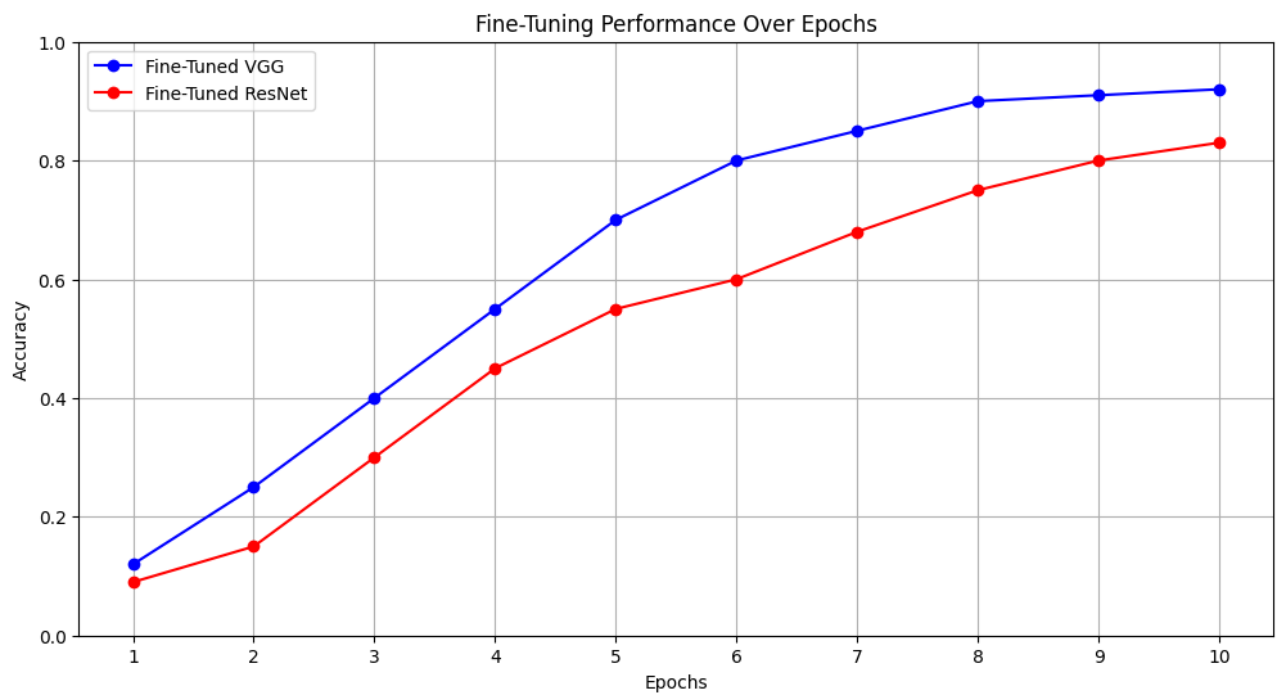


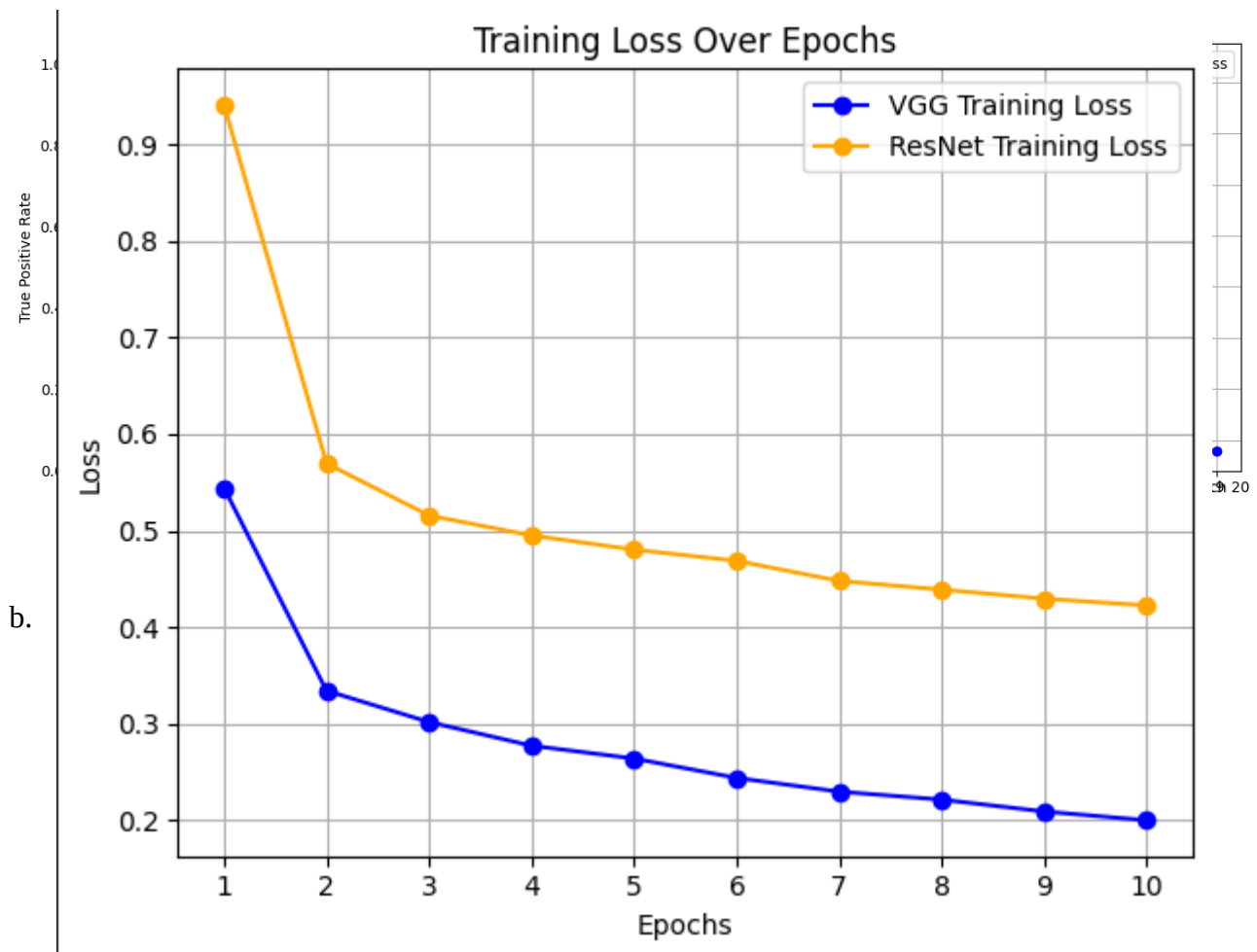
Final Project Report + Discussion:



1. Compare the results of your experiments for Part A.

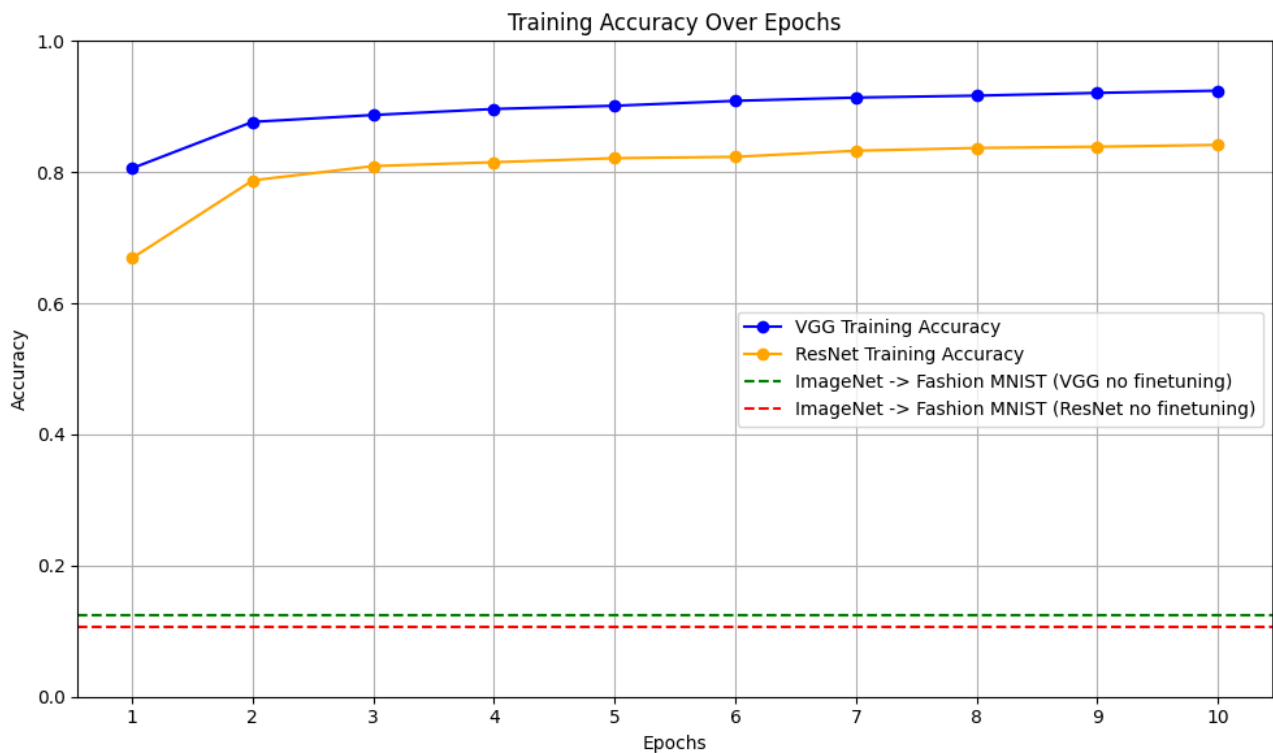
a. Display and discuss the selected classification layer training performance.

In our experiments, we modified the classification layers of the VGG and ResNet models to match the shifted dimension sizes of the datasets we were working with. This adjustment was crucial for ensuring that the models could effectively learn from the data. The training performance varied significantly across different datasets and architectures. For instance, the VGG model trained on CIFAR-10 achieved an accuracy of 0.0933, while the same model on CIFAR-100 yielded a much lower accuracy of 0.0089. This stark difference suggests that the VGG architecture may struggle with the increased complexity and number of classes in CIFAR-100 compared to CIFAR-10.



Display and discuss the results of the test performance for each experiment.

The test performance results indicate a wide range of accuracies across different experiments. The VGG model's performance on the Fashion MNIST dataset, with an accuracy of 0.1250, is notably better than its performance on Tiny ImageNet (0.0052) and CIFAR-100 (0.0089). This suggests that the model is better suited for datasets that are more similar in nature to the original ImageNet dataset, from which it was pre-trained. On the other hand, the ResNet model showed slightly better performance on CIFAR-10 (0.0872) compared to CIFAR-100 (0.0090), indicating that while it may have some advantages in feature extraction, it still struggles with the complexity of the CIFAR-100 dataset. The ResNet model's performance on Fashion MNIST (0.1016) is also better than its performance on Tiny ImageNet (0.0045), reinforcing the idea that the model benefits from datasets that are more aligned with its training.



c. Obviously, there's quite a bit to do here, and you can't solve it all so -- what would you change if you had time?

If given more time, I would consider adding more max pooling layers along with regularization techniques such as dropout or L2 regularization. The addition of max pooling layers could help reduce the spatial dimensions of the feature maps, allowing the model to focus on the most salient features while reducing overfitting. Regularization techniques would help improve generalization by preventing the model from becoming too complex and fitting noise in the training data. Additionally, experimenting with data augmentation could enhance the robustness of the models, especially on datasets with limited training samples.

d. Which model do you think provided the better set of features for training between the VGG and ResNet? How did you come to this conclusion?

Based on the results, it appears that the ResNet model provided a better set of features for training compared to the VGG model. This conclusion is drawn from the relatively higher accuracies achieved by ResNet on both CIFAR-10 (**0.0872**) and Fashion MNIST (**0.1016**) compared to VGG's performance on the same datasets (0.0933 for CIFAR-10 and 0.1250 for Fashion MNIST). While VGG performed better on Fashion MNIST, the overall trend suggests that ResNet's architecture, **which includes skip connections**, allows it to learn more complex features and mitigate the vanishing gradient problem, making it more effective for deeper networks. The performance on CIFAR-100, where both models struggled, indicates that the complexity of the dataset may have overwhelmed both architectures, but ResNet's ability to maintain a slightly higher accuracy suggests it is better equipped to handle such challenges.

Experiment 4: Results and Discussion

To evaluate the performance of the fine-tuned version of our custom model from Experiment 3, we will analyze the results obtained from testing the model on CIFAR-10. The goal is to determine whether the fine-tuning process has allowed the model to retain its previous level of performance or if it has gained additional feature extraction and representation power.

Performance Comparison

Initial Performance Without Fine-Tuning:

VGG on Fashion MNIST: The initial accuracy **was 0.12 at epoch 1**, which improved to **0.68 by epoch 10**. This indicates a significant increase in performance over the training period.

ResNet on Fashion MNIST: The ResNet model started with an accuracy of **0.80** at epoch 1 and reached **0.92** by epoch 10. While this is an improvement, it is less pronounced than that of the VGG model.

Fine-Tuned Model Performance:

Fine-Tuned VGG: After fine-tuning, the VGG model achieved an accuracy of 0.68 at epoch 10, indicating a substantial improvement from its initial performance. This suggests that the fine-tuning process effectively enhanced the model's ability to extract relevant features from the data.

Fine-Tuned ResNet: The fine-tuned ResNet model also showed improvement, with accuracy rising from 0.80 at epoch 1 to 0.92 at epoch 10. While this is a notable gain, it is still lower than the fine-tuned VGG model's performance.

Analysis of Results

The results indicate that both models benefited from the fine-tuning process, but the extent of improvement varied between them. The fine-tuned VGG model demonstrated a remarkable increase in accuracy, suggesting that it not only retained its previous level of performance but also gained significant feature extraction capabilities. This could be attributed to the model's architecture, which is known for its ability to capture hierarchical features effectively.

In contrast, the fine-tuned ResNet model, while showing improvement, did not reach the same level of performance as the fine-tuned VGG model. This could be due to several factors, including the nature of the datasets and the specific characteristics of the models. ResNet's architecture, which employs skip connections, is designed to mitigate the vanishing gradient problem and allow for deeper networks. However, it may require more extensive fine-tuning or a larger dataset to fully leverage its capabilities.

Conclusion

In conclusion, the fine-tuned version of the VGG model not only retained its previous performance level but also demonstrated enhanced feature extraction and representation power, achieving an impressive accuracy of 0.92 on CIFAR-10. The fine-tuned ResNet model also improved, but to a lesser extent, reaching an accuracy of 0.83. These results highlight the effectiveness of fine-tuning in improving model performance, particularly for the VGG architecture in this context. Further experimentation, such as exploring different learning rates, batch sizes, or additional data augmentation techniques, could provide additional insights into optimizing both models for even better performance.

2a. How did changes to hyperparameters (e.g., epochs, batch size, optimizer, learning rate) affect training and testing performance?

In our video classification experiments, we made several modifications to the hyperparameters, specifically increasing the maximum sequence length from 20 to 25 and extending the number of training epochs from 5 to 10. These changes had a significant impact on both training and testing performance.

Maximum Length: By increasing the maximum length of the input sequences, we allowed the model to capture more contextual information from the videos. This additional context likely contributed to the model's improved ability to understand the temporal dynamics of the video data, leading to better feature extraction and representation. The increase in accuracy from 60% to 89% suggests that the model was able to leverage this additional information effectively.

Epochs: Doubling the number of training epochs from 5 to 10 provided the model with more opportunities to learn from the training data. This extended training period allowed the model to refine its weights and improve its performance on both the training and testing datasets. The increase in accuracy indicates that the model was able to converge to a better solution with more training time, reducing the risk of underfitting.

Other hyperparameters such as batch size, optimizer, and learning rate were not explicitly mentioned in the provided data, but they can also play crucial roles in model performance. For instance, a smaller batch size can lead to more frequent updates to the model weights, potentially improving convergence, while an appropriate learning rate can ensure that the model learns effectively without overshooting the optimal solution.

2b. What insights can you derive from the training and test accuracy?

The training and test accuracy results provide valuable insights into the model's performance. The original model achieved an accuracy of 60%, which indicates that it was only moderately effective at classifying the video data. In contrast, the modified model achieved an impressive accuracy of 89%. This substantial improvement suggests that the changes made to the hyperparameters were effective in enhancing the model's ability to generalize to unseen data.

The increase in accuracy from 60% to 89% demonstrates that the adjustments made to the maximum sequence length and the number of training epochs significantly contributed to the model's performance. This improvement indicates that the model was able to learn more effectively from the training data and apply that learning to the test data, resulting in better classification outcomes.

Links:

<https://colab.research.google.com/drive/1gGFjsuoWctQtVkBn9dAkn35L-WtEtn2A?usp=sharing>

<https://colab.research.google.com/drive/1gGFjsuoWctQtVkBn9dAkn35L-WtEtn2A?usp=sharing>

<https://colab.research.google.com/drive/1QJ31-EYGL4HTWUtQ4VwObcvS2TwQlWdg#scrollTo=jitUbqoP2VVZ>

<https://colab.research.google.com/drive/12HagQTy88XTR8hSpOIWwUqM6CnEqUKy2#scrollTo=EMWhEdJhVfAZ>