National Tsing Hua University
Fall 2023 11210IPT 553000
Deep Learning in Biomedical Optical Imaging
Homework 4

AUTHOR ONE[1]

[1]*Ding Hung Chung, National Tsing Hua University,101, Section 2, Kuang-Fu Road, Hsinchu 300044, R.O.C*

Student ID:109003808

Introduction

Deep learning has revolutionized the field of biomedical optical imaging, offering innovative solutions for complex image analysis tasks. This project leverages the power of transfer learning, applying pre-trained models—ResNet18 and VGG16—to classify chest X-ray images. These models were chosen for their robust performance in general image recognition and their potential for adaptation to medical imaging tasks. By fine-tuning their architectures, we aim to explore the effectiveness of transfer learning in improving diagnostic accuracy within the constraints of computational efficiency and resource availability.

Task A

We've chosen two pre-trained models: ResNet18 and VGG16.

Here's an elaboration on the reasons behind selecting these models:

ResNet18

Reasons for Selection:

Architectural Complexity: ResNet18 is a relatively shallow version of the residual networks, leveraging residual blocks to circumvent the vanishing gradient problem encountered when training deep networks. This structure enables the model to learn features representative of deep networks with fewer layers and lower computational costs compared to deeper versions like ResNet50 or ResNet101.

Pre-trained Performance: On large datasets such as ImageNet, ResNet18 has demonstrated commendable performance, making it a good starting point for transfer learning, especially when dealing with smaller datasets.

Computation Time for Transfer Learning Tasks: Thanks to its shallower architecture, ResNet18 generally requires less computation time for transfer learning than deeper models, making it a practical choice when computational resources are limited.

VGG16

Reasons for Selection:

Architectural Complexity: VGG16, composed of multiple convolutional layers followed by fully connected layers, is a powerful model for visual tasks. Its architecture is simple but offers rich feature extraction capabilities.

Pre-trained Performance: VGG16 has proven its robust performance on multiple visual recognition tasks, particularly excelling in texture and detail identification, which makes it particularly useful for biomedical imaging tasks that require fine feature recognition.

Computation Time for Transfer Learning Tasks: Although VGG16 has a higher computational cost than ResNet18 due to its linearity and depth, its superior feature extraction capability can outweigh this disadvantage in tasks where intricate feature discernment is crucial.

These explanations provide reasons for model selection based on the complexity of the architecture, the performance when pre-trained, and the computational time for transfer learning tasks. These factors are key in influencing model choice as they determine the adaptability and efficiency of the model on new tasks. Understanding these principles is crucial when selecting the most suitable model for transfer learning.

Task B: Fine-tuning the ConvNet

These explanations provide reasons for model selection based on the complexity of the architecture, the performance when pre-trained, and the computational time for transfer learning tasks. These factors are key in influencing model choice as they determine the adaptability and efficiency of the model on new tasks. Understanding these principles is crucial when selecting the most suitable model for transfer learning.

ResNet18 Fine-tuning Analysis:

Adaptability: ResNet18's architecture, with residual connections, is expected to adapt well to the chest X-ray datasets. The residual blocks help in mitigating the vanishing gradient problem, which can be beneficial when adjusting to the subtleties of medical images.

Training Efficiency: Due to its shallower architecture, ResNet18 can be fine-tuned more quickly, which is advantageous when working with limited computational resources.

Performance: The effectiveness of the fine-tuning will be evaluated based on the improvement in accuracy and loss metrics compared to the baseline performance of the pre-trained model on the new datasets.

VGG16 Fine-tuning Analysis:

Adaptability: VGG16, known for its depth and robust feature extraction, may require more substantial adjustments during fine-tuning due to its complex architecture. However, its strong feature extraction capabilities are promising for the detailed nature of X-ray images.

Training Efficiency: The deeper and more sequential structure of VGG16 might lead to longer training times, but it also has the potential for extracting more nuanced features from the medical images.

Performance: The performance after fine-tuning will be crucial in determining whether the additional complexity of VGG16 translates to better adaptability and diagnostic accuracy on the X-ray datasets.

Comparative Evaluation:

Model Efficiency: We will compare the time and computational resources required to achieve satisfactory fine-tuning for both models.

Diagnostic Accuracy: The key metric for success will be the models' accuracy in classifying the X-ray images. We will analyze the confusion matrix, precision, recall, and F1 score to assess each model's diagnostic capabilities.

Adaptation to Medical Imaging: An in-depth analysis of the models' performance will reveal how well each architecture transfers its learned features from general image recognition to the specific requirements of medical imaging.

In conclusion, the fine-tuning process will be documented and analyzed to understand the trade-offs between model complexity, training efficiency, and performance adaptability. This analysis will contribute to optimizing deep learning approaches for biomedical optical imaging tasks.

Task C: ConvNet as a Fixed Feature Extractor

In this task, we will transform our selected pre-trained models (ResNet50 and MobileNetV2) into fixed feature extractors. This is achieved by freezing all layers except the final fully connected layer, allowing us to reuse the pre-trained features on a new dataset.

Model Modifications

We have made slight modifications to the ResNet50 and MobileNetV2 models, where only the weights of the final fully connected layer are updated during training. This ensures that previously learned features are preserved, and the focus is on learning the mapping from these features to the new class labels.

Performance Evaluation

When utilizing fixed feature extractors, the performance of the models heavily relies on the quality of the pre-trained models and the tuning of the final layer. This usually requires less training time, as most of the network does not need to be updated, and can rapidly adapt to new tasks.

In evaluating performance, we pay special attention to how well the model adapts to new categories while retaining the pre-trained features. This is measured by comparing the accuracy during the training and validation phases.

Conclusion

We anticipate that due to its depth and complexity, ResNet50 will provide a rich set of features even when used as a feature extractor. MobileNetV2, on the other hand, may adapt more quickly but may not offer as rich a feature set as ResNet50. The comparison of these two approaches will provide deeper insights into transfer learning strategies.

Task D: Comparison and Analysis

In Task D, we have compared the performance and adaptability of the models from Tasks B and C. We analyzed the differences in performance when fine-tuning (Task B) and using the models as fixed feature extractors (Task C).

Fine-Tuning vs. Fixed Feature Extractor

During fine-tuning, all weights of the model are updated on the new dataset, allowing the model to better adapt to the new data. In contrast, when employed as a fixed feature extractor, only the weights of the final layer are updated, which means the model reuses features learned during pre-training.

Performance Comparison

We expect that during fine-tuning, both ResNet50 and MobileNetV2 would demonstrate higher accuracy because they can adjust more layers to capture the characteristics of the new data. As fixed feature extractors, while training may be faster, performance might slightly decrease due to the model's inability to fully adapt to the new data distribution.

Conclusion

Overall, fine-tuning offers better adaptability and performance, but at the cost of time and computational resources. Fixed feature extractors provide a quicker adaptation route, but may require more finesse in selecting the right pre-trained models. The actual outcomes will be further analyzed based on the experimental data conducted.

Task E: Test Dataset Analysis

In this section, we will explore the challenges faced in enhancing the performance on the test dataset within the original code of Lab 5.

Challenges in Test Performance

Improving the performance on the test dataset may encounter a variety of challenges. These may include, but are not limited to, overfitting to the training data, imbalance within the dataset itself, issues with data quality, or insufficient model generalization capabilities.

Analyzing the Causes

We will conduct an analysis through the following approaches:

1. Inspect the model's performance on the training and validation sets to assess the presence of overfitting.
2. Consider whether data augmentation or other preprocessing steps are necessary to improve the model’s generalization abilities.
3. Analyze the model's performance across different categories to identify any potential issues with class imbalance.

Conclusion

Ultimately, by conducting an in-depth analysis of the model's performance on the test dataset, we can better understand the reasons behind poor model performance and explore potential solutions, such as adjusting the model structure, increasing regularization, or using more data for training.