Transfer Learning and Fine Tuning for Image Classification

1 Introduction

Deep learning models trained from scratch require large amounts of labeled data and computational resources. Transfer learning addresses this limitation by reusing models that have already learned general visual representations from large scale datasets such as ImageNet. By adapting these models to new tasks through fine tuning, strong performance can be achieved even when task specific data is limited.

In this work, transfer learning and fine tuning techniques are applied to two image classification problems. The first task is facial expression recognition using the FER2013 dataset. The second task is weed species classification using the DeepWeeds dataset. For both tasks, different fine tuning strategies are evaluated to understand their impact on generalization performance.

2 Datasets

2.1 FER2013

The FER2013 dataset consists of grayscale facial images categorized into multiple emotional expressions such as happiness, sadness, anger, and fear. The dataset is moderately sized and exhibits class imbalance, making macro averaged evaluation metrics particularly important.

2.2 DeepWeeds

The DeepWeeds dataset contains images of various weed species captured in real world outdoor environments. The images show large variation in background, lighting, and soil conditions. This dataset is well suited for evaluating how robust a model is to environmental noise and visual diversity.

3 Model Selection

For the facial expression recognition task, a ResNet50 model pretrained on ImageNet was selected. Residual networks are known to perform well on structured visual tasks and are capable of capturing subtle spatial patterns required for facial analysis.

For the weed classification task, an EfficientNet B0 model pretrained on ImageNet was used. EfficientNet models provide a strong balance between accuracy and parameter efficiency, making them suitable for outdoor imagery where texture and color variation are important.

4 Data Augmentation

Data augmentation was applied during training to improve generalization.

For facial expression recognition, mild augmentations such as random cropping, horizontal flipping, and small rotations were used to preserve facial structure while introducing variability.

For the DeepWeeds dataset, stronger augmentations were applied including random resized cropping, horizontal flipping, small rotations, and mild color jitter. These augmentations help the model become robust to lighting and background variations common in outdoor images.

Validation data was not augmented and only standard resizing and normalization were applied.

5 Fine Tuning Strategies

Two fine tuning strategies were evaluated for both datasets.

5.1 Strategy A Partial Fine Tuning

In Strategy A, the early and intermediate layers of the pretrained network were frozen. Only the final convolutional block and the classification head were allowed to update during training. This approach limits the number of trainable parameters and preserves general visual features learned from ImageNet.

5.2 Strategy B Full Fine Tuning

In Strategy B, all layers of the network were unfrozen and trained on the target dataset using a smaller learning rate. This approach allows the model to fully adapt to the new task but significantly increases the risk of overfitting and instability.

6 Evaluation Metrics

Model performance was evaluated using the following metrics on the validation set.

Accuracy measures the overall proportion of correctly classified samples.

Macro precision, macro recall, and macro F1 score were used to account for class imbalance by giving equal importance to all classes.

Confusion matrices were also analyzed to better understand class specific performance.

7 Results and Analysis

Across both datasets, Strategy A consistently outperformed Strategy B, particularly in macro averaged metrics.

7.1 Feature Reuse Versus Feature Corruption

Pretrained models learn strong low level features such as edges, textures, and shapes from ImageNet. In both tasks, these features are directly relevant. Facial images share consistent geometric structure, while plant images share common texture and shape patterns.

Strategy A preserves these high quality features by freezing early layers. Strategy B allows these features to be updated using a smaller dataset, which can distort previously learned representations and reduce generalization.

7.2 Small Data Overfitting

Full fine tuning exposes millions of parameters to training. When applied to datasets that are small relative to ImageNet, this increased capacity allows the model to memorize dataset specific artifacts rather than learning general concepts.

In the DeepWeeds dataset, this can lead to reliance on background cues such as soil color or lighting. In facial expression recognition, it can lead to memorization of specific facial identities. Partial fine tuning restricts this behavior and encourages learning of meaningful features.

7.3 Catastrophic Forgetting

Full fine tuning is also vulnerable to catastrophic forgetting. Large gradient updates during early training stages can overwrite foundational knowledge in the network.

By restricting training to higher level layers, Strategy A protects the pretrained visual representations while still allowing adaptation to the target task.

8 Conclusion

This study demonstrates that partial fine tuning is often more effective than full fine tuning when working with medium sized datasets. Strategy A achieved better generalization by balancing stability and adaptation, preserving pretrained visual knowledge while avoiding overfitting and catastrophic forgetting.

These results highlight the importance of selecting fine tuning strategies based on dataset size and task complexity rather than assuming that full fine tuning will always yield superior performance.