1. Design Choices for Architecture

1.1 VGG-16

VGG-16 is a convolutional neural network recognized for its straightforward design and efficiency in image classification tasks. The architecture consists of 16 layers, comprising 13 convolutional layers and 3 fully connected layers.

1.2 ResNet50

ResNet50 employs a deep residual architecture that facilitates the efficient training of very deep networks by utilizing skip connections. This approach effectively mitigates the vanishing gradient problem that is often encountered in deep neural networks.

1.3 EfficientNet

EfficientNet uniquely balances depth, width, and image resolution to enhance the model's performance while minimizing computational costs. It excels in feature extraction and frequently achieves state-of-the-art results with a reduced number of parameters.

2. Merits and Demerits of Design Choices

2.1 VGG-16

Merits:

- Achieves high accuracy on various computer vision tasks, including image classification (92.7% top-5 accuracy on ImageNet)4.
- Straightforward architecture makes it easy to understand and implement4.

Demerits:

- Requires significant computational resources and time for training due to its depth4.
- Tends to be slow to train; original models were reported to take 2-3 weeks on powerful GPUs4.

2.2 ResNet50

Merits:

- Introduces skip connections that facilitate training deeper networks, avoiding the vanishing gradient problem6.
- Proven to perform exceedingly well in various applications, including object detection and emotion recognition6.

Demerits:

- High computational and memory demands due to its deep architecture6.
- Training can be time-consuming and complex, often taking weeks to complete6.

2.3 EfficientNet

Merits:

- Achieves a good balance between accuracy and computational efficiency, with reduced model sizes and faster performance, achieving state-of-the-art results7.
- Adaptable for deployment in low-resource environments while maintaining performance7.

Demerits:

- While efficient, it may struggle with the vanishing gradient problem similar to other deep architectures7.
- More complex scaling can complicate model tuning and optimization7.

3. Fine-tuning Deeper Layers vs. Shallower Layers

Fine-tuning deeper layers of a pre-trained convolutional neural network (CNN) is typically more advantageous for image classification tasks. By fine-tuning the higher-level layers, the model can learn and adapt complex, task-specific features that are well-aligned with the target dataset. Deep layers in CNNs are capable of capturing intricate representations, enabling the model to effectively adjust to variations in the data.In contrast, while fine-tuning shallower layers can be beneficial, it may limit the model's ability to learn the specific nuances relevant to the target task. This can potentially lead to suboptimal performance, as the model may not be able to fully capture the unique characteristics of the new dataset.Furthermore, research has shown that fine-tuning all layers or skipping just the first one or two layers often yields comparable performance levels. However, focusing on the deeper layers frequently provides superior enhancements in accuracy, particularly for complex image classification tasks. This is because the deeper layers have learned more specific and relevant features that can be fine-tuned to the target problem, leading to better overall performance.