110000169 劉國揚
電機資訊學院學士班

National Tsing Hua University
Fall 2025 11410IPT 553000
Deep Learning in Biomedical Optical Imaging
Homework 4 Analysis Report

1.1 Task A: Model Selection (20%)

The models I will select for transfer learning tasks are ResNet-50 and MobileNetV3 (Large).

Weight	Acc@1	Acc@5	Params	GFLOPS	Recipe
ResNet50_Weights.IMAGENET1K_V1	76.13	92.862	25.6M	4.09	link
ResNet50_Weights.IMAGENET1K_V2	80.858	95.434	25.6M	4.09	link
MobileNet_V3_Large_Weights.IMAGENET1K_V1	74.042	91.34	5.5M	0.22	link
MobileNet_V3_Large_Weights.IMAGENET1K_V2	75.274	92.566	5.5M	0.22	link

Fig.1.1 Model comparison table from *torchvision.models*

Reasons for selecting ResNet-50:

ResNet-50 uses “Shortcut Connections” that allows effective training of networks with more than 100 layers to overcome the vanishing gradient problem, and its 50 layers makes a medium-to-high complexity model. Besides, it achieves 3.57% error on the ImageNet and becomes a classic architecture with performance which surpasses all previous architectures such as VGG and GoogLeNet. Compared to deeper networks like ResNet-101 or resNet-152, ResNet-50 has lower computational cost and become suitable for transfer learning especially for fine-tuning on GPUs.

Residual connections also help to learn robust features which generalize well across different tasks, showing strong potential for transfer learning. The last reason supports my choice is that ResNet model is widely used across tasks like classification, detection and segmentation and this makes it easy to find the fine-tuning examples and optimization methods. (relatively more resources).

Reasons for selecting MobileNetV3:

MobileNetV3 employs Depthwise Separable Convolutions with Squeeze-and-Excitation (SE) blocks for higher efficiency and low computational cost. It has also fewer parameters than ResNet-50 while maintaining good performance. On ImageNet, it achieves performance similar to ResNet-50 but with significantly lower computational cost. Therefore, MobileNetV3 model is suitable for resource-limited environments (edge devices or GPUs for etc) and allowing fast fine-tuning.

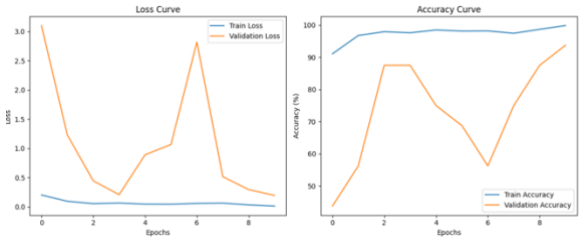
MobileNetV3 provides other several benefits such as efficiency, flexibility and transferability. It is suitable for practical deployment scenarios (mobile or embedded systems) and available in both large and small versions which can be chosen based on hardware requirements. The paper also states that it is effective adaptation to other tasks on mobile hardware and this demonstrate its flexibility for low-resource transfer learning.

From the Comparison Table provided in *torchvision.models*, we might have a better model which is outperform across all metrics, but I would still select ResNet-50 and MobileNetV3 models by considering the balance between the performance and efficiency, transfer learning feasibility and resource constraints.

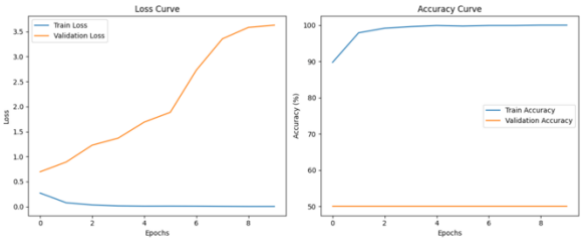
1.2 Task B: Fine-tuning the ConvNet (30%)

Fine-tune the selected pre-trained models on our 2-class x-ray dataset. They need to modify the models to fit the new dataset and analyze the effectiveness of fine-tuning, considering the architecture and depth of the networks.

ı **Discussion (30%, 15% for each)**: Analyze the performances of the fine-tuned models. Include a comparative evaluation, focusing on how effectively fine-tuning facilitated their adaptation to the new task.




```
Model weights loaded successfully.
Test Accuracy: 82.50%
Test Loss: 0.7611
```




```
Model weights loaded successfully.
Test Accuracy: 63.44%
Test Loss: 0.6627
```

Fig.1.2 Results with fine-tuning ResNet-50 and MobileNetV3

The fine-tuned ResNet-50 model achieved a test accuracy of 82.50% and a test loss of 0.7611, indicating reasonably strong performance on the two-class X-ray classification task. The training curves show that the model's training loss steadily decreased and stabilized at a low value, while the validation loss exhibited noticeable fluctuations across epochs. This suggests that the model learned the training data effectively but its generalization capability to unseen data was inconsistent. The accuracy plot also demonstrates that training accuracy quickly approached nearly the ceiling whereas validation accuracy oscillated significantly, implying mild overfitting. These observations are typical for deep networks like ResNet-50, which possess many parameters and can easily overfit smaller or domain-specific datasets if not regularized properly.

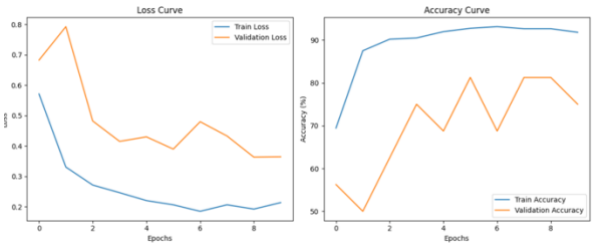
The fine-tuned MobileNetV3 model achieved lower test accuracy of 63.44% with a test loss of 0.6627, demonstrating weaker generalization compared to ResNet-50. The learning curves indicate that the training loss consistently decreased and training accuracy rapidly reached near-perfect levels, but the validation accuracy remained almost constant around 50%, and the validation loss continued to increase throughout training. This pattern clearly reflects overfitting, that is, the model memorized the training data but failed to adapt to new samples. Its lightweight and compact architecture is advantageous for efficiency but also provides limited representational capacity and is less suited to large domain shifts such as from natural to medical imagery. These results indicate that MobileNetV3 struggled to benefit effectively from transfer learning under the current setup.

The observed results differed from my expectations, as fine-tuning pre-trained models such as ResNet-50 and MobileNetV3 is generally anticipated to yield high performance on binary classification tasks. For this situation, I would suggest that it is caused by the domain difference between natural images (on which these models were originally trained, e.g., ImageNet) and medical X-ray images. X-rays possess distinct visual characteristics that pre-trained filters may not readily capture without extensive adaptation. Secondly, dataset limitations can lead to unstable validation results and hinder the models from learning representative features. Inappropriate learning rates, inadequate layer freezing strategies, or insufficient regularization can also cause overfitting or prevent effective convergence.

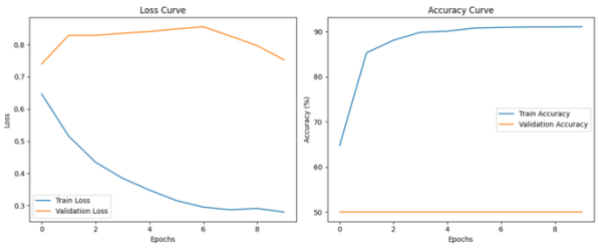
1.3 Task C: ConvNet as Fixed Feature Extractor (30%)

You will transform the selected models into fixed feature extractors by freezing all layers except the final one and evaluate their performance.

ı **Discussion (30%, 15% for each)**: Similar to Task B, provides a comprehensive analysis of the models' performances. The focus should again be on a comparative evaluation of their effectiveness in the new task, though this time as fixed feature extractors.




```
Model weights loaded successfully.
Test Accuracy: 85.47%
Test Loss: 0.3375
```




```
Model weights loaded successfully.
Test Accuracy: 63.44%
Test Loss: 0.6812
```

Fig.1.3 Results with fine-tuning ResNet-50 and MobileNetV3 at Fixed Feature Extractors

When used as a fixed feature extractor, the ResNet-50 model demonstrated strong performance and achieved a test accuracy of 85.47% and a test loss of 0.3375. The loss curve showed a consistent downward trend for both training and validation, indicating effective learning and stable convergence. Although the validation accuracy fluctuated slightly across epochs, it generally remained high, suggesting that the features extracted from the pretrained ResNet-50 layers were highly transferable to the new classification task. This outcome highlights the robustness of ResNet-50's deep residual architecture, which can capture rich and discriminative representations even when most of the network layers remain frozen. This observation also implies that ResNet-50's pretrained features generalize well and can serve as an effective foundation for downstream tasks with limited fine-tuning.

MobileNetV3 model achieved a test accuracy of 63.44% and a test loss of 0.6812, showing considerably weaker performance compared to ResNet-50 under the same fixed feature extraction setup. The training loss decreased steadily, but the validation loss remained high, and the validation accuracy remained around 50%, indicating poor generalization. This pattern suggests that although the model could adapt to the training data, the frozen MobileNetV3 features were not sufficiently expressive for the target dataset. Since MobileNetV3 is a lightweight and efficiency-oriented architecture with fewer parameters and less depth, its learned representations tend to be less transferable in complex or high-variance tasks when not fine-tuned end-to-end. Its limited performance implies the trade-off between model efficiency and representational richness in transfer learning scenarios.

1.4 Task D: Comparison and Analysis (10%)

By completing Task B and C, contrast the performance outcomes and adaptability of the models when subjected to the two distinct transfer learning approaches.

l **Discussion (10%)**: Offer a succinct analysis that highlights the differences in performance and adaptability observed when the models are fine-tuned versus when used as fixed feature extractors.

The results demonstrate that the choice of transfer learning strategy had a significant impact on model performance and generalization. By the experimentation, we noticed that the fixed feature extraction produced more stable learning behaviour and higher test accuracy compared to fine-tuning, especially for deeper architectures like ResNet-50. This suggests that freezing pretrained layers can better preserve learned representations and improve model robustness when working with limited or domain-specific datasets.

For ResNet-50, applying it as a fixed feature extractor yielded superior results, achieving 85.47% accuracy with a test loss of 0.3375, compared to 82.50% accuracy and 0.7611 loss under fine-tuning. The frozen configuration helped prevent overfitting and leveraged the rich, pretrained ImageNet features effectively then resulted in consistent convergence and high validation performance. Fine-tuning all layers caused noticeable fluctuations in validation accuracy and loss, likely due to over-adjusting the pretrained weights on a relatively small X-ray dataset.

For MobileNetV3, the model consistently underperformed in both configurations, with test accuracy around 63%. Fine-tuning led to clear overfitting, where training accuracy approached the ceiling, but validation accuracy remained near 50%. Similarly, the fixed feature extraction setup failed to achieve substantial improvement, indicating that MobileNetV3's compact and efficiency-oriented architecture lacked sufficient representational depth for the complex visual features present in x-ray task.

We can conclude that ResNet-50 demonstrated stronger adaptability and feature transferability, while MobileNetV3's lightweight design limited its generalization in both approaches. The findings highlight that that fixed feature extraction can offer a more stable and effective transfer learning strategy for small, domain-shifted datasets such as medical X-rays.

1.5 Task E: Test Dataset Analysis (10%)

In the original Lab 4's code, you may have encountered challenges in enhancing the performance on the test dataset.

l **Discussion (10%)**: Elucidate your perspective on this phenomenon. Provide an analysis explaining the reasons behind the difficulty in improving the test dataset performance.

As I mentioned my observation at *Section 1.2*, the output results are really differed from my expectation which selected model can provides significant helps on performance enhancement. The difficulty in improving the test dataset performance primarily stems from domain mismatch, data limitations, and model optimization challenges.

We know that there is a significant domain difference between the ImageNet dataset used for pretraining and the X-ray images used in this task. ImageNet contains natural colour images with diverse objects, while X-rays are grayscale and have entirely different texture and structural patterns. As a result, the pretrained filters in models like ResNet-50 and MobileNetV3 may not transfer well to the medical domain. Without sufficient fine-tuning or adaptation, the models struggle to extract meaningful features from the X-ray data, leading to suboptimal generalization.

The dataset size and diversity also play a crucial role. When the training dataset is relatively small or imbalanced, the model may be overfit to specific patterns in the training samples and fail to generalize to unseen data. This is evident from the fluctuating validation accuracy and rising validation loss during training (Refer to Fig.1.2).

Besides, training hyperparameters and strategies also influence model's performance. Improper learning rates or incomplete freezing/unfreezing of layers can interrupt effective fine-tuning. Insufficient regularization techniques (dropout, weight decay, or data augmentation) allow the model to memorize training data rather than learn robust and generalizable representations.

Model architecture differences might be a factor as well. ResNet-50, being deeper, captures complex hierarchical features but risks overfitting more easily on small datasets. MobileNetV3, which is efficient but has limited representational capacity, making it less effective for complex medical image patterns.

In this assignment, we learnt that the challenge in improving test performance is largely due to the combination of domain gap, limited data, overfitting tendencies, and suboptimal training configurations, all of which restrict the models' ability to generalize effectively to unseen X-ray images.

References:

1) Deep Residual Learning for Image Recognition (ResNet-50 Model)
https://arxiv.org/pdf/1512.03385
2) Searching for MobileNetV3 (MobileNetV3 Model)
https://arxiv.org/pdf/1905.02244
3) https://www.geeksforgeeks.org/machine-learning/ml-introduction-to-transfer-learning/
4) https://www.sciencedirect.com/science/article/pii/S0957417423033092