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

WISARUD YONGBANJERD[1]

109003888:

1. Data Analysis

Our dataset comprises histological images designed to facilitate the automated classification of tissue textures commonly associated with cancer histology. The dataset contains a total of 3,150 images, divided into 2,550 for training and 600 for validation. Additionally, there are 600 test images. Each image has dimensions of 150 x 150 pixels and is in RGB format.

The dataset features six distinct classes: Tumor, Stroma, Complex, Lympho, Debris, and Mucosa. The classes are balanced, with each having approximately 425 images. The visualizations provided in Figure 1 offer insight into the dataset's composition.

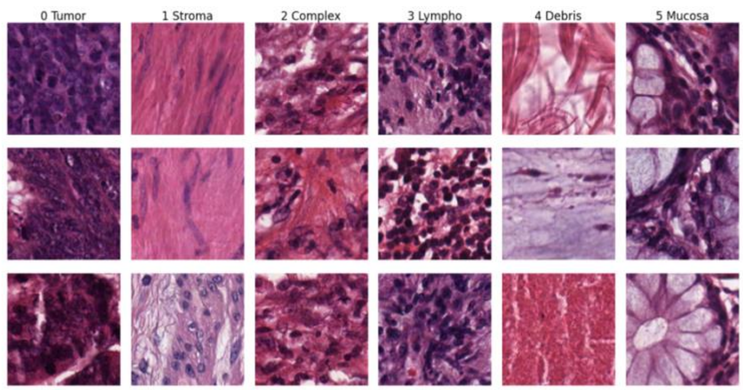


Fig. 1. The result of each of the images according to the class.

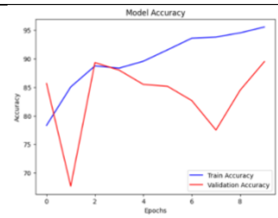
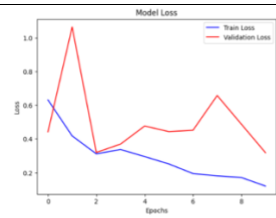
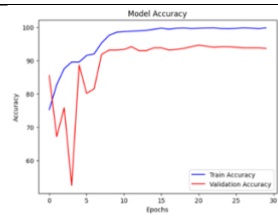
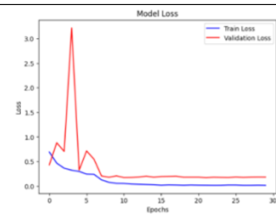
2. Choice of model

Considering the task at hand—classifying histological images of cancerous tissues—Convolutional Neural Networks (CNNs) were the natural choice due to their long history of successful applications in image classification tasks, including those in medical image analysis. CNNs are known for their computational efficiency and their ability to achieve commendable results with relatively less data, attributes that are particularly beneficial when leveraging architectures like ResNet. Utilizing pre-trained CNN models on datasets such as ImageNet allows for fine-tuning on specific datasets, which can lead to robust performance with a reduced need for extensive training from scratch.

3. Training and Testing

The model was initialized using the pre-trained ResNet18 architecture, which was then fine-tuned for our specific task of six-class classification. To maintain consistency across datasets, identical preprocessing steps—such as resizing and normalization—were applied to the training, validation, and test datasets. Initial indications of overfitting led to the implementation of dropout layers and L2 regularization within the model's fully connected layer. After experimentation, the dropout rate was set to 0.3, and the weight decay for the optimizer was established at 5e-4. Additionally, a learning rate scheduler was employed to methodically reduce the learning rate by a factor of 0.1 every 7 epochs, thereby enabling more precise weight adjustments during the course of training. The training duration was extended to 30 epochs, up from the initial 10, to afford the model additional learning opportunities. The impact of these modifications is depicted in Table 1, which contrasts the results before and after the implementation of these enhancements.

Table 1. Showing results from running Resnet18 model

			Test Accuracy
Training with Restnet 18 without finetuning any parameters			90.83%
Training with Restnet 18 after finetuning mentioned parameters			94.67%

3.1 Pre-Fine-tuning Training Performance:

Before fine-tuning, the training accuracy began just under 80%, quickly rising to peak at approximately 95% by the third epoch, with subsequent fluctuations leading to a near 97% by the final epoch. Conversely, the validation accuracy started at roughly 75%, spiked to almost 90% by the second epoch, and then followed a more erratic pattern, concluding the training above 80% but below its peak.

The initial training loss presented just above 0.6, rapidly descending to below 0.2 in the second epoch and then gently reducing to about 0.1 by the end of training. The validation loss, starting just below 0.8, experienced a sharp decline, a spike surpassing 1.0, and ended just below 0.4, with a series of fluctuations. These patterns in the validation metrics may indicate overfitting, suggesting that while the model learned effectively from the training data, it was less adept at generalizing to new data. This overfitting is evidenced by significant shifts in performance, possibly exacerbated by a limited validation set.

3.2 Post-Fine-tuning Training Performance:

After fine-tuning the parameters, the training accuracy demonstrated a consistent upward trajectory until about the 15th epoch, where it plateaued, ultimately reaching approximately 100%. The validation accuracy, despite initial volatility, began to stabilize after roughly ten epochs, aligning more closely with the training accuracy, which implies enhanced generalization capabilities.

The training loss mirrored this improvement, sharply declining and maintaining a low level throughout the remaining epochs. The validation loss, while initially peaking, similarly leveled out, paralleling the stabilization seen in the validation accuracy. These adjustments in the loss curve suggest a model less prone to overfitting than before. The narrowed discrepancy between training and validation accuracy and the more consistent validation loss reflect the effectiveness of the strategies implemented to mitigate overfitting.

4. Discussion

The implemented changes led to a marked improvement in the model's performance. Test accuracy soared to 94.67%, a notable enhancement from the baseline accuracy. The graphs depicting training and validation accuracy demonstrate greater congruence, suggesting that the degree of overfitting has been substantially reduced. Nevertheless, the persistence of minor fluctuations in validation accuracy points to the possibility of further refining the model.

The introduction of dropout and L2 regularization was pivotal in bolstering the model's ability to generalize. Moreover, the incorporation of a learning rate scheduler ensured a more nuanced convergence, contributing to the observed rise in test accuracy. While the potential benefits of data augmentation were acknowledged, its implementation was deferred, primarily due to apprehensions regarding the preservation of data integrity. This decision leaves room for future investigative efforts into the viability and impact of augmentation techniques.

5. Conclusion

The iterative process of refinement undertaken in this study has yielded a robust model adept at classifying histological images. This is evidenced by the notable increase in test accuracy. The techniques utilized here, particularly the systematic tuning of hyperparameters and the application of regularization methods, have proven to be effective. They not only enhance the model's performance but also provide a valuable framework for future improvements. These methods highlight the critical role that careful calibration of model parameters plays in the success of machine learning workflows.

References

1. Pytorch Contributors. (2023). MODELS AND PRE-TRAINED WEIGHTS. https://pytorch.org/vision/main/models.html#table-of-all-available-quantized-classification-weights.
2. Srivastava, N., et al. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research."
3. Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural Networks: Tricks of the Trade.
4. Ng, A. (n.d.). Deep Learning Specialization. Coursera. Retrieved [November 20, 2023], from https://www.coursera.org/specializations/deep-learning.