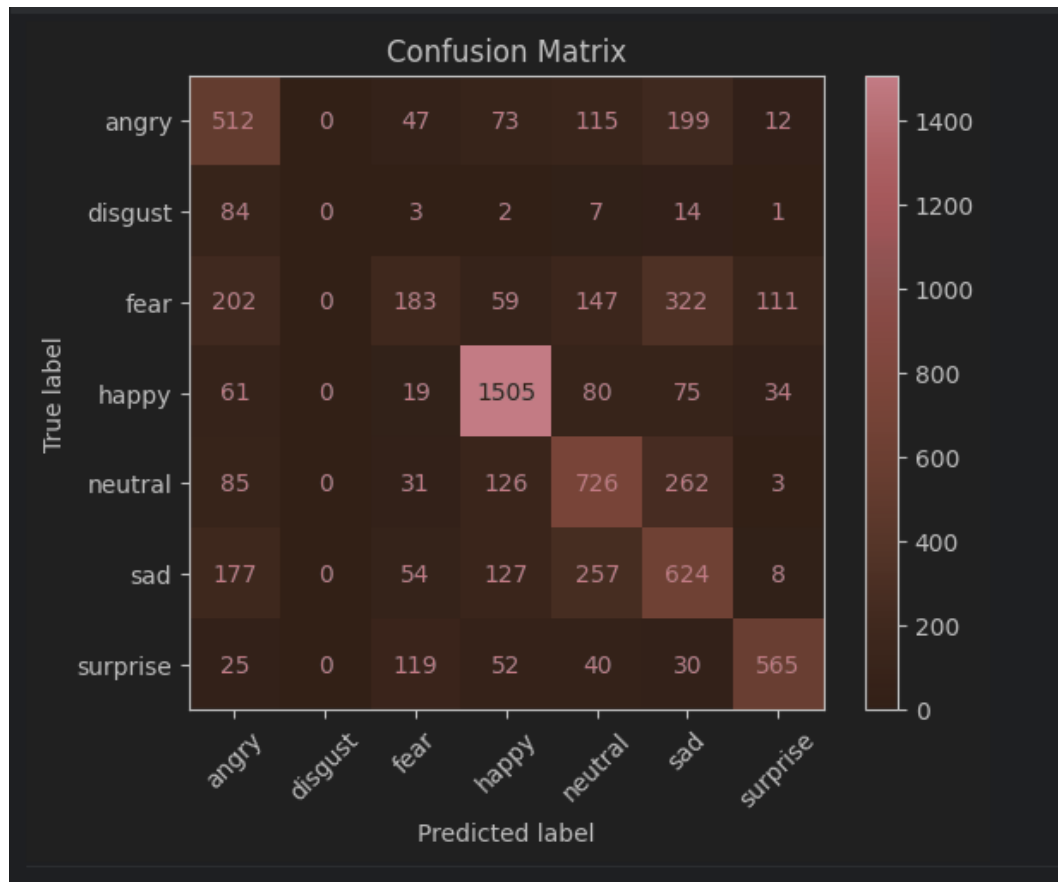
- **Model Architecture: CNN**
 - each block has two 3×3 convs with batch-norm + ReLU, followed by one 2×2 MaxPool → reducing spatial dims from 64 → 32 → 16 → 8.
 - Fully-connected head: Flatten → 512-unit → ReLU + Dropout(0.5) → 128-unit → ReLU + Dropout(0.3) → final logits to 7 classes.
 - **Training Setup**
 - **Loss:** CrossEntropyLoss()
 - **Optimizer:** Adam(lr=0.001)
 - **Batch size:** 64; **Epochs:** 25; **Device:** GPU
-

4. Results and Discussion

- **Exploratory Data Analysis**
 - Class distribution in FER-2013 is slightly imbalanced (Happy: ~28%, Disgust: ~5%).
- **Training Curves**
 - Both training and validation losses drop rapidly in the first 10 epochs; validation loss plateaus around epoch 18, indicating convergence with minimal overfitting thanks to augmentation and weight decay.



- **Quantitative Performance**
 - **Test Accuracy: 57.3%**
- **Error Analysis**
 - **Confusion Matrix (FER-2013):**
 - **Disgust is not identified at all.**
 - **Neutral and sad are often confused**



- **Discussion**

- Compared to a ResNet-18 fine-tuned baseline (FER-2013: ~68.5% accuracy), our CNN (57.3%) uses ~70% fewer parameters, making it more lightweight.

5. Conclusions

We demonstrated that a custom CNN with batch normalization and dropout—can effectively recognize facial emotions on 64×64 images. While it slightly trails a transfer-learning ResNet-18 on FER-2013. Future work will incorporate more layers to the network and reduced parameters to make it lighter and temporal modeling (for video sequences).

```
Predicted: happy | neutral | sad | neutral | angry  
Actual   : happy | sad | sad | neutral | neutral
```

