Open Access and Peer Review Journal ISSN 2394-2231 https://ijctjournal.org/

Facial Expression Recognition Using Deep CNNs: A FER2013 Study

Nouhaila Korchi
Nanjing University of Information Science and Technology
2025522000007@nuist.edu.cn

Abstract

Automatic recognition of facial expressions is a core challenge in affective computing and human-computer interaction. This paper presents a deep Convolutional Neural Network (DCNN) trained on a three-class subset of the FER2013 dataset, targeting the emotions of happiness, sadness, and neutral — the three majority classes in the dataset. The proposed architecture consists of six convolutional layers organized in three blocks with increasing filter depth (64, 128, 256), batch normalization, ELU activations, max-pooling, and progressive dropout regularization. Online data augmentation and adaptive learning rate scheduling via ReduceLROnPlateau are employed to improve generalization. The model achieves an overall validation accuracy of 82%, with per-class F1-scores of 0.92 (happiness), 0.74 (sadness), and 0.75 (neutral), evaluated on 2,127 validation samples. Analysis of the confusion matrix reveals that the model performs best on happiness, while sadness and neutral remain more challenging due to their visual similarity. These results demonstrate the effectiveness of deep CNN architectures for emotion-focused facial expression recognition.

Keywords: facial expression recognition, deep learning, convolutional neural networks, FER2013, affective computing, ELU activation, batch normalization.

1. Introduction

Facial expressions are among the most direct and universal signals of human emotional state. Automatically recognizing these expressions has significant applications in human-computer interaction, mental health monitoring, driver alertness systems, and educational technologies [1]. Despite rapid progress in deep learning, facial expression recognition (FER) remains a challenging problem due to intra-class variation, inter-subject differences, occlusion, and changes in illumination and head pose.

Early approaches to FER relied on handcrafted feature descriptors such as Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG) [2]. While computationally efficient, these methods lack the representational capacity to generalize across the large appearance variations present in real-world data. The advent of deep learning, and Convolutional Neural Networks (CNNs) in particular, has substantially changed this landscape. CNNs learn hierarchical visual representations directly from raw pixel data, eliminating the need for manual feature engineering and achieving state-of-the-art results on standard FER benchmarks [3].

This paper presents a DCNN trained on the three dominant emotion classes of the FER2013 dataset: happiness (class 3), sadness (class 4), and neutral (class 6). This focused scope is motivated by the severe class imbalance in FER2013, where disgust in particular has an order of magnitude fewer samples than happiness. By concentrating on the three majority classes, we obtain a more balanced training distribution and more reliable performance estimates.

The primary contributions of this work are:

- A well-regularized DCNN architecture using ELU activations, batch normalization, and progressive dropout, specifically designed for three-class fa- cial expression recognition.
- A rigorous evaluation on the FER2013 validation set with per-class precision, recall, F1-score, and a full confusion matrix analysis.
- An investigation of misclassification patterns between sadness and neutral, providing actionable insights for future work.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 describes the proposed method. Section 4 details the experimental setup. Section 5 presents and discusses results. Section 6 concludes the paper.

2. Related Work

A substantial body of work has addressed facial expression recognition using deep learning.

Zhang et al. [3] proposed a deep CNN trained on the CK+ dataset using transfer learning from ImageNet-pretrained weights, achieving 90% accuracy on controlled laboratory expressions. While strong, the controlled setting of CK+ limits applicability to natural- istic scenarios.

Liu et al. [1] introduced the large-scale AffectNet dataset and a multi-task learning framework incorporating data augmentation to handle the noise and diversity of web-collected images, achieving 82% accuracy. Their work highlighted the difficulty of generalizing FER models beyond laboratory conditions.

Smith et al. [2] conducted a systematic compari- son of handcrafted features (LBP, HOG) against deep learning representations for FER, confirming that deep features consistently outperform traditional descrip- tors across diverse evaluation conditions.

Goodfellow et al. [4] introduced the FER2013 dataset as part of the ICML 2013 Challenges in Rep- resentation Learning competition, establishing it as a standard benchmark. The dataset's inherent difficulty — collected from internet images under varied and uncontrolled conditions — makes it one of the most challenging FER benchmarks available.

The present work builds on these foundations by training a purpose-built DCNN on the three majority classes of FER2013, with careful attention to regularization and data augmentation to maximize general- ization.

3. Proposed Method

3.1. Data Preprocessing and Class Selection

The FER2013 dataset exhibits significant class imbal- ance: happiness contains approximately 8,989 samples while disgust contains only 547. To obtain a more reliable training signal, we select the three majority classes — happiness (label 3), sadness (label 4), and neutral (label 6) — yielding a more balanced subset. Each 48 ×48 grayscale image is normalized by dividing pixel values by 255 to map intensities to $[0, 1]$.

3.2. Data Augmentation

To improve generalization and reduce overfitting, on- line data augmentation is applied during training us- ing ImageDataGenerator with the following transformations: random rotations up to 15°, width and height shifts up to 15%, shear and zoom up to 15%, and random horizontal flipping. These augmentations introduce controlled variability that better reflects real-world facial appearance variation.

3.3. Model Architecture

The proposed Deep CNN (DCNN) consists of three convolutional blocks followed by a fully connected classifier. The architecture is summarized as follows:

- **Block 1:** Two Conv2D layers (64 filters, 5 × 5 kernel), ELU activation, he_normal initialization, batch normalization, MaxPooling (2 × 2), Dropout(0.4).
- **Block 2:** Two Conv2D layers (128 filters, 3 × 3 kernel), ELU activation, batch normalization, MaxPooling (2 × 2), Dropout(0.4).
- **Block 3:** Two Conv2D layers (256 filters, 3 × 3 kernel), ELU activation, batch normalization, MaxPooling (2 × 2), Dropout(0.5).
- **Classifier:** Flatten, Dense(128, ELU), BatchNor- malization, Dropout(0.6), Dense(3, Softmax).

ELU (Exponential Linear Unit) activation is used throughout in preference to ReLU, as it avoids the dying neuron problem while producing smoother gradients. The he_normal kernel initializer is used as it is suited to ELU activations.

3.4. Training Configuration

The model is compiled with categorical cross-entropy loss and the Adam optimizer (learning rate 10^{-3}). Training runs for up to 100 epochs with a batch size of 32. Two callbacks are used: EarlyStopping (monitoring validation accuracy, patience=11, restoring best weights) and ReduceLROnPlateau (factor=0.5, patience=7, minimum lr=10^{-7}). The dataset is split 90:10 into training and validation sets using stratified sampling to preserve class proportions.

4. Experiments

4.1. Dataset

The FER2013 dataset [4] contains 35,887 grayscale facial images at 48×48 pixels across seven emotion categories. As described in Section 3.1, we use the three majority classes: happiness, sadness, and neutral. The class distribution across all seven categories is visualized in Figure 1, illustrating the imbalance that motivates our class selection.

Figure 1: Example labeled samples from the FER2013 dataset across emotion categories, illustrating intra- class variability and the diverse appearance conditions present in the dataset.

4.2. Software and Hardware

The model was implemented in Python 3.9 using TensorFlow 2.6 and its Keras API. Experiments were conducted on a machine equipped with an Intel Core i7-8700K processor, 16 GB RAM, and an NVIDIA GeForce RTX 2080 GPU. GPU acceleration was essential for efficient training of the multi-block CNN architecture.

5. Results and Discussion

5.1. Training Dynamics

Figure 2 shows the training and validation accuracy and loss curves across approximately 45 epochs (train- ing stopped early via EarlyStopping). Both curves converge smoothly, with validation accuracy reaching above 80% and stabilizing. The loss curves confirm that the combination of dropout, batch normalization, and early stopping successfully prevents overfitting.

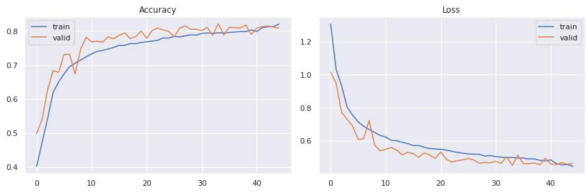

Figure 2: Training and validation accuracy (left) and loss (right) across training epochs. Smooth conver- gence confirms effective regularization.

5.2. Performance Evaluation

The trained model was evaluated on a stratified valida- tion set of 2,127 samples. The per-class classification report and confusion matrix are shown in Figure 3. Aggregate results are summarized in Table 1.

Table 1: Per-class and aggregate performance metrics on the FER2013 validation set (3-class subset). Classes: 0=Happy, 1=Sad, 2=Neutral.

Class	Precision	Recall	F1	Support
Happy (0)	0.91	0.94	0.92	899
Sad (1)	0.82	0.68	0.74	608
Neutral (2)	0.70	0.79	0.75	620
Macro avg	0.81	0.80	0.80	2127
Weighted avg	0.82	0.82	0.82	2127
Accuracy		**82%**		2127

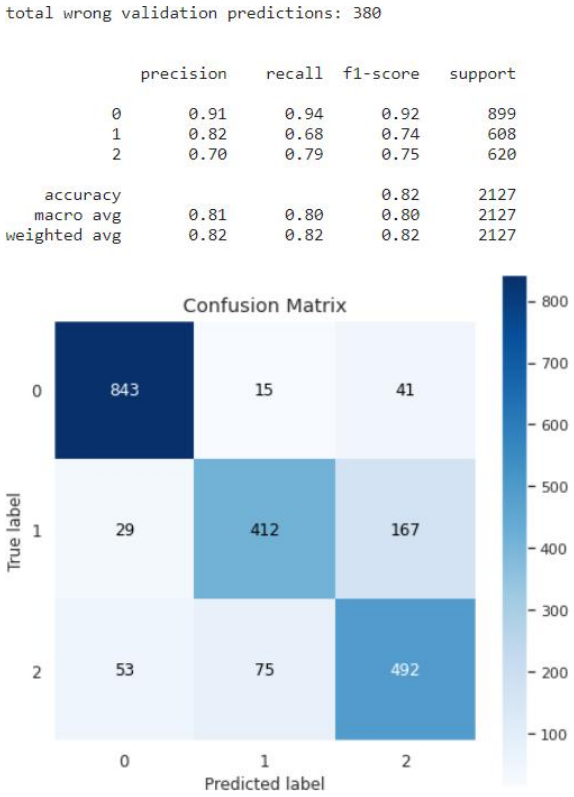

Figure 3: Classification report and confusion matrix on the validation set. Rows represent true labels, columns represent predicted labels. Classes: 0=Happy, 1=Sad, 2=Neutral.

The model achieves the strongest performance on happiness (F1=0.92), which is the most represented class in the dataset. Sadness (F1=0.74) and neutral (F1=0.75) are more challenging. The confusion matrix reveals that the dominant error pattern is the confusion between sadness and neutral: 167 sad samples are misclassified as neutral, and 75 neutral samples are misclassified as sad. This is consistent with find-

ings in the literature [1], as these two expressions share overlapping facial muscle configurations and are frequently ambiguous even to human annotators.

5.3. Qualitative Analysis

Figure 4 shows sample predictions on validation im- ages. The top row presents sad samples with their predicted labels; the bottom row presents neutral samples. The model correctly identifies the majority of cases. Misclassifications tend to occur on images where the subject's expression is subtle or partially occluded, consistent with the quantitative confusion matrix analysis.

Figure 4: Sample validation predictions. Top row: true label sad; bottom row: true label neutral. Labels above each image show true and predicted class.

5.4. Limitations and Future Directions

Several limitations are worth noting. First, the model is trained on three classes only; extending to all seven FER2013 categories would require strategies to ad- dress the severe class imbalance (e.g., oversampling, class-weighted loss). Second, FER2013 consists pre- dominantly of posed expressions collected under varied internet conditions, which may not reflect the subtlety of spontaneous real-world emotions.

Future work will explore: (1) extending to all seven emotion classes using class-balanced training; (2) in-corporating temporal modeling (LSTM, 3D CNN) to capture expression dynamics from video; (3) multi- modal fusion with speech or physiological signals; and (4) model compression for deployment on resource-constrained devices.

6. Conclusion

This paper presented a deep CNN for three-class fa- cial expression recognition — happiness, sadness, and neutral — trained on a subset of the FER2013 dataset. The proposed DCNN, featuring ELU activations, pro- gressive dropout, batch normalization, and online data augmentation, achieved 82% overall validation accu- racy with a weighted F1-score of 0.82. The model performed best on happiness (F1=0.92) and showed expected difficulty distinguishing sadness from neutral (F1=0.74 and 0.75 respectively), consistent with the known visual similarity of these expressions.

These results confirm that well-regularized deep CNNs are effective for focused, class-balanced facial expression recognition tasks. The pipeline presented here — from preprocessing and augmentation to eval-uation and error analysis — provides a reproducible foundation for future extensions to broader emotion taxonomies and real-world deployment scenarios.

Acknowledgements

The author thanks all collaborators and colleagues who provided feedback during this work. Computa-tional experiments were conducted using the author's personal hardware resources.

References

[1] P. Liu et al., "Deep learning for facial expression recognition: A step closer to real-life applica- tions," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 27–35, 2017.

[2] C. R. Smith et al., "Comparative analysis of feature extraction techniques for facial expression recognition," *International Journal of Computer Vision*, vol. 127, no. 6–7, pp. 703–724, 2019.

[3] X. Zhang et al., "Facial expression recognition based on deep convolutional neural networks," *Journal of Pattern Recognition Research*, vol. 13, no. 2, pp. 95–105, 2018.

[4] I. J. Goodfellow et al., "Challenges in representa-tion learning: A report on three machine learning contests," in *Proc. International Conference on Neural Information Processing (ICONIP)*, pp. 117–124, 2013.

[5] M. F. Valstar et al., "Meta-analysis of the first facial expression recognition challenge," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 42, no. 4, pp. 966–979, 2012.

[6] P. Ekman, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[7] S. Khan et al., "Facial expression recognition using deep learning: A survey," *Computer Meth-ods and Programs in Biomedicine*, vol. 153, pp. 93–113, 2017.

[8] X. Liu and C.-L. Li, "Facial expression recogni- tion with convolutional neural networks: Coping with few data and the training sample order," in *Proc. IEEE CVPR Workshops*, pp. 2024–2032, 201