Transfer Learning and Fine-Tuning for Image Classification

DeepWeeds and Facial Expression Recognition

Vaibhav Chourasia

January 6, 2026

1 Introduction

Convolutional Neural Networks (CNNs) have achieved state-of-the-art performance in image classification by learning hierarchical feature representations directly from data. However, training deep CNNs from scratch is computationally expensive and requires large labeled datasets. Transfer learning addresses this challenge by reusing models pre-trained on large-scale datasets such as ImageNet and adapting them to new tasks.

In this project, transfer learning and fine-tuning techniques are applied to two challenging image classification problems: weed species classification using the DeepWeeds dataset and facial expression recognition using the FER2013 dataset. The objective is to analyze and compare fine-tuning strategies while balancing classification performance and computational efficiency.

2 Datasets

2.1 DeepWeeds Dataset

The DeepWeeds dataset contains approximately 17,500 images of weed species collected in natural outdoor environments. The dataset includes nine classes such as *Chinee Apple*, *Lantana*, *Parthenium*, and *Snake Weed*. Images exhibit variations in lighting, background clutter, and viewpoint, making the classification task non-trivial.

2.2 FER2013 Dataset

The Facial Expression Recognition (FER2013) dataset consists of grayscale facial images categorized into seven emotion classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The dataset is characterized by low-resolution images, class imbalance, and subtle inter-class differences, which collectively increase the difficulty of the classification task.

3 Data Preprocessing and Augmentation

To improve generalization and reduce overfitting, data augmentation techniques were applied to the training sets of both datasets. Test sets were kept unchanged to ensure fair evaluation.

3.1 Augmentation Techniques

- Random resized cropping
- Horizontal flipping
- Random rotation
- Pixel normalization using ImageNet statistics

These augmentations increase sample diversity and encourage the model to learn robust, invariant features.

4 Model Selection

ResNet18 was selected as the base architecture for both tasks. ResNet models utilize residual connections that alleviate the vanishing gradient problem and enable stable training.

ResNet18 was chosen due to:

- Proven effectiveness in image classification
- Lower computational complexity compared to deeper variants
- Suitability for CPU-based training environments

5 Transfer Learning and Fine-Tuning Strategies

Two fine-tuning strategies were employed to study the trade-off between accuracy and computational cost.

5.1 DeepWeeds: Partial Fine-Tuning

For the DeepWeeds dataset, ImageNet pre-trained weights were used. The convolutional backbone was frozen, and only the final fully connected layer was trained. This approach leverages general visual features learned from ImageNet while reducing training time.

5.2 FER2013: Partial Layer Fine-Tuning

For FER2013, multiple fine-tuning strategies were explored, including:

- Fully frozen backbone (classifier-only training)
- Partial fine-tuning by unfreezing the final convolutional block

Due to CPU-only constraints and the large dataset size, full fine-tuning was computationally infeasible. Partial fine-tuning of the final block provided a balance between feature adaptation and manageable training time.

6 Training Configuration

Both models were trained using the Adam optimizer and cross-entropy loss.

- Optimizer: Adam
- Loss Function: Cross-Entropy Loss
- Batch Size: 32
- Epochs: 10 (DeepWeeds), 5 (FER2013)

Training was stopped once accuracy improvements began to plateau.

7 Results and Evaluation

Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrices.

7.1 DeepWeeds Results

The DeepWeeds model achieved strong classification performance using partial fine-tuning of a pre-trained ResNet18 backbone. Accuracy increased steadily across epochs and plateaued after the eighth epoch.

Table 1: DeepWeeds Classification Performance

Metric	Value
Accuracy	78.0%
Weighted F1-score	0.78
Macro F1-score	0.71

Class-wise analysis shows high recall for dominant classes such as *Negative*, while moderate confusion occurs between visually similar weed species. Overall, the results demonstrate effective transfer of ImageNet features to natural outdoor imagery.

7.2 FER2013 Results

The FER2013 model exhibited gradual improvement with partial fine-tuning of the final convolutional block. Accuracy improved consistently across epochs, reaching 44.0% after five epochs.

Table 2: FER2013 Classification Performance

Metric	Value
Accuracy	44.0%
Weighted F1-score	0.41
Macro F1-score	0.34

The confusion matrix reveals strong recall for the *Happy* class due to dataset imbalance, while minority classes such as *Disgust* remain challenging. These outcomes align with known characteristics of the FER2013 dataset.

8 Discussion

The experiments demonstrate that partial fine-tuning can achieve a favorable balance between performance and computational efficiency. DeepWeeds benefited significantly from pre-trained ImageNet features, whereas FER2013 required limited feature adaptation to improve accuracy. Increasing the depth of fine-tuning yielded diminishing returns under CPU-only constraints.

9 Conclusion

This project successfully applied transfer learning and fine-tuning techniques to two real-world image classification tasks. By carefully selecting model architectures and fine-tuning strategies, reasonable performance was achieved within limited computational resources. The comparison highlights the importance of adapting training strategies to dataset characteristics and hardware constraints.

Future work may involve GPU-based training, deeper architectures, and techniques to address class imbalance in FER2013.