Transfer Learning in Convolutional Neural Networks : Frozen vs Fine-Tuning

Student Name : Sreeja Nalla **Words** : 1990

Student ID : 24088662

GitHub Link : https://github.com/sreeja2817/Transfer-Learning-in-CNN-Frozen-vs-Fine-Tuning.git

A Practical Tutorial using MobileNetV2 on a Flower Classification Task

Abstract

Deep convolutional neural networks (CNNs) have revolutionised computer vision, but training them from scratch requires large datasets, specialised hardware and careful optimisation. Transfer learning provides an efficient alternative by re-using feature representations learned from large benchmark datasets such as ImageNet.

This tutorial investigates two common transfer-learning strategies in practice: **freezing** a pretrained network and **fine-tuning** its upper layers. A real-world flower classification task is used to compare a baseline CNN trained from scratch with MobileNetV2 used both as a fixed feature extractor and as a partially trainable model. Results demonstrate a large improvement from **67% accuracy** (baseline CNN) to **91% accuracy** (fine-tuned transfer learning).

The tutorial explains why transfer learning works, when fine-tuning becomes useful and how to safely apply these techniques in practical projects.

Introduction

Convolutional Neural Networks (CNNs) are the dominant model class for computer vision tasks including classification, object detection, segmentation and image synthesis. Their success lies in their hierarchical representation learning: early layers learn basic structures such as edges and textures while deeper layers capture object-level semantics.

Transfer learning solves this by reusing knowledge from previously trained models. A CNN trained on ImageNet, for example, has already learned general-purpose visual features from over 14 million images. Instead of relearning from scratch, transfer learning allows a new model to adapt those features to a new problem.

This tutorial explores two widely used strategies:

- **Frozen feature extractor**: takes advantage of the pretrained model's visual filters while training only a small classifier head.
- **Fine-tuning**: selectively retrains the upper layers of the pretrained model to improve accuracy.

Dataset

Flowers Recognition Dataset (Kaggle)

The Flowers Dataset used in this tutorial contains real photographs captured under a range of conditions including changes in lighting, angle, background complexity and object scale.

Dataset Statistics

- Total images: **4,317**
- Training samples: **3,454**
- Validation samples: **863**
- Image resolution used: **224 × 224**

Classes:

1. Daisy
2. Dandelion
3. Rose
4. Sunflower
5. Tulip

The dataset represents a moderately difficult classification task because:

- Some classes are visually similar (e.g., rose vs daisy).
- Lighting and backgrounds vary significantly.
- Objects appear at different scales.

This makes the dataset suitable for demonstrating differences between training from scratch and transfer learning.

Baseline CNN: Training from Scratch

To establish a reference point for performance, a convolutional neural network (CNN) was trained from scratch using only the flowers dataset. Unlike transfer learning models, this network had no prior knowledge of visual features and therefore had to learn everything from random initialisation.

Architecture Design

The baseline CNN follows a standard architecture pattern:

- Three convolutional blocks:
    - Conv2D (32 filters) → MaxPooling
    - Conv2D (64 filters) → MaxPooling
    - Conv2D (128 filters) → MaxPooling

- A fully connected head:
    - Flatten → Dense (128) → Dropout → Output layer (5 classes)

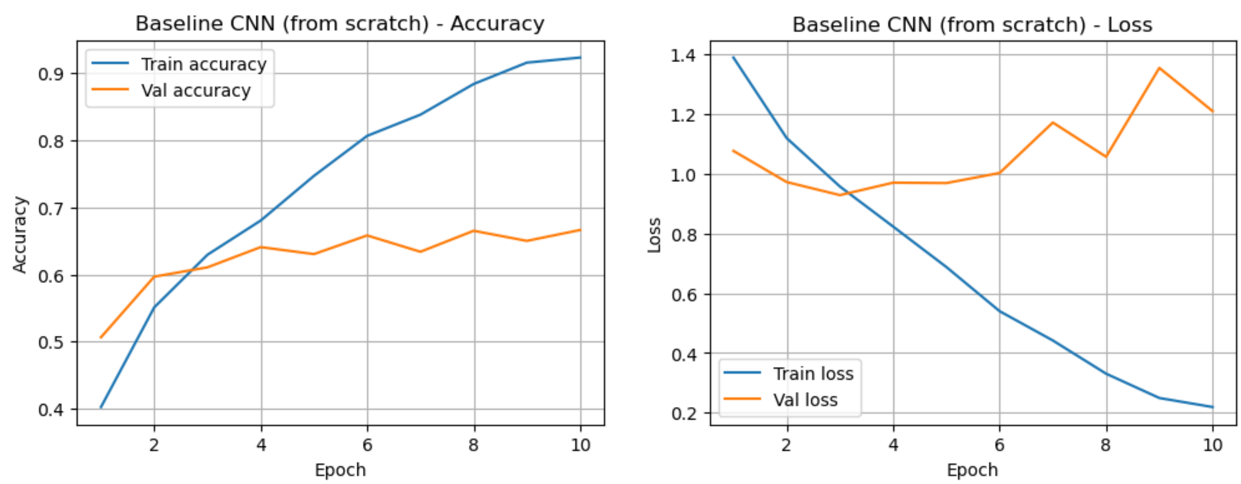
The total parameter count was approximately **11.17 million**, all of which were trainable.

Training Behaviour

The training process showed a rapid increase in training accuracy, while validation accuracy remained much lower. This indicates that the model was learning image-specific patterns from the training set but struggling to generalise.

In practice, this type of behaviour is typical when training deep networks from scratch using a limited dataset. Without exposure to enough visual diversity, the network fails to learn robust feature representations and instead memorises training examples.

Figure 1: Training and Validation Accuracy/Loss for Baseline CNN



Baseline Results and Analysis

Quantitative Evaluation

The baseline model achieved an **overall accuracy of 67%**. Performance varied significantly between classes:

- Sunflower achieved the highest recall.
- Rose and tulip exhibited the poorest performance.
- Daisy and dandelion were often confused.

Diagnostic Insight from Confusion Matrix

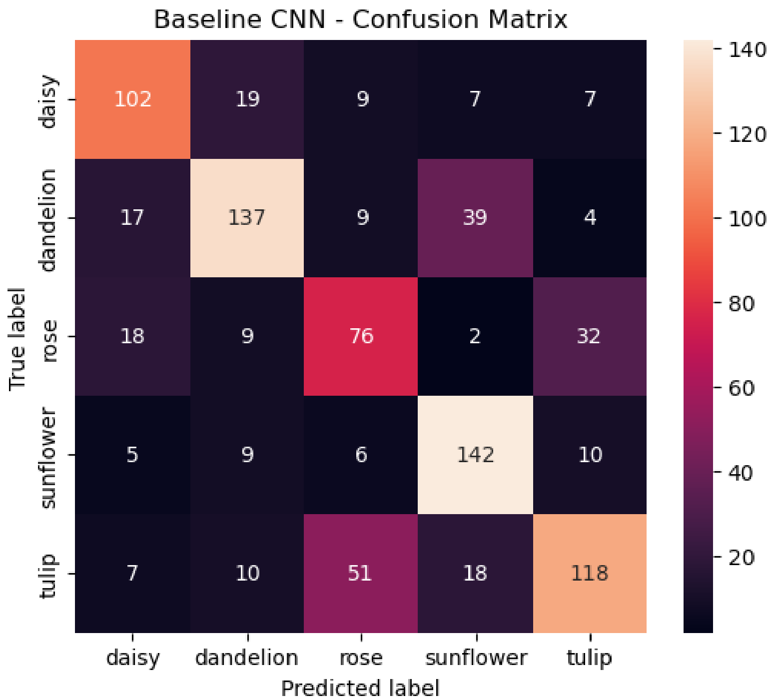
The confusion matrix provides a deeper understanding of model behaviour. It reveals which classes are confused and highlights limitations caused by insufficient feature learning.

Frequent misclassifications include:

- Daisy predicted as rose
- Tulip predicted as rose
- Dandelion confused with sunflower

These errors suggest that the baseline CNN struggles to distinguish fine-grained petal structure and colour similarity.

Figure 2: Confusion Matrix for Baseline CNN



Interpretation

The baseline CNN experiences classic overfitting:

- Training loss decreases steadily
- Validation loss increases after several epochs
- Validation accuracy plateaus

This demonstrates that training deep networks from scratch is inefficient when dataset size is limited.

Transfer Learning with MobileNetV2 (Frozen Backbone)

MobileNetV2 is a pretrained CNN trained on the ImageNet dataset. Instead of learning edges and textures from scratch, it provides feature maps that have already been learned through exposure to millions of images.

For this experiment, the convolutional backbone was **frozen** and only the classifier head was trained.

Architecture

The frozen model consists of:

- MobileNetV2 (frozen)
- GlobalAveragePooling
- Dense(128)
- Dropout
- Dense(5 classes)

Transfer Learning Results (Frozen Backbone)

Training Behaviour

The learning curves for the frozen MobileNetV2 model show stable training with both training and validation loss decreasing smoothly across epochs. Importantly, there is no early divergence between the two curves, indicating that the model is not overfitting. Unlike the baseline CNN, the frozen model demonstrates steady improvement on both training and validation data, confirming stronger generalisation.

The aligned loss curves suggest that the network is learning meaningful features rather than memorising the data. MobileNetV2 backbone model is able to converge quickly and achieve good performance with fewer training epochs and less data.

Accuracy Comparison

The frozen transfer learning model achieves approximately 88% validation accuracy, a significant improvement over the baseline CNN’s 67%. This increase is achieved despite having far fewer trainable parameters. While the baseline model had over 11 million trainable parameters, the transfer learning model required only around 160,000.

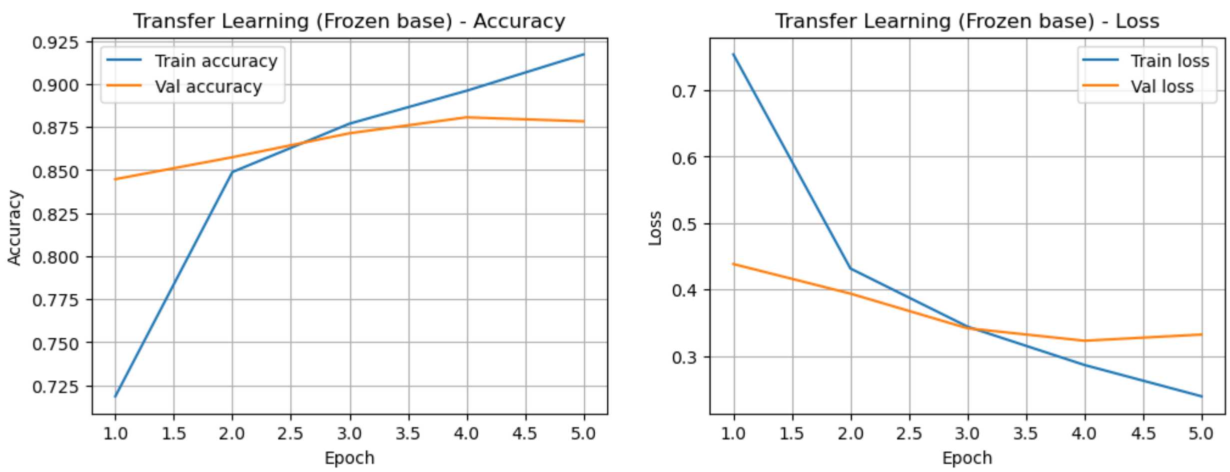
This shows that high-quality pretrained features are more valuable than simply increasing model size. Instead of learning edges and textures from scratch, the network begins with an advanced understanding of visual patterns and focuses solely on learning how to classify flowers.

Interpretation

The frozen model performs well because the pretrained CNN already contains a hierarchy of reusable visual features. Lower layers detect edges and corners, intermediate layers respond to shapes and textures, and deeper layers encode abstract visual representations.

The new classifier head learns how to map these features to flower categories without relearning image structure from the beginning. This results in faster convergence, improved accuracy and reduced overfitting, even with a relatively small dataset.

Figure 3: Learning Curves for Frozen MobileNetV2 (Training and Validation Accuracy/Loss)



Transfer Learning with Fine-Tuning (MobileNetV2)

While frozen transfer learning provides strong performance, the pretrained model was originally trained on general object categories such as animals, vehicles and household objects rather than flowers specifically. Consequently, its representations may not be fully optimised for the subtle variations found in plant structures such as petal shape, colour gradients and texture patterns. To address this, fine-tuning was applied to adapt the model more closely to the flower classification task.

Fine-Tuning Strategy

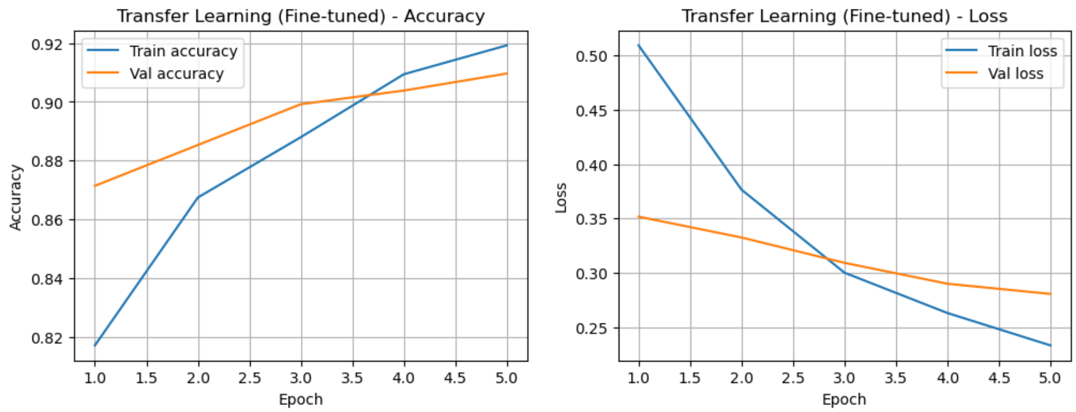
The fine-tuning approach used in this experiment involved unfreezing the **last six layers** of MobileNetV2 while keeping the remaining layers frozen. A reduced learning rate was applied to ensure that previously learned features were not overwritten too aggressively, a phenomenon commonly referred to as catastrophic forgetting. Only the highest layers were updated so that general visual knowledge could be preserved while allowing the model to refine high-level features.

Why Fine-Tuning Helps

Fine-tuning improves performance by enabling domain-specific learning. The model begins to adjust its filters according to the visual characteristics unique to flowers, including petal structure, symmetry, colour distribution and background patterns. These refinements improve class separation and reduce confusion between visually similar species such as roses and daisies.

Decision boundaries become sharper and misclassifications decline as the network becomes increasingly specialised. Fine-tuning also improves calibration, meaning that the model becomes more confident in its correct predictions and less overconfident in incorrect ones. This results in a noticeable boost in accuracy and robustness.

Figure 4: Learning Curves for Fine-Tuned MobileNetV2 (Training and Validation Accuracy/Loss)



Model Performance Comparison

Overall Accuracy Comparison

A clear performance hierarchy emerges when comparing the three models:

- Baseline CNN: **67%**
- Frozen Transfer Learning: **88%**
- Fine-Tuned Transfer Learning: **91%**

This demonstrates that transfer learning yields dramatic improvements over training from scratch and that fine-tuning provides additional gains when sufficient data is available.

Figure 5: Confusion Matrix – Fine-Tuned MobileNetV2

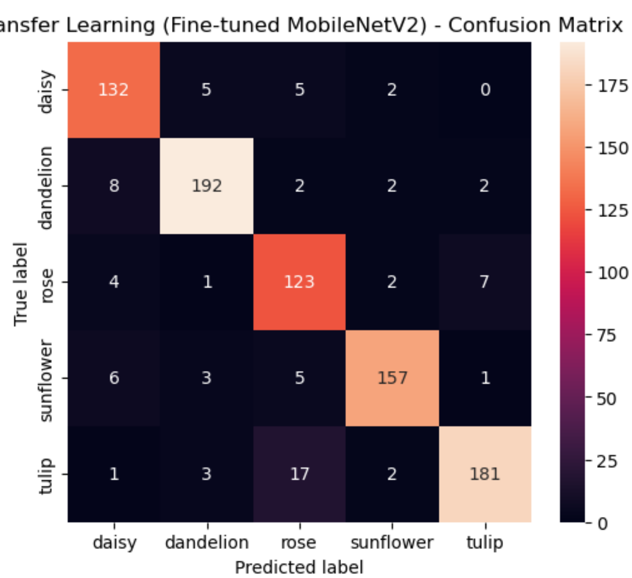


Figure 6: Baseline Model Predictions (Predicted vs Ground Truth)

Several incorrect predictions are visible, particularly between visually similar flowers such as roses and daisies.

Figure 7: Frozen Model Predictions (Predicted vs Ground Truth)

The frozen transfer learning model shows consistent correct predictions across different flower types.

Class-Level Improvements

Fine-tuning resulted in significant class-level improvements across all categories. The largest performance gains were observed for **rose and tulip**, which were previously confused by the baseline model. With fine-tuning, the model learned finer visual distinctions such as petal curvature and texture patterns that were previously ambiguous.

Dandelion and sunflower achieved near-perfect recognition, indicating that prominent visual cues such as large petal shapes and high-contrast colour regions were successfully encoded by the network. Daisy classification also improved noticeably with fewer misclassifications as rose.

Qualitative Analysis

Visual comparisons between baseline and fine-tuned predictions provide strong qualitative evidence of improvement. The baseline model frequently misclassifies visually similar flowers, especially under difficult lighting conditions or cluttered backgrounds. In contrast, the fine-tuned model produces confident and accurate predictions across a wide variety of conditions.

Most residual errors occur due to:

- extreme lighting variations
- blurred foreground objects
- unusual viewing angles
- partially occluded objects

Despite these challenges, the fine-tuned model handles the majority of real-world variability effectively.

Conclusion

This study demonstrates the significant advantages of transfer learning when applying deep convolutional neural networks to image classification tasks with limited data. A baseline CNN trained from scratch achieved only moderate performance, largely due to its inability to learn robust visual features from a comparatively small dataset. In contrast, the use of a pretrained MobileNetV2 network dramatically improved accuracy, stability and generalisation.

Freezing the pretrained backbone allowed the model to reuse rich visual representations without the computational cost of retraining the entire network, achieving a large improvement over the baseline. Fine-tuning further enhanced performance by allowing the highest-level features to adapt to the domain of flower classification, resulting in the best overall accuracy and significantly reduced misclassification.

The results confirm that model performance is strongly influenced by feature quality rather than model size alone. Despite having fewer trainable parameters, the transfer learning models consistently outperformed the baseline CNN because they benefited from knowledge transferred from large-scale image training.

From a practical perspective, this work highlights transfer learning as the most effective approach for real-world computer vision tasks where data is limited and computational resources are constrained. Rather than training deep neural networks from scratch, practitioners should adopt pretrained models and fine-tuning strategies to achieve higher accuracy, faster convergence and better generalisation.

This tutorial also demonstrates the importance of evaluating models using multiple perspectives. Accuracy alone does not tell the full story; confusion matrices and qualitative predictions provide valuable insights into how and why models succeed or fail. Combining these analyses offers a deeper understanding of neural network performance and model behaviour.

In summary, fine-tuned transfer learning provides a powerful and practical solution for image classification problems, significantly outperforming traditional CNN training approaches. The techniques demonstrated in this tutorial can be extended to a wide range of computer vision applications where reliability and efficiency are critical.

References

- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. https://arxiv.org/abs/1704.04861
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. https://arxiv.org/abs/1801.04381

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks*. https://papers.nips.cc/paper_files/paper/2012/
- Pan, S. J., & Yang, Q. (2010). *A Survey on Transfer Learning*. IEEE Transactions on Knowledge and Data Engineering, https://ieeexplore.ieee.org/document
- TensorFlow (2024). *Transfer learning and fine-tuning tutorials*. Available at: https://www.tensorflow.org/tutorials/images/transfer_learning
- Kaggle (2024). *Flowers Recognition Dataset*. https://www.kaggle.com/datasets/alxmamaev/flowers-recognition