CNNs on CIFAR-10 Using PyTorch **Danial Zohourian**

The objective of this project was to develop and compare three distinct models for image classification on the CIFAR-10 dataset. The task involved training the models on the provided training data and evaluating their performance on a separate test dataset.

Data Split:

The dataset was split into training and test sets. Specifically, 80% of the train data was allocated for training purposes, and the remaining 20% was utilized as validation data during training.

Data Augmentation:

To enhance the diversity and robustness of the training dataset, data augmentation techniques were employed. These included applying transformations such as **Random Horizontal Flip** and **Random Crop** to the images in the training set.

Hyperparameters, Criterion and Optimizer:

The models were trained using the following hyperparameters:

- **Learning Rate:** 0.01
- **Momentum:** 0.9
- **Weight Decay:** 5e-4
- **Optimizer:** SGD
- **Criterion:** Cross Entropy

Early Stopping:

An early stopping mechanism was implemented during the training process. The criterion for early stopping was based on the validation loss not showing improvement for more than 8 consecutive epochs. Once this condition was met, the training process was halted to prevent overfitting.

Confusion Matrix:

It shows the model's performance for each class. Rows represent the actual classes; columns represent predicted classes.

Diagonal elements represent correct predictions, while off-diagonal elements show misclassifications.

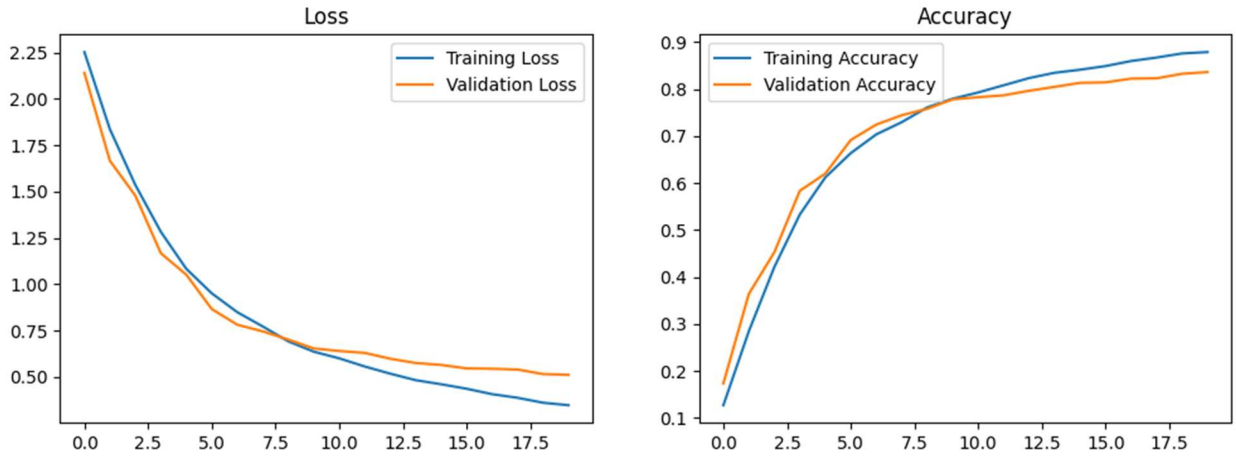
Model 1: VGG8

Training Progress:

- Epochs: 20
- Training Accuracy: Gradually increases from 12.69% to 87.90% over 20 epochs.
- Validation Accuracy: Increases from 17.34% to 83.64% over 20 epochs.
- Training Loss: Decreases steadily from 2.2532 to 0.3475.
- Validation Loss: Similarly decreases from 2.1405 to 0.5115.

Test Accuracy:

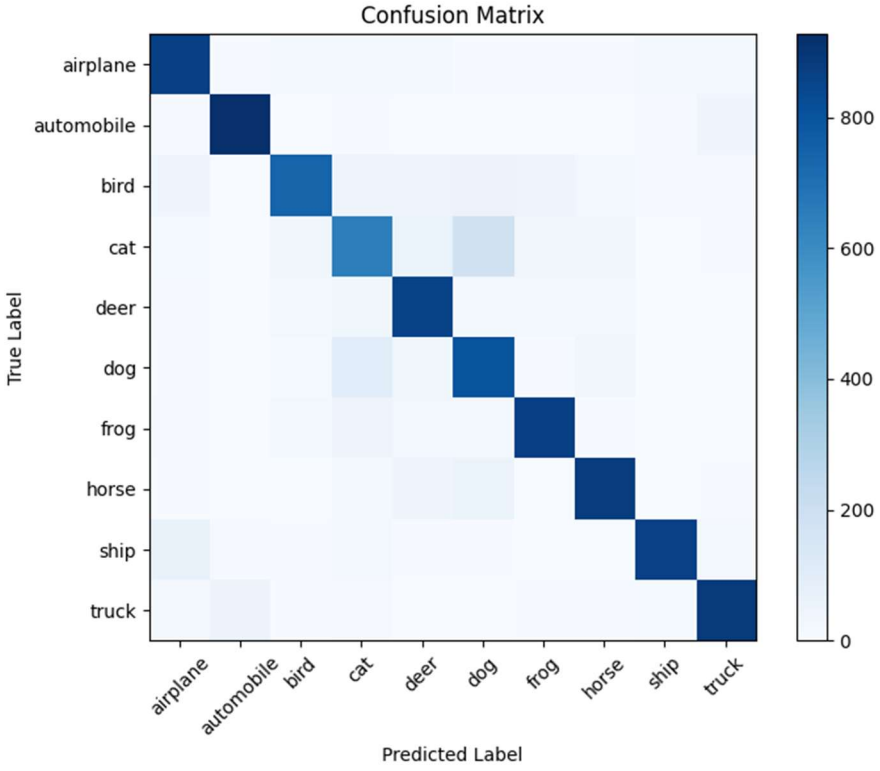
Achieved an accuracy of 83.32% on the test dataset, which is reasonably good.



Confusion Matrix:

There seem to be higher numbers along the diagonal, indicating the model performs better on certain classes (e.g., 'frog', 'deer', 'ship') compared to others (e.g., 'airplane', 'bird', 'cat').

The model appears to struggle with distinguishing classes like 'cat' and 'dog', 'truck' and 'automobile'.



Considerations:

- The model would have performed better with higher number of epochs, as seen in the improving validation numbers

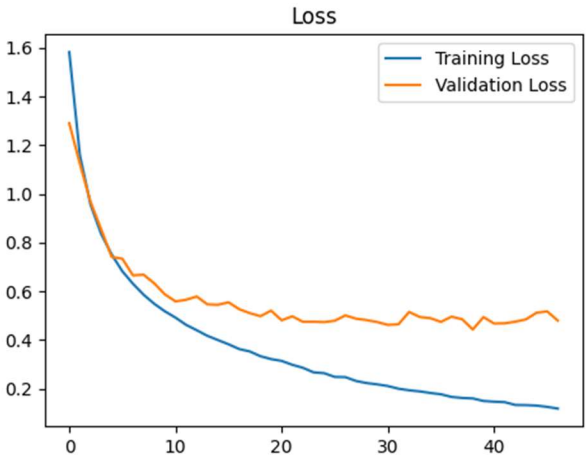
Model 2: ResNet18

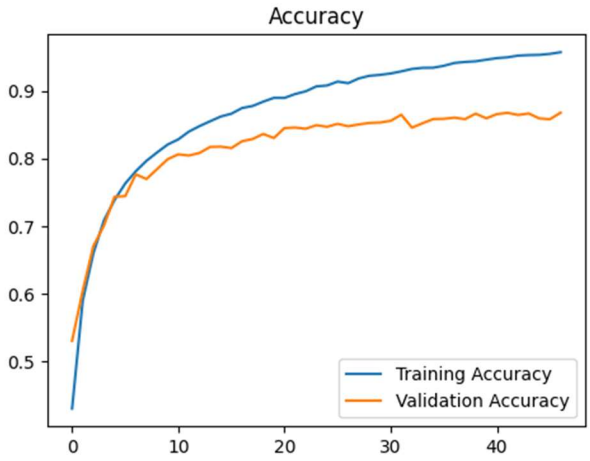
Training Progress:

- **Epochs**: Training was stopped early at epoch 47 due to possibly reaching the optimal performance.
- **Training Accuracy**: Gradually increases from 43.07% to 95.68% over 47 epochs.
- **Validation Accuracy**: Increases from 53.07% to 86.73% over 47 epochs.
- **Training Loss**: Decreases from 1.5824 to 0.1182.
- **Validation Loss**: Decreases from 1.2905 to 0.4799.

Test Accuracy:

Achieved an accuracy of 86.29% on the test dataset, which is notably higher than Model 1's accuracy.

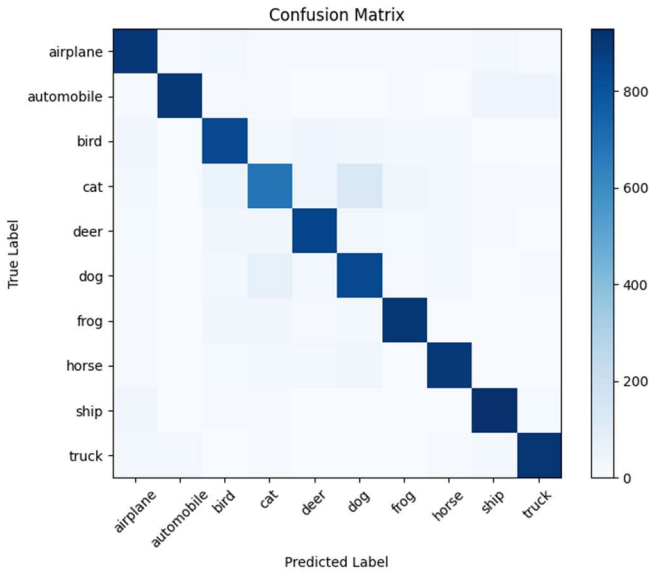




Confusion Matrix:

Generally higher numbers along the diagonal indicate better predictions.

Similar to Model 1, certain classes ('ship', 'automobile', 'bird') have higher accuracy compared to others ('cat', 'dog').



Comparison with Model 1:

Model 2 (ResNet18) achieved a higher test accuracy (86.29%) compared to Model 1 (VGG8 - 83.32%).

ResNet18 seems to perform slightly better in differentiating between classes, as indicated by the confusion matrix.

Considerations:

- Early stopping at epoch 47 might have prevented overfitting but further fine-tuning or regularization techniques could potentially improve performance without risking overfitting.

Model 3: ResNet18 with Transfer Learning

Structure:

- Changed the first default convolution to match the size of CIFAR-10 images.
- Allowed updating gradients to first layer of convolutions.
- Changed the fully connected layer to match the number of classes of CIFAR-10 data.

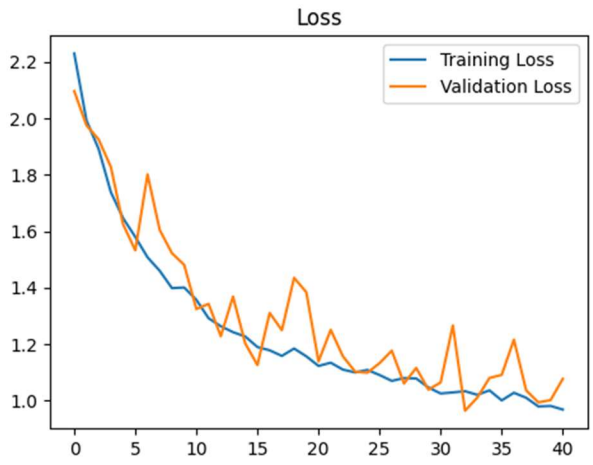
Training Progress:

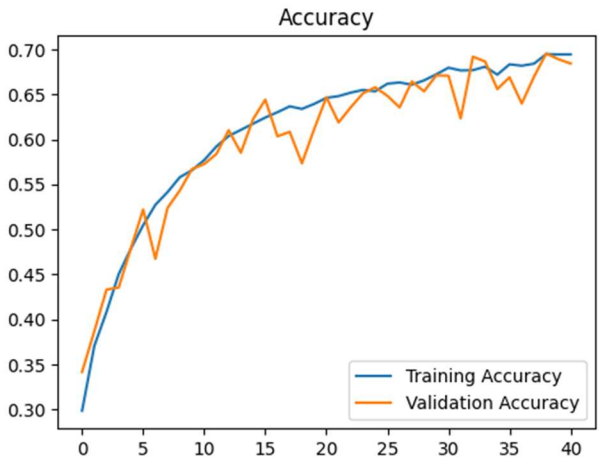
- **Epochs**: Training was stopped early at epoch 41.
- **Training Accuracy**: Gradually increases from 29.85% to 69.45% over 41 epochs.
- **Validation Accuracy**: Increases from 34.17% to 68.44% over 41 epochs.

- **Training Loss**: Decreases from 2.2299 to 0.9685.
- **Validation Loss**: Decreases from 2.0967 to 1.0772.

Test Accuracy:

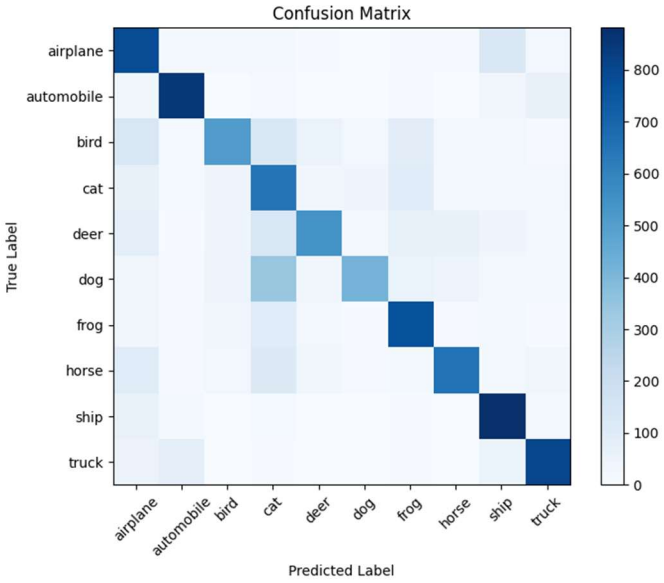
- Achieved an accuracy of 68.48% on the test dataset, which is lower than both Model 1 and Model 2.





Confusion Matrix:

Some classes ('airplane', 'bird', 'cat') seem to have relatively lower accuracy compared to others ('ship', 'horse').



Comparison with Previous Models:

Model 3's performance (Transfer ResNet18) is lower than both Model 1 (VGG8) and Model 2 (ResNet18) in terms of test accuracy.

Considerations:

- The model might not have been fine-tuned extensively, or the specific transfer learning strategy might not have been highly effective for this dataset.
- Further exploration of fine-tuning strategies or adjusting the transfer learning approach could potentially enhance the model's performance.
- Early stopping might have prevented the model to train better, as the plots seem to show an overall upward trend.

Conclusion:

In summary, the three models exhibited distinct performances on the CIFAR-10 dataset. While Model 2 (ResNet18) emerged as the top performer with the highest test accuracy, Models 1 (VGG8) and 3 (Transfer ResNet18) showcased differing strengths and weaknesses, indicating potential areas for further improvement.