National Tsing Hua University
Deep Learning in Biomedical Optical Imaging
Homework 4

Task A: Model Selection

In this task, I chose ResNet50 and VGG16 as pre-trained models for transfer learning. Both models feature well-designed architectures and have high accuracy from pre-training on the ImageNet dataset, making them suitable for transfer learning on new tasks. Below is a detailed explanation for each model.

Model Selection

- Chosen Models: ResNet50 and VGG16.
- Reason: These two models are common convolutional neural network (CNN) architectures, pre-trained on large datasets, widely used, and effective for transfer learning applications. ◦

Model Explanation

ResNet50

- **Architecture Complexity:** ResNet50 uses a 50-layer deep residual network. Its residual block structure addresses the vanishing gradient problem common in deep networks, allowing the network to maintain depth while effectively training. Residual blocks enable information to flow directly from earlier to later layers, making it more flexible for learning higher-level features.
- **Pre-training Performance:** Pre-trained on ImageNet with high accuracy, ResNet50 excels in extracting complex features from images, making it suitable for fine-grained classification tasks.
- **Inference Time:** In experiments, ResNet50 achieved an inference time of approximately 578.4 milliseconds, demonstrating a good balance between computational efficiency and accuracy. This makes it well-suited for transfer learning tasks requiring precise feature extraction.

VGG16

- **Architecture Complexity:** VGG16 is a 16-layer convolutional network using a consistent 3x3 filter size across all convolutional layers. This simple, modular structure is easy to understand and modify, but requires higher computational resources due to its number of layers and filters. ◦
- **Pre-training Performance:** VGG16 is also pre-trained on ImageNet, achieving strong classification performance. Its straightforward, effective architecture ensures stable

performance in image classification tasks, making it ideal for transfer learning applications where computational resources may be limited and frequent adjustments are unnecessary.

- **Inference Time:** VGG16's inference time was approximately 616.8 milliseconds, slightly longer than ResNet50. While stable and easy to adapt, it may be less efficient for tasks requiring high computational performance.

Conclusion

- **ResNet50** is better suited for high-precision tasks requiring fine-grained classification. Its deep residual network and efficient inference time allow it to adapt flexibly to new tasks in transfer learning.
- **VGG16** has a slightly longer inference time but offers a simple and stable design, making it ideal for applications needing consistent performance and shorter computational times. Its reliable feature extraction performance is especially beneficial for resource-limited environments.

Both models offer distinct advantages. Depending on the feature requirements and computational constraints of a new task, either model can be chosen to achieve optimal transfer learning performance.

Task B: Fine-tuning the ConvNet

In this task, we fine-tuned the pre-trained ResNet50 and VGG16 models to adapt them to a simulated binary X-ray dataset. By adjusting the output layer of each model to meet binary classification requirements, we trained both models for 5 epochs to evaluate their performance on this new task.。

Fine-tuning Process and Loss Performance

1. ResNet50 Fine-tuning：

- Over 5 training epochs, the loss for ResNet50 steadily decreased, from 0.7468 in the first epoch to 0.6972 in the fifth epoch, indicating a stable convergence trend.
- The model achieved a test accuracy of 54.5%, showing some classification capability, though the error rate remained relatively high.

2. VGG16 Fine-tuning：

- Starting with a loss of 1.5614, VGG16’s loss gradually decreased to 0.7078, but its convergence rate was slower and its final loss slightly higher than that of ResNet50.
- The test accuracy for VGG16 was 46.0%, significantly lower than ResNet50, demonstrating weaker adaptability in this binary classification task.

Performance Comparison and Analysis

Impact of Model Architecture:

- **ResNet50:** The use of residual blocks suits a deeper network structure, allowing it to retain learned features better and capture subtle feature variations during training. This capability enables ResNet50 to adapt more effectively to the binary classification task, resulting in higher test accuracy compared to VGG16.
- **VGG16:** With a shallower architecture and fewer convolutional layers, VGG16 performs well in simpler classification tasks. However, its relatively limited depth and feature extraction capacity are inadequate for tasks requiring fine-grained feature extraction, which restricts its adaptability to this binary classification task.

Convergence During Training:

ResNet50 demonstrated stable convergence across the 5 epochs, with a strong capacity to adapt and a consistent decrease in loss. In contrast, VGG16 converged more slowly, with greater fluctuations in loss, indicating more difficulty in learning new features for this task.

Test Performance Differences:

The test accuracy of ResNet50 was significantly higher than that of VGG16, highlighting the flexibility of deep residual networks in capturing and learning new task features. VGG16's shallow architecture, however, struggled to effectively adapt to new data distributions in the short term.

Conclusion

- **ResNet50** is better suited for this binary X-ray image classification task. Its depth and residual structure enhance its ability to adapt to new features, achieving superior performance after fine-tuning compared to VGG16.
- **VGG16**, though effective in basic classification tasks, falls short in tasks requiring robust feature extraction, with a fine-tuning performance that does not match that of ResNet50. ◦

For further improvement in accuracy, we could consider increasing the number of training epochs or applying data augmentation techniques to improve VGG16's adaptability and enhance the generalization performance of ResNet50.

Task C: ConvNet as Fixed Feature Extractor

In this task, I used the ResNet50 and VGG16 models as fixed feature extractors by freezing all convolutional layers and training only the final linear layer to adapt to the binary X-ray dataset. This approach aims to leverage pre-trained feature extraction capabilities and evaluate transfer learning performance without full model retraining.

Training Process and Loss Performance

ResNet50 as a Fixed Feature Extractor:

- Over 5 training epochs, the loss for ResNet50 gradually decreased from 0.7388 in the first epoch to 0.6350 in the fifth epoch, indicating a convergence trend.
- The test accuracy reached 50.0%, which fell short of expected classification performance, suggesting limited adaptability of ResNet50 in this fixed feature extraction setting.

VGG16 as a Fixed Feature Extractor:

- VGG16's loss also decreased across 5 epochs, from 0.7757 to 0.7481, though the loss fluctuations indicated a slower and less stable convergence.
- The test accuracy for VGG16 was 54.0%, slightly higher than ResNet50, suggesting a slight advantage for VGG16 as a fixed feature extractor in this binary classification task.

Performance Comparison and Analysis

Impact of Model Architecture:

- **ResNet50:** The deep residual network of ResNet50 is advantageous for deeper feature learning; however, as a fixed feature extractor, its deep structure might not be fully utilized, relying solely on the pre-trained features for classification, resulting in suboptimal performance.
- **VGG16:** The simpler convolutional structure of VGG16 proved more stable in the fixed feature extraction task, making it more suitable for direct feature classification and reflecting a stable feature extraction capability.

Convergence During Training:

- **ResNet50** showed a stable decline in training loss, indicating good convergence, but its test accuracy remained low, suggesting limited adaptability in the fixed feature extraction mode.
- **VGG16**, while exhibiting greater loss fluctuations, achieved slightly higher test accuracy, indicating slightly better adaptability in fixed feature extraction tasks.

Test Performance Differences:

The test accuracy for VGG16 in fixed feature extraction mode was marginally higher than ResNet50, suggesting it may be more suited to new tasks in this mode, while the deeper residual features of ResNet50 were not as advantageous in this application.

Conclusion

- **VGG16** showed a slight advantage as a fixed feature extractor for this binary classification task, benefiting from stable feature extraction and adapting relatively well to the new task without altering its underlying features.
- **ResNet50** demonstrated weaker adaptability in the fixed feature extraction mode; its deeper feature structure did not fully utilize its potential when only the final layer was trained. 。

Both models show different strengths in fixed feature extraction tasks: if stable classification performance is required and computational resources are limited, VGG16 offers practicality. However, if a new task requires deeper feature learning, it is recommended to adopt a fine-tuning strategy to fully leverage ResNet50's advantages.

Task D: Comparison and Analysis

In this task, we compared the performance and adaptability of the ResNet50 model using two transfer learning methods: fine-tuning and fixed feature extraction on a binary X-ray image dataset.

Model Performance Comparison

1. **Fine-tuning Mode:**

- **Training Process:** All layers were updated, with training loss quickly decreasing from 0.7954 to 0.2963 within five epochs, showing good convergence.
- **Test Accuracy:** Achieved 93.0%, indicating strong adaptability and accurate feature learning for classification.

2. **Fixed Feature Extraction Mode:**

- **Training Process:** Only the final layer was updated, resulting in slower convergence with loss decreasing from 0.7516 to 0.6634.
- **Test Accuracy:** Achieved 55.5%, showing limited ability to learn fine details, impacting classification performance.

Adaptability Analysis

- **Fine-tuning Mode:** Updating all layers allows the model to adapt effectively to new tasks, capturing subtle features in X-ray images, making it highly adaptable.
- **Fixed Feature Extraction Mode:** With frozen convolutional layers, adaptability is limited as the model relies on pre-trained features, making it less suitable for tasks requiring new data adaptation.

Conclusion

- **Fine-tuning Mode:** Outperforms fixed feature extraction by enabling layer flexibility, crucial for tasks requiring detailed feature detection, achieving 93.0% test accuracy.
- **Fixed Feature Extraction Mode:** Suitable for basic classification tasks or resource-limited settings but lacks adaptability for complex tasks, achieving only 55.5% accuracy.

Overall, fine-tuning demonstrated superior adaptability and accuracy, confirming its suitability for transfer learning on new, feature-rich tasks.

Task E: Test Dataset Analysis

In this experiment, I trained the ResNet50 model and evaluated its performance on both the training and test datasets. The results showed a training accuracy of 85.5% and a slightly lower test accuracy of 85.0%. Although these accuracies are similar, certain factors may still be affecting test set performance. Below is an analysis of potential factors limiting test performance.

Analysis of Factors Limiting Test Set Performance

1. Overfitting:

The model may have over-learned details in the training data, resulting in high training accuracy but weaker generalization on the test set. Overfitting is especially likely when the dataset is small or has high feature redundancy, leading the model to memorize training features rather than learning patterns generalizable to broader data.

2. Differences in Data Distribution:

Even with simulated data, slight distribution differences may exist between the training and test sets, affecting the model's generalization to the test set. In real applications, the test set may contain features unseen in the training set, impacting classification performance.

3. Dataset Size Limitations:

The limited number of test samples may not fully represent the target data's true distribution, causing fluctuations in test performance and reducing result stability.

4. Model Architecture Limitations:

Although ResNet50 is a powerful deep convolutional neural network, it may still struggle to capture certain fine features. For example, subtle structural changes in X-ray images may be challenging for the model to extract and classify effectively.

Conclusion

Challenges in improving test set performance may arise from overfitting, data distribution differences, dataset size limitations, and model architecture suitability. To enhance test set performance, the following strategies can be considered:

1. **IIncrease Training Data Diversity:** Use data augmentation techniques to simulate varied feature distributions, helping the model learn more generalizable features.
2. **Expand Test Set Sample Size:** Adding more test samples can better represent the target data distribution, potentially improving accuracy and result stability.
3. **Apply Regularization Techniques:** Techniques like Dropout or L2 regularization can reduce overfitting, enhancing the model's generalization capability.

By implementing these measures, the model's generalization performance on the test set can likely be further improved.