# 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 demonstrated remarkable success in image classification by learning hierarchical feature representations directly from pixel data. However, training deep CNNs from scratch is computationally expensive and typically requires very large labeled datasets.

Transfer learning addresses this limitation by reusing knowledge learned from large-scale datasets such as ImageNet and adapting it to new tasks. Fine-tuning strategies further allow selective adaptation of pre-trained features depending on dataset characteristics and computational constraints.

In this project, transfer learning and fine-tuning are applied to two distinct image classification problems:

- Weed species classification using the DeepWeeds dataset
- Facial expression recognition using the FER2013 dataset

The goal is to analyze how different fine-tuning strategies affect performance while maintaining computational feasibility.

## 2 Datasets

### 2.1 DeepWeeds Dataset

The DeepWeeds dataset contains approximately 17,500 images of weed species collected in natural outdoor environments. The dataset consists of nine classes including *Chinee Apple*, *Lantana*, *Parthenium*, and *Snake Weed*.

Images exhibit significant variation in lighting conditions, background clutter, and camera viewpoints. Such variability makes DeepWeeds a realistic and challenging dataset well-suited for evaluating transfer learning from ImageNet-trained models.

### 2.2 FER2013 Dataset

The Facial Expression Recognition (FER2013) dataset contains grayscale facial images categorized into seven emotion classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

FER2013 presents several challenges:

- Low-resolution images
- Subtle inter-class differences
- Strong class imbalance (e.g., dominance of the Happy class)

These characteristics make facial expression recognition particularly difficult and sensitive to model design and fine-tuning strategy.

# 3 Data Preprocessing and Augmentation

Data preprocessing and augmentation were applied to improve generalization and reduce overfitting. Augmentation was applied only to training data, while test sets were left unchanged to ensure fair evaluation.

## 3.1 Augmentation Techniques

The following augmentation techniques were employed:

- Random resized cropping
- Horizontal flipping
- Random rotation
- Pixel normalization using ImageNet mean and standard deviation

For DeepWeeds, additional color jittering was used to model illumination changes common in outdoor environments. For FER2013, simpler augmentations were applied due to the grayscale nature and facial alignment of the images.

# 4 Model Selection

ResNet18 was selected as the backbone architecture for both datasets. Residual networks introduce skip connections that mitigate the vanishing gradient problem and enable stable optimization.

ResNet18 was chosen specifically because:

- It provides strong representational capacity with relatively low computational cost
- It is well-suited for CPU-based training environments
- It allows clear experimentation with different fine-tuning strategies

Deeper variants such as ResNet50 were avoided due to increased training time and hardware limitations.

# 5 Transfer Learning and Fine-Tuning Strategies

Different fine-tuning strategies were adopted for the two datasets based on their characteristics and computational constraints.

## 5.1 DeepWeeds: Feature Extraction with Frozen Backbone

For DeepWeeds, a ResNet18 model pre-trained on ImageNet was used. All convolutional layers were frozen, and only the final fully connected classification layer was trained.

This strategy treats the pre-trained network as a fixed feature extractor. ImageNet-trained filters capture generic visual patterns such as edges, textures, and shapes, which transfer effectively to natural outdoor imagery like weeds.

This approach significantly reduces training time and the risk of overfitting while still achieving strong performance.

## 5.2 FER2013: Partial Fine-Tuning of Final Convolutional Block

For FER2013, a different strategy was adopted. The network was initialized without ImageNet weights, and only the final convolutional block (layer4) along with the classifier was trained.

Using ImageNet pre-trained weights was avoided due to the domain mismatch between natural images and grayscale facial expressions. Instead, partial fine-tuning was used to allow higher-level features to adapt specifically to facial patterns while keeping earlier layers frozen to limit computational cost.

This strategy provides a compromise between full training and purely fixed feature extraction.

# 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)

Learning rate scheduling was applied for DeepWeeds to encourage stable convergence. 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 performance using frozen ImageNet features with a newly trained classifier.

Table 1: DeepWeeds Classification Performance

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

Class-wise evaluation indicates high recall for dominant classes such as *Negative*, while moderate confusion occurs among visually similar weed species. Overall, results demonstrate effective transfer of ImageNet features to real-world agricultural imagery.

### 7.2 FER2013 Results

The FER2013 model showed gradual improvement with partial fine-tuning, reaching a final accuracy of 44.0%.

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 performance for the *Happy* class, reflecting dataset imbalance, while minority classes such as *Disgust* remain challenging. These results are consistent with prior studies on FER2013.

## 8 Discussion

The experiments highlight the importance of selecting dataset-appropriate fine-tuning strategies. Feature extraction with frozen backbones proved highly effective for DeepWeeds, where ImageNet features closely align with the target domain.

In contrast, FER2013 benefited from limited feature adaptation due to its specialized facial domain and class imbalance. Full fine-tuning was not feasible under CPU-only constraints, making partial fine-tuning a practical compromise.

## 9 Conclusion

This study demonstrated the effectiveness of transfer learning and fine-tuning across two distinct image classification tasks. By tailoring fine-tuning strategies to dataset characteristics and hardware limitations, meaningful performance was achieved without excessive computational cost.

Future work may explore GPU-based training, deeper architectures, and techniques such as class-balanced loss functions to further improve facial expression recognition performance.