

Assignment 4

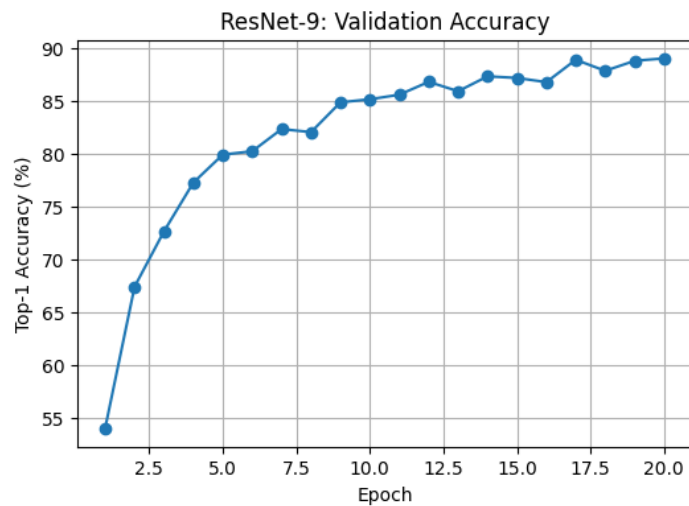
Bharath Prakash
bharathp@usc.edu

I have tried 3 models, one with the normal ResNet-9 classifier, one after fine-tuning it with full ViT model and one with fine-tuning only the classifier head of the ViT model. I have shared the results for all 3 below.

Results for ResNet-9 classifier



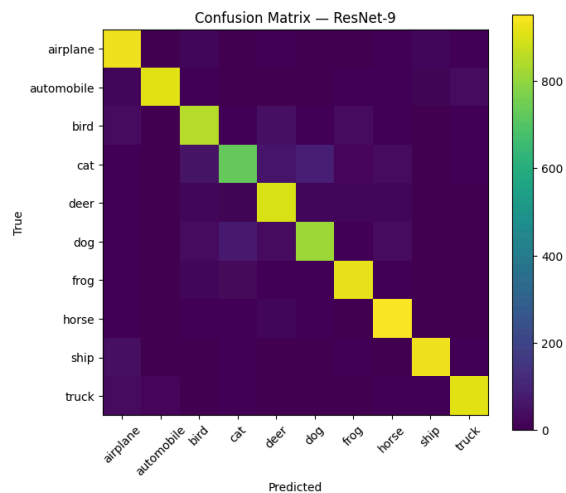
Training loss variation per step



Validation accuracy variation per epoch

	precision	recall	f1-score	support
airplane	0.8499	0.9340	0.8899	1000
automobile	0.9672	0.9130	0.9393	1000
bird	0.8456	0.8490	0.8473	1000
cat	0.8307	0.7260	0.7748	1000
deer	0.8410	0.8990	0.8690	1000
dog	0.8528	0.8110	0.8314	1000
frog	0.9039	0.9220	0.9129	1000
horse	0.8815	0.9520	0.9154	1000
ship	0.9463	0.9340	0.9401	1000
truck	0.9393	0.9130	0.9260	1000
accuracy			0.8853	10000
macro avg	0.8858	0.8853	0.8846	10000
weighted avg	0.8858	0.8853	0.8846	10000

Classification report depicting each class



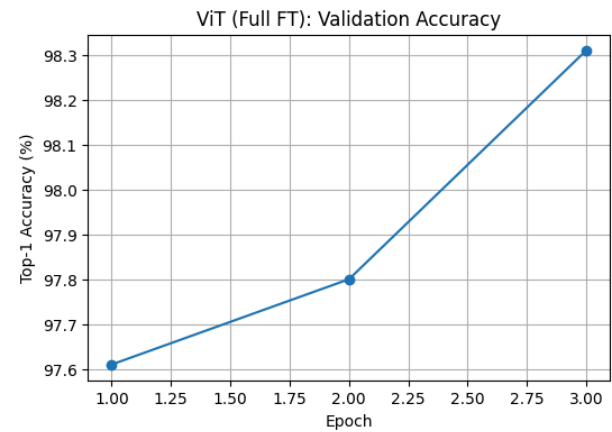
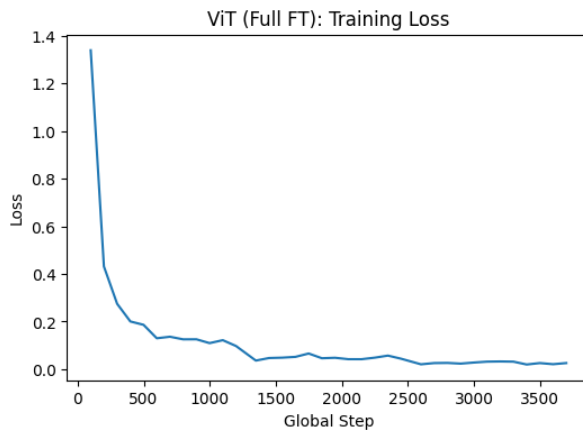
Confusion matrix

Best Validation Accuracy = 89.04%

Test Accuracy = 88.53%

Since we have a balanced dataset, the classification accuracy for each class is more or less similar.

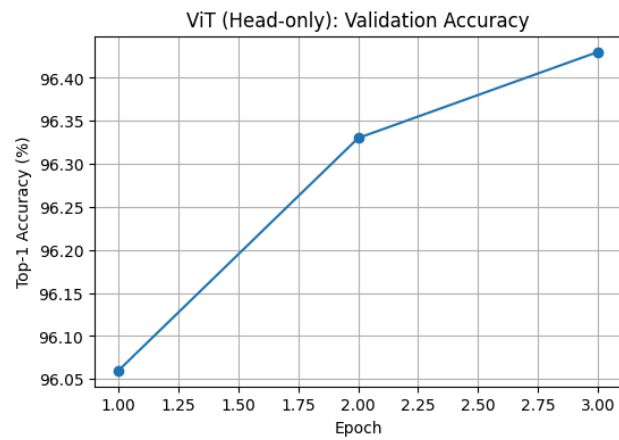
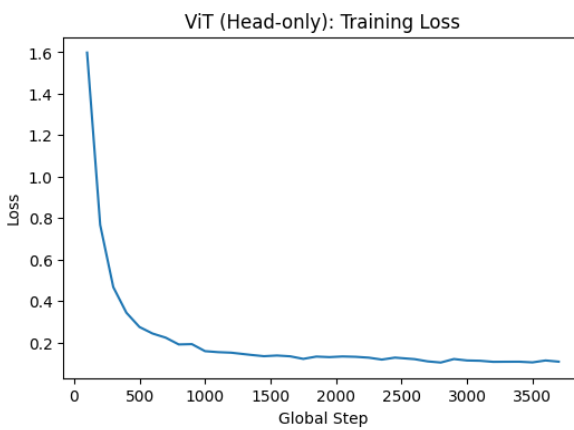
Results for full ViT and head-only ViT classifier



Training loss per step and validation accuracy for 3 epoch training for full ViT classifier

Best Validation Accuracy = 98.31%

Test Accuracy = 97.93%



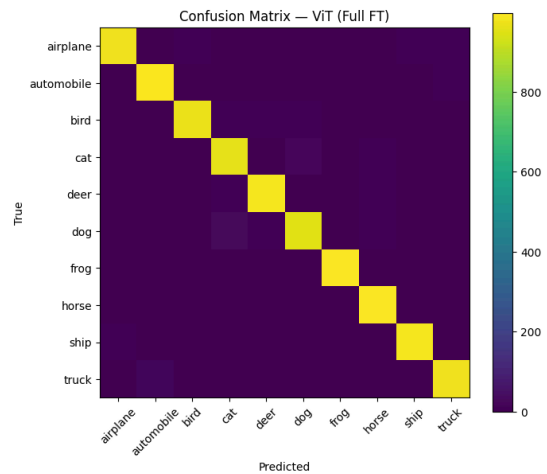
Training loss per step and validation accuracy for 3 epoch training for head-only ViT classifier

Best Validation Accuracy = 96.43%

Test Accuracy = 96.40%

Classification report (ViT Full):

	precision	recall	f1-score	support
airplane	0.9899	0.9770	0.9834	1000
automobile	0.9783	0.9920	0.9851	1000
bird	0.9908	0.9700	0.9803	1000
cat	0.9514	0.9600	0.9557	1000
deer	0.9762	0.9850	0.9806	1000
dog	0.9636	0.9530	0.9583	1000
frog	0.9930	0.9950	0.9940	1000
horse	0.9755	0.9960	0.9857	1000
ship	0.9900	0.9870	0.9885	1000
truck	0.9849	0.9780	0.9814	1000
accuracy			0.9793	10000
macro avg	0.9794	0.9793	0.9793	10000
weighted avg	0.9794	0.9793	0.9793	10000

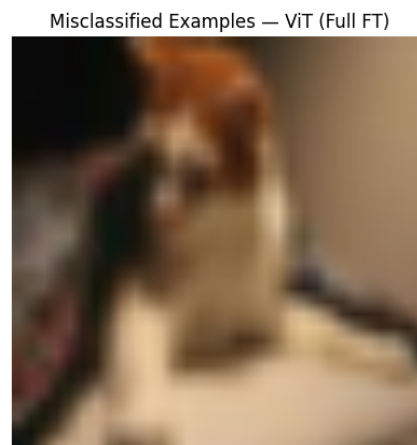


Classification report and confusion matrix for full-ViT classifier depicting each class

Here are some examples of misclassified samples:



True value: Dog
Predicted value: Deer



True value: Cat
Predicted value: Dog

Summary

The ResNet-9 model gives decent accuracy for a 10-class classification, if we train for more than 20 epochs, the accuracy has potential to improve as the loss kept reducing and the validation accuracy kept increasing.

Using the ViT classifier massively improved the accuracy for just 3 epochs, although the training time was comparatively high. This is where fine-tuning the head is massively useful as it took lesser training time to achieve almost similar accuracy. Below is the final summary of all 3 models.

Model	Epoch	Best Validation Accuracy	Test Accuracy
ResNet-9	20	89.04%	88.53%
Full ViT	3	98.31%	97.93%
Head-only ViT	3	96.43%	96.40%