

1. Design Choices for the Architecture

a) Full Model Fine-Tuning

- Fine-tune all layers of the pre-trained model.

b) Freeze the Pre-Trained Model and Add Custom Classification Head

- Keep the pre-trained model layers frozen and add a new classification head (usually a few fully connected layers) at the end.

c) Partial Fine-Tuning (Selective Layer Fine-Tuning)

- Freeze the initial layers and fine-tune only the deeper layers of the pre-trained model.

2. Merits and Demerits of the Design Choices

a) Full Model Fine-Tuning

Merits:

- Full adaptability to the new dataset, allowing the model to learn task-specific features from all layers.
- Can lead to higher accuracy if the new dataset is significantly different from the original pre-training dataset.

Demerits:

- Computationally expensive and time-consuming due to backpropagation across the entire model.
- Requires a large amount of labeled data to avoid overfitting, especially in deep architectures.
- More sensitive to hyperparameters like learning rate.

b) Freeze Pre-Trained Model and Add Custom Classification Head

Merits:

- Faster and less computationally expensive since only the new layers are trained.
- Reduced risk of overfitting, especially on small datasets.
- Useful when the new dataset is similar to the one used for pre-training, making the pre-trained features highly transferable.

Demerits:

- The model might not learn task-specific features adequately, especially if the new dataset differs significantly from the pre-training dataset.
- Limited flexibility, as frozen layers may not adapt well to domain-specific patterns.

c) Partial Fine-Tuning

Merits:

- Balances computational efficiency and model adaptability.
- Only adjusts deeper layers, which are more task-specific, while retaining generic features learned by earlier layers.
- Suitable for datasets that are somewhat similar to the original pre-training dataset but still require specialized learning.

Demerits:

- Requires careful selection of layers to freeze and fine-tune, which may demand more experimentation.
- Can still overfit if the dataset is too small or dissimilar without careful tuning.

3. Fine-Tuning Shallower Layers vs. Deeper Layers

Deeper layers should generally be fine-tuned in an image classification problem, rather than the shallower layers.

Shallower layers in pre-trained models capture low-level features like edges, textures, and basic shapes, which are usually universal across different datasets. These layers are generally applicable across various tasks and don't need major adjustments.

Deeper layers, on the other hand, capture more task-specific, abstract features, such as object parts and complex patterns. These are more likely to differ between tasks, especially if your new dataset is significantly different from the one used for pre-training (e.g., different objects, styles, or domains).

Fine-tuning the deeper layers allows the model to adjust its high-level feature extraction to be more aligned with the new dataset, improving performance on the target task. Freezing the shallower layers ensures that the model retains robust, general feature extraction without unnecessary re-training, making the process more efficient and reducing overfitting risks.