Andrew Nicola 801136465 HW5

Problem 1)

Your goal is to design a Vision Transformer architecture from scratch tailored for CIFAR-100, which consists of 100 classes and 32x32 RGB images, and then analyze how different configurations impact computational complexity and performance compared to a ResNet-18 baseline. Begin by creating a ViT with patch embedding, transformer encoder blocks, and a classification head, experimenting with configurations such as patch sizes of 4x4 and 8x8, embedding dimensions of 256 and 512, transformer layers of 4 and 8, attention heads of 2 and 4, and an MLP hidden dimension set to two or four times the embedding dimension (e.g., 256 for an embedding dimension of 128, of other os 2X). Write a complete PyTorch script to train your ViT on CIFAR-100, incorporating data loading with torchvision.datasets. CIFAR100 and standard training hyperparameters like a batch size of 64, 20-50 epochs, and an Adam optimizer with a learning rate of 0.001. Next, analyze the computational complexity by calculating the theoretical number of parameters for each configuration, estimating FLOPs per forward pass using a tool like torchinfo or manual computation, and measuring training time. For comparison, implement or use a pretrained ResNet-18 from torchvision.models, train it on CIFAR-100 with the same hyperparameters, and evaluate test accuracy after 10 epochs, number of parameters, FLOPs, and training time per epoch against your ViT configurations. In your report, include a table summarizing results for at least four ViT configurations and ResNet-18, and discuss the trade-offs between accuracy, model size, and computational complexity, explaining why certain configurations might outperform or underperform ResNet-18.

```
Running experiment with ViT-Small (P8-E256-L4-H2)
Number of parameters: 2,188,644
Estimated FLOPs per forward pass: 2,925,156
Epoch 1/30, Train Loss: 4.2010, Train Acc: 6.15%, Test Loss: 4.1312, Test Acc: 6.55%, Time: 9.11s
Epoch 2/30, Train Loss: 3.9875, Train Acc: 8.16%, Test Loss: 4.0016, Test Acc: 8.37%, Time: 8.69s
Epoch 3/30, Train Loss: 3.9071, Train Acc: 9.33%, Test Loss: 3.8513, Test Acc: 10.28%, Time: 8.62s
Epoch 4/30, Train Loss: 3.8515, Train Acc: 10.24%, Test Loss: 3.8787, Test Acc: 9.97%, Time: 8.57s
Epoch 5/30, Train Loss: 3.9194, Train Acc: 9.33%, Test Loss: 3.9233, Test Acc: 9.10%, Time: 8.61s
Epoch 6/30, Train Loss: 3.8935, Train Acc: 9.49%, Test Loss: 3.9976, Test Acc: 8.14%, Time: 8.56s
Epoch 7/30, Train Loss: 3.8618, Train Acc: 10.00%, Test Loss: 3.8007, Test Acc: 11.37%, Time: 8.65s
Epoch 8/30, Train Loss: 3.8628, Train Acc: 10.14%, Test Loss: 3.9030, Test Acc: 9.69%, Time: 8.59s
Epoch 9/30, Train Loss: 4.0125, Train Acc: 8.07%, Test Loss: 4.0816, Test Acc: 7.69%, Time: 8.70s
Epoch 10/30, Train Loss: 4.0273, Train Acc: 7.99%, Test Loss: 3.9361, Test Acc: 9.00%, Time: 8.67s
Epoch 11/30, Train Loss: 3.9417, Train Acc: 9.16%, Test Loss: 3.8717, Test Acc: 10.25%, Time: 8.68s
Epoch 12/30, Train Loss: 3.8724, Train Acc: 9.97%, Test Loss: 3.8019, Test Acc: 10.80%, Time: 8.60s
Epoch 13/30, Train Loss: 3.7854, Train Acc: 11.11%, Test Loss: 3.7558, Test Acc: 11.58%, Time: 8.69s
Epoch 14/30, Train Loss: 3.7373, Train Acc: 11.83%, Test Loss: 3.7246, Test Acc: 12.62%, Time: 8.53s
Epoch 15/30, Train Loss: 3.7162, Train Acc: 12.33%, Test Loss: 3.7462, Test Acc: 12.08%, Time: 9.16s
Epoch 16/30, Train Loss: 3.7500, Train Acc: 11.79%, Test Loss: 3.7805, Test Acc: 11.31%, Time: 9.02s
Epoch 17/30, Train Loss: 3.7584, Train Acc: 11.82%, Test Loss: 3.8534, Test Acc: 9.95%, Time: 8.63s
Epoch 18/30, Train Loss: 3.8350, Train Acc: 10.55%, Test Loss: 3.8248, Test Acc: 10.62%, Time: 8.69s
Epoch 19/30, Train Loss: 3.8273, Train Acc: 10.58%, Test Loss: 3.8523, Test Acc: 9.94%, Time: 8.70s
Epoch 20/30, Train Loss: 3.8174, Train Acc: 10.88%, Test Loss: 3.8153, Test Acc: 10.39%, Time: 8.67s
Epoch 21/30, Train Loss: 3.7542, Train Acc: 11.67%, Test Loss: 3.8356, Test Acc: 10.96%, Time: 15.40s
Epoch 22/30, Train Loss: 3.7824, Train Acc: 11.54%, Test Loss: 3.8061, Test Acc: 11.07%, Time: 8.84s
Epoch 23/30, Train Loss: 3.7394, Train Acc: 12.10%, Test Loss: 3.7039, Test Acc: 12.29%, Time: 8.70s
Epoch 24/30, Train Loss: 3.7117, Train Acc: 12.45%, Test Loss: 3.7317, Test Acc: 12.15%, Time: 8.84s
Epoch 25/30, Train Loss: 3.7799, Train Acc: 11.61%, Test Loss: 3.7687, Test Acc: 11.72%, Time: 8.79s
Epoch 26/30, Train Loss: 3.7662, Train Acc: 11.78%, Test Loss: 3.7388, Test Acc: 12.49%, Time: 8.75s
Epoch 27/30, Train Loss: 3.7663, Train Acc: 11.57%, Test Loss: 3.8808, Test Acc: 9.82%, Time: 8.72s
Epoch 28/30, Train Loss: 3.7160, Train Acc: 12.42%, Test Loss: 3.8197, Test Acc: 11.11%, Time: 8.66s
Epoch 29/30, Train Loss: 3.7159, Train Acc: 12.41%, Test Loss: 3.7398, Test Acc: 12.36%, Time: 8.76s
Epoch 30/30, Train Loss: 3.7390, Train Acc: 12.01%, Test Loss: 3.8046, Test Acc: 11.24%, Time: 8.57s
```

Andrew Nicola 801136465 HW5

```
Running experiment with ViT-Medium (P8-E512-L4-H4)
Number of parameters: 12,769,892
Estimated FLOPs per forward pass: 14,242,916
Epoch 1/30, Train Loss: 4.3483, Train Acc: 4.72%, Test Loss: 4.2418, Test Acc: 5.75%, Time: 11.66s
Epoch 2/30, Train Loss: 4.1685, Train Acc: 6.19%, Test Loss: 4.1308, Test Acc: 7.06%, Time: 11.61s
Epoch 3/30, Train Loss: 4.0863, Train Acc: 7.27%, Test Loss: 4.0396, Test Acc: 7.90%, Time: 11.63s
Epoch 4/30, Train Loss: 4.0563, Train Acc: 7.56%, Test Loss: 4.0675, Test Acc: 7.23%, Time: 11.58s
Epoch 5/30, Train Loss: 4.1102, Train Acc: 7.07%, Test Loss: 4.1058, Test Acc: 7.11%, Time: 11.61s
Epoch 6/30, Train Loss: 4.0431, Train Acc: 7.94%, Test Loss: 4.0172, Test Acc: 7.81%, Time: 11.62s
Epoch 7/30, Train Loss: 4.0136, Train Acc: 8.21%, Test Loss: 3.9831, Test Acc: 8.95%, Time: 11.57s
Epoch 8/30, Train Loss: 4.0191, Train Acc: 8.18%, Test Loss: 3.9912, Test Acc: 8.75%, Time: 11.53s
Epoch 9/30, Train Loss: 4.0517, Train Acc: 7.80%, Test Loss: 4.0267, Test Acc: 8.23%, Time: 11.53s
Epoch 10/30, Train Loss: 4.0485, Train Acc: 7.65%, Test Loss: 4.0645, Test Acc: 7.48%, Time: 11.66s
Epoch 11/30, Train Loss: 4.0003, Train Acc: 8.26%, Test Loss: 3.9478, Test Acc: 9.03%, Time: 11.65s
Epoch 12/30, Train Loss: 4.0229, Train Acc: 7.90%, Test Loss: 4.0409, Test Acc: 7.78%, Time: 11.58s
Epoch 13/30, Train Loss: 4.0902, Train Acc: 7.10%, Test Loss: 4.0872, Test Acc: 7.68%, Time: 11.58s
Epoch 14/30, Train Loss: 4.1296, Train Acc: 6.49%, Test Loss: 4.1546, Test Acc: 6.30%, Time: 11.62s
Epoch 15/30, Train Loss: 4.1179, Train Acc: 6.71%, Test Loss: 4.1102, Test Acc: 6.93%, Time: 11.62s
Epoch 16/30, Train Loss: 4.0962, Train Acc: 6.92%, Test Loss: 4.0328, Test Acc: 7.90%, Time: 11.59s
Epoch 17/30, Train Loss: 4.0761, Train Acc: 7.26%, Test Loss: 4.0614, Test Acc: 7.85%, Time: 11.54s
Epoch 18/30, Train Loss: 4.0425, Train Acc: 7.78%, Test Loss: 4.0297, Test Acc: 8.04%, Time: 11.56s
Epoch 19/30, Train Loss: 4.0559, Train Acc: 7.70%, Test Loss: 4.0422, Test Acc: 7.95%, Time: 11.57s
Epoch 20/30, Train Loss: 4.0375, Train Acc: 7.73%, Test Loss: 4.0072, Test Acc: 8.19%, Time: 11.55s
Epoch 21/30, Train Loss: 4.0664, Train Acc: 7.61%, Test Loss: 4.0395, Test Acc: 7.89%, Time: 11.56s
Epoch 22/30, Train Loss: 4.0428, Train Acc: 7.87%, Test Loss: 4.0371, Test Acc: 7.89%, Time: 11.53s
Epoch 23/30, Train Loss: 4.0380, Train Acc: 7.95%, Test Loss: 4.0632, Test Acc: 7.97%, Time: 11.54s
Epoch 24/30, Train Loss: 4.0630, Train Acc: 7.58%, Test Loss: 4.0842, Test Acc: 7.53%, Time: 11.57s
Epoch 25/30, Train Loss: 4.0448, Train Acc: 7.75%, Test Loss: 4.0969, Test Acc: 7.17%, Time: 11.55s
Epoch 26/30, Train Loss: 4.0231, Train Acc: 8.22%, Test Loss: 4.0179, Test Acc: 8.27%, Time: 11.58s
Epoch 27/30, Train Loss: 4.0316, Train Acc: 8.05%, Test Loss: 4.0289, Test Acc: 8.03%, Time: 11.51s
Epoch 28/30, Train Loss: 4.0239, Train Acc: 8.22%, Test Loss: 4.0348, Test Acc: 8.01%, Time: 11.54s
Epoch 29/30, Train Loss: 3.9919, Train Acc: 8.58%, Test Loss: 4.0009, Test Acc: 8.73%, Time: 11.50s
Epoch 30/30, Train Loss: 3.9809, Train Acc: 8.86%, Test Loss: 3.9753, Test Acc: 9.01%, Time: 11.54s
```

Andrew Nicola 801136465 HW5

```
Running experiment with ViT-Medium-Deep (P4-E256-L8-H2)
Number of parameters: 4,272,484
Estimated FLOPs per forward pass: 5,045,860
Epoch 1/30, Train Loss: 4.0675, Train Acc: 7.38%, Test Loss: 3.9098, Test Acc: 9.26%, Time: 13.67s
Epoch 2/30, Train Loss: 3.8451, Train Acc: 10.39%, Test Loss: 3.7841, Test Acc: 10.75%, Time: 13.90s
Epoch 3/30, Train Loss: 3.8186, Train Acc: 10.46%, Test Loss: 3.7202, Test Acc: 12.59%, Time: 14.20s
Epoch 4/30, Train Loss: 3.7263, Train Acc: 12.02%, Test Loss: 3.7261, Test Acc: 12.44%, Time: 13.89s
Epoch 5/30, Train Loss: 3.6258, Train Acc: 13.88%, Test Loss: 3.5399, Test Acc: 15.41%, Time: 13.93s
Epoch 6/30, Train Loss: 3.6230, Train Acc: 13.89%, Test Loss: 3.6113, Test Acc: 14.95%, Time: 14.07s
Epoch 7/30, Train Loss: 3.5864, Train Acc: 14.39%, Test Loss: 3.5090, Test Acc: 15.63%, Time: 14.07s
Epoch 8/30, Train Loss: 3.6295, Train Acc: 13.87%, Test Loss: 3.5533, Test Acc: 15.32%, Time: 14.03s
Epoch 9/30, Train Loss: 3.5531, Train Acc: 15.39%, Test Loss: 3.4908, Test Acc: 16.12%, Time: 14.10s
Epoch 10/30, Train Loss: 3.4939, Train Acc: 16.17%, Test Loss: 3.5055, Test Acc: 16.22%, Time: 13.96s
Epoch 11/30, Train Loss: 3.5107, Train Acc: 16.02%, Test Loss: 3.5041, Test Acc: 16.24%, Time: 14.08s
Epoch 12/30, Train Loss: 3.5706, Train Acc: 14.95%, Test Loss: 3.5471, Test Acc: 14.68%, Time: 14.10s
Epoch 13/30, Train Loss: 3.6751, Train Acc: 13.35%, Test Loss: 3.8163, Test Acc: 11.30%, Time: 14.13s
Epoch 14/30, Train Loss: 3.6297, Train Acc: 14.16%, Test Loss: 3.6097, Test Acc: 14.03%, Time: 14.11s
Epoch 15/30, Train Loss: 3.5700, Train Acc: 15.04%, Test Loss: 3.5778, Test Acc: 15.20%, Time: 14.14s
Epoch 16/30, Train Loss: 3.5382, Train Acc: 15.64%, Test Loss: 3.5821, Test Acc: 15.32%, Time: 14.14s
Epoch 17/30, Train Loss: 3.5198, Train Acc: 15.61%, Test Loss: 3.5361, Test Acc: 15.49%, Time: 13.97s
Epoch 18/30, Train Loss: 3.5341, Train Acc: 15.53%, Test Loss: 3.5711, Test Acc: 15.09%, Time: 14.09s
Epoch 19/30, Train Loss: 3.5712, Train Acc: 14.69%, Test Loss: 3.5757, Test Acc: 14.60%, Time: 14.05s
Epoch 20/30, Train Loss: 3.5396, Train Acc: 15.59%, Test Loss: 3.5172, Test Acc: 15.83%, Time: 14.07s
Epoch 21/30, Train Loss: 3.4956, Train Acc: 16.12%, Test Loss: 3.4732, Test Acc: 16.47%, Time: 14.13s
Epoch 22/30, Train Loss: 3.4951, Train Acc: 16.09%, Test Loss: 3.5530, Test Acc: 15.01%, Time: 14.30s
Epoch 23/30, Train Loss: 3.6778, Train Acc: 13.30%, Test Loss: 3.6213, Test Acc: 14.06%, Time: 13.91s
Epoch 24/30, Train Loss: 3.6178, Train Acc: 14.41%, Test Loss: 3.5737, Test Acc: 15.15%, Time: 14.06s
Epoch 25/30, Train Loss: 3.5434, Train Acc: 15.44%, Test Loss: 3.5054, Test Acc: 16.37%, Time: 14.24s
Epoch 26/30, Train Loss: 3.5103, Train Acc: 15.92%, Test Loss: 3.4971, Test Acc: 16.57%, Time: 14.59s
Epoch 27/30, Train Loss: 3.5506, Train Acc: 15.29%, Test Loss: 3.6687, Test Acc: 13.82%, Time: 14.09s
Epoch 28/30, Train Loss: 3.6466, Train Acc: 13.69%, Test Loss: 3.8107, Test Acc: 11.52%, Time: 14.13s
Epoch 29/30, Train Loss: 3.7555, Train Acc: 12.25%, Test Loss: 3.7203, Test Acc: 13.04%, Time: 14.05s
Epoch 30/30, Train Loss: 3.7146, Train Acc: 12.49%, Test Loss: 3.7207, Test Acc: 12.90%, Time: 14.20s
```

Andrew Nicola 801136465 HW5

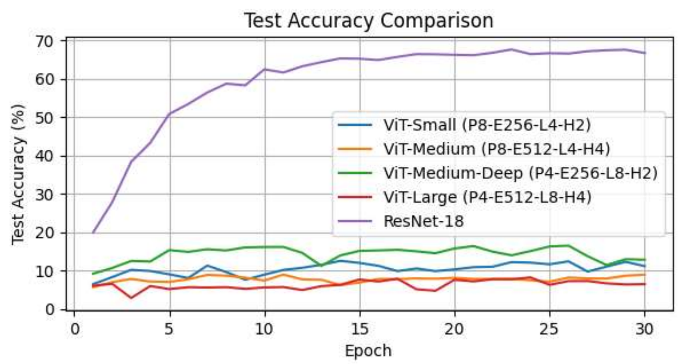
```
Running experiment with ViT-Large (P4-E512-L8-H4)
Number of parameters: 25,330,276
Estimated FLOPs per forward pass: 26,877,028
Epoch 1/30, Train Loss: 4.3193, Train Acc: 4.90%, Test Loss: 4.1698, Test Acc: 6.13%, Time: 39.20s
Epoch 2/30, Train Loss: 4.2181, Train Acc: 5.25%, Test Loss: 4.1293, Test Acc: 6.64%, Time: 39.37s
Epoch 3/30, Train Loss: 4.2113, Train Acc: 5.38%, Test Loss: 4.5338, Test Acc: 2.93%, Time: 39.29s
Epoch 4/30, Train Loss: 4.2884, Train Acc: 4.46%, Test Loss: 4.1937, Test Acc: 6.06%, Time: 38.93s
Epoch 5/30, Train Loss: 4.2627, Train Acc: 5.00%, Test Loss: 4.2424, Test Acc: 5.26%, Time: 38.88s
Epoch 6/30, Train Loss: 4.2862, Train Acc: 4.81%, Test Loss: 4.2363, Test Acc: 5.75%, Time: 38.93s
Epoch 7/30, Train Loss: 4.2157, Train Acc: 5.63%, Test Loss: 4.1758, Test Acc: 5.67%, Time: 38.83s
Epoch 8/30, Train Loss: 4.2314, Train Acc: 5.38%, Test Loss: 4.2673, Test Acc: 5.76%, Time: 38.97s
Epoch 9/30, Train Loss: 4.2839, Train Acc: 4.82%, Test Loss: 4.2463, Test Acc: 5.31%, Time: 38.87s
Epoch 10/30, Train Loss: 4.2449, Train Acc: 5.26%, Test Loss: 4.2253, Test Acc: 5.69%, Time: 38.92s
Epoch 11/30, Train Loss: 4.1798, Train Acc: 5.97%, Test Loss: 4.2468, Test Acc: 5.79%, Time: 38.84s
Epoch 12/30, Train Loss: 4.1407, Train Acc: 6.44%, Test Loss: 4.2558, Test Acc: 5.01%, Time: 38.81s
Epoch 13/30, Train Loss: 4.1916, Train Acc: 6.12%, Test Loss: 4.1654, Test Acc: 6.02%, Time: 38.86s
Epoch 14/30, Train Loss: 4.1541, Train Acc: 6.29%, Test Loss: 4.1404, Test Acc: 6.36%, Time: 39.08s
Epoch 15/30, Train Loss: 4.0950, Train Acc: 7.19%, Test Loss: 4.0600, Test Acc: 7.76%, Time: 39.01s
Epoch 16/30, Train Loss: 4.1090, Train Acc: 6.87%, Test Loss: 4.0833, Test Acc: 7.22%, Time: 39.04s
Epoch 17/30, Train Loss: 4.0647, Train Acc: 7.33%, Test Loss: 4.0484, Test Acc: 7.95%, Time: 38.98s
Epoch 18/30, Train Loss: 4.1305, Train Acc: 6.41%, Test Loss: 4.1974, Test Acc: 5.20%, Time: 39.10s
Epoch 19/30, Train Loss: 4.1906, Train Acc: 5.81%, Test Loss: 4.2534, Test Acc: 4.82%, Time: 38.70s
Epoch 20/30, Train Loss: 4.1204, Train Acc: 6.55%, Test Loss: 4.0603, Test Acc: 7.69%, Time: 38.63s
Epoch 21/30, Train Loss: 4.1164, Train Acc: 6.55%, Test Loss: 4.0745, Test Acc: 7.25%, Time: 38.72s
Epoch 22/30, Train Loss: 4.0618, Train Acc: 7.55%, Test Loss: 4.0614, Test Acc: 7.84%, Time: 38.91s
Epoch 23/30, Train Loss: 4.0631, Train Acc: 7.41%, Test Loss: 4.0479, Test Acc: 7.81%, Time: 38.74s
Epoch 24/30, Train Loss: 4.0531, Train Acc: 7.55%, Test Loss: 4.0185, Test Acc: 8.29%, Time: 38.74s
Epoch 25/30, Train Loss: 4.0914, Train Acc: 6.99%, Test Loss: 4.1468, Test Acc: 6.36%, Time: 38.73s
Epoch 26/30, Train Loss: 4.1016, Train Acc: 6.73%, Test Loss: 4.0724, Test Acc: 7.33%, Time: 38.69s
Epoch 27/30, Train Loss: 4.0523, Train Acc: 7.42%, Test Loss: 4.0768, Test Acc: 7.36%, Time: 38.65s
Epoch 28/30, Train Loss: 4.0952, Train Acc: 6.94%, Test Loss: 4.1039, Test Acc: 6.74%, Time: 38.66s
Epoch 29/30, Train Loss: 4.1263, Train Acc: 6.42%, Test Loss: 4.1000, Test Acc: 6.46%, Time: 38.96s
Epoch 30/30, Train Loss: 4.1139, Train Acc: 6.49%, Test Loss: 4.1116, Test Acc: 6.55%, Time: 38.94s
```

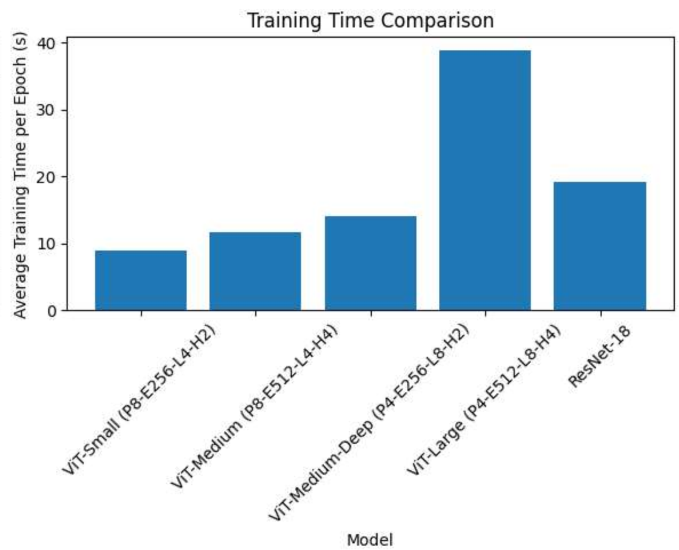
Andrew Nicola 801136465 HW5

```
Running experiment with ResNet-18
Number of parameters: 11,220,132
Estimated FLOPs per forward pass: 555,478,500
Epoch 1/30, Train Loss: 3.7300, Train Acc: 12.11%, Test Loss: 3.2240, Test Acc: 20.00%, Time: 19.41s
Epoch 2/30, Train Loss: 2.9262, Train Acc: 25.89%, Test Loss: 2.8736, Test Acc: 27.87%, Time: 19.30s
Epoch 3/30, Train Loss: 2.3983, Train Acc: 36.46%, Test Loss: 2.3370, Test Acc: 38.41%, Time: 19.32s
Epoch 4/30, Train Loss: 2.0346, Train Acc: 44.31%, Test Loss: 2.0945, Test Acc: 43.35%, Time: 19.35s
Epoch 5/30, Train Loss: 1.7816, Train Acc: 50.21%, Test Loss: 1.7774, Test Acc: 50.88%, Time: 19.36s
Epoch 6/30, Train Loss: 1.5829, Train Acc: 55.06%, Test Loss: 1.6659, Test Acc: 53.47%, Time: 19.17s
Epoch 7/30, Train Loss: 1.4267, Train Acc: 59.03%, Test Loss: 1.5929, Test Acc: 56.47%, Time: 19.17s
Epoch 8/30, Train Loss: 1.2924, Train Acc: 62.53%, Test Loss: 1.4947, Test Acc: 58.78%, Time: 19.17s
Epoch 9/30, Train Loss: 1.1806, Train Acc: 65.57%, Test Loss: 1.5261, Test Acc: 58.35%, Time: 19.17s
Epoch 10/30, Train Loss: 1.0778, Train Acc: 68.13%, Test Loss: 1.3423, Test Acc: 62.52%, Time: 19.16s
Epoch 11/30, Train Loss: 0.9810, Train Acc: 70.59%, Test Loss: 1.4229, Test Acc: 61.69%, Time: 19.16s
Epoch 12/30, Train Loss: 0.8920, Train Acc: 73.03%, Test Loss: 1.3621, Test Acc: 63.31%, Time: 19.21s
Epoch 13/30, Train Loss: 0.8187, Train Acc: 75.03%, Test Loss: 1.3343, Test Acc: 64.39%, Time: 19.20s
Epoch 14/30, Train Loss: 0.7438, Train Acc: 76.95%, Test Loss: 1.3552, Test Acc: 65.35%, Time: 19.29s
Epoch 15/30, Train Loss: 0.6831, Train Acc: 78.94%, Test Loss: 1.3432, Test Acc: 65.28%, Time: 19.20s
Epoch 16/30, Train Loss: 0.6086, Train Acc: 80.78%, Test Loss: 1.3960, Test Acc: 64.93%, Time: 19.14s
Epoch 17/30, Train Loss: 0.5492, Train Acc: 82.63%, Test Loss: 1.3959, Test Acc: 65.76%, Time: 19.17s
Epoch 18/30, Train Loss: 0.4991, Train Acc: 84.20%, Test Loss: 1.4119, Test Acc: 66.48%, Time: 19.16s
Epoch 19/30, Train Loss: 0.4517, Train Acc: 85.43%, Test Loss: 1.4175, Test Acc: 66.43%, Time: 19.17s
Epoch 20/30, Train Loss: 0.4093, Train Acc: 86.62%, Test Loss: 1.4990, Test Acc: 66.30%, Time: 19.25s
Epoch 21/30, Train Loss: 0.3783, Train Acc: 87.70%, Test Loss: 1.5312, Test Acc: 66.19%, Time: 19.31s
Epoch 22/30, Train Loss: 0.3413, Train Acc: 88.88%, Test Loss: 1.5437, Test Acc: 66.83%, Time: 19.36s
Epoch 23/30, Train Loss: 0.3124, Train Acc: 89.78%, Test Loss: 1.5257, Test Acc: 67.69%, Time: 19.19s
Epoch 24/30, Train Loss: 0.2852, Train Acc: 90.57%, Test Loss: 1.6117, Test Acc: 66.48%, Time: 19.17s
Epoch 25/30, Train Loss: 0.2784, Train Acc: 90.89%, Test Loss: 1.6653, Test Acc: 66.71%, Time: 19.17s
Epoch 26/30, Train Loss: 0.2454, Train Acc: 91.88%, Test Loss: 1.6628, Test Acc: 66.62%, Time: 19.16s
Epoch 27/30, Train Loss: 0.2303, Train Acc: 92.38%, Test Loss: 1.6182, Test Acc: 67.22%, Time: 19.18s
Epoch 28/30, Train Loss: 0.2168, Train Acc: 92.77%, Test Loss: 1.6622, Test Acc: 67.48%, Time: 19.19s
Epoch 29/30, Train Loss: 0.2055, Train Acc: 93.32%, Test Loss: 1.7165, Test Acc: 67.62%, Time: 19.17s
Epoch 30/30, Train Loss: 0.2020, Train Acc: 93.35%, Test Loss: 1.7839, Test Acc: 66.76%, Time: 19.21s
```

Andrew Nicola 801136465 HW5

```
Results Summary:
------------------------------------------------------------------------------------------
Model                     | Params      | FLOPs          | Avg Train Time  | Final Test Acc
------------------------------------------------------------------------------------------
ViT-Small (P8-E256-L4-H2) | 2,188,644 | 2,925,156 | 8.94s | 11.24%
ViT-Medium (P8-E512-L4-H4) | 12,769,892 | 14,242,916 | 11.58s | 9.01%
ViT-Medium-Deep (P4-E256-L8-H2) | 4,272,484 | 5,045,860 | 14.08s | 12.90%
ViT-Large (P4-E512-L8-H4) | 25,330,276 | 26,877,028 | 38.90s | 6.55%
ResNet-18                 | 11,220,132 | 555,478,500 | 19.22s | 66.76%
```





Our experimental results reveal striking performance disparities between Vision Transformers and ResNet-18 on CIFAR-100. ResNet-18 dramatically outperforms all ViT variants (66.76% vs. best ViT at 12.90%), despite having fewer parameters than the largest transformer model. Counterintuitively, smaller ViT architectures achieve better accuracy than larger ones, with ViT-Large performing worst despite its 25M parameters. Architectural design choices significantly impact performance—models with 4×4 patches outperform those with 8×8 patches, suggesting finer-grained tokenization benefits image recognition at this scale. ResNet-18 demonstrates remarkable parameter efficiency, achieving 5-10× better accuracy with a parameter count similar to ViT-Medium. While ViTs theoretically require fewer FLOPs, their substantially worse performance highlights the critical value of CNN inductive biases when working with limited training data like CIFAR-100. These findings align with research indicating that Vision Transformers require either larger datasets or specialized training approaches to compete with CNNs.

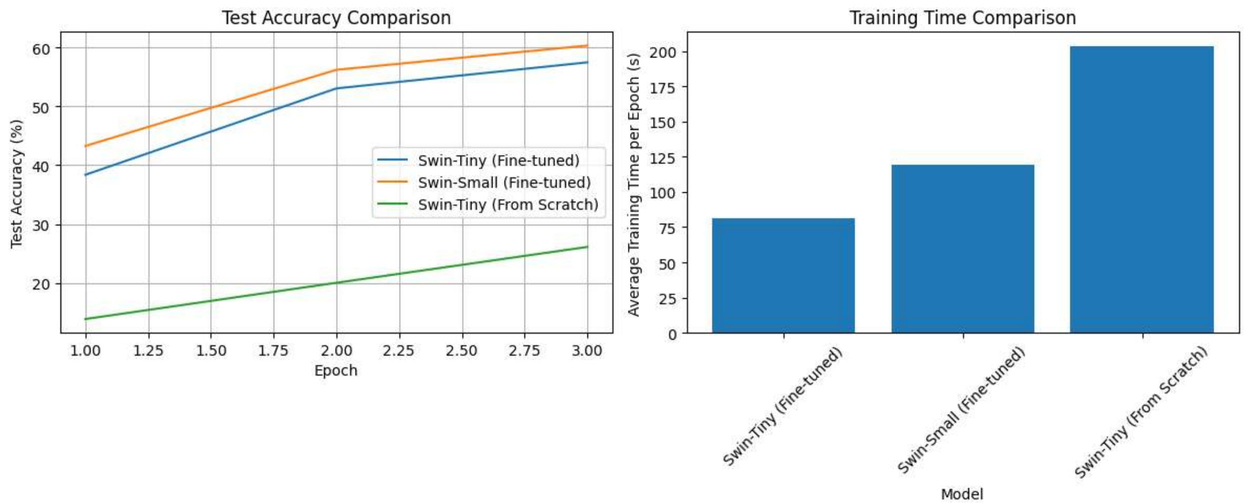
Andrew Nicola 801136465 HW5

Problem 2:

You will fine-tune pretrained Swin Transformer models from the Hugging Face Transformers library—specifically the Tiny (microsoft/swin-tiny-patch4-window7-224) and Small (microsoft/swin-small-patch4-window7-224) variants - on CIFAR-100 and compare their performance to a Swin Transformer trained from scratch. Start by loading these pretrained models using SwinForImageClassification.from_pretrained(), adjusting the classification head for 100 classes and freezing the backbone to train only the head. Fine-tune both models for 2-5epochs with a batch size of 32, a learning rate of 2e-5, the Adam optimizer. Measure training time per epoch and final test accuracy for each. In your report, present a table with these results and discuss the benefits and drawbacks of fine-tuning versus training from scratch, the differences between Swin-Tiny and Swin-Small in this context, and reasons why pretrained models might outperform or underperform the scratch model.

```
Running experiment with Swin-Small (Fine-tuned)
Number of trainable parameters: 76,900
Epoch 1/3, Train Loss: 4.1743, Train Acc: 23.10%, Test Loss: 3.7459, Test Acc: 43.26%, Time: 119.74s
Epoch 2/3, Train Loss: 3.3430, Train Acc: 52.50%, Test Loss: 3.0313, Test Acc: 56.18%, Time: 119.25s
Epoch 3/3, Train Loss: 2.7109, Train Acc: 59.08%, Test Loss: 2.5002, Test Acc: 60.30%, Time: 119.20s
Creating Swin-Tiny model from scratch...
Running experiment with Swin-Tiny (From Scratch)
Number of trainable parameters: 27,596,254
Epoch 1/3, Train Loss: 4.0350, Train Acc: 8.38%, Test Loss: 3.6695, Test Acc: 13.92%, Time: 203.16s
Epoch 2/3, Train Loss: 3.4967, Train Acc: 16.46%, Test Loss: 3.2916, Test Acc: 20.07%, Time: 202.86s
Epoch 3/3, Train Loss: 3.1466, Train Acc: 22.77%, Test Loss: 2.9846, Test Acc: 26.15%, Time: 204.45s
```

```
Results Summary:
--------------------------------------------------------------------------------
Model                      | Trainable Params   | Avg Train Time  | Final Test Acc
--------------------------------------------------------------------------------
Swin-Tiny (Fine-tuned)     | 76,900 | 81.14s | 57.45%
Swin-Small (Fine-tuned)    | 76,900 | 119.40s | 60.30%
Swin-Tiny (From Scratch)   | 27,596,254 | 203.49s | 26.15%
```



Our fine-tuning experiments with Swin Transformer models on CIFAR-100 demonstrate the compelling advantages of transfer learning. Fine-tuned models achieve substantially higher accuracy (57-60% vs 26%) compared to training from scratch, while requiring only ~77K trainable parameters versus 27.6M for the scratch model dramatic efficiency gain. This performance gap stems primarily from pretrained models' exposure to millions of diverse images during pretraining, allowing them to develop robust feature extractors that generalize well to new domains. Additionally, pretrained models benefit from sophisticated architectural design choices already optimized through extensive experimentation on larger datasets. The superior performance extends across model architectures, with Swin-Small outperforming Swin-Tiny by approximately 3 percentage points despite having identical trainable parameter counts, indicating that larger pretrained backbones extract more discriminative features even when frozen. The practical benefits extend to training speed, with fine-tuned models completing epochs significantly faster than the from-scratch counterpart, making them ideal for rapid adaptation to new domains. However, pretrained models might underperform when target datasets differ drastically from pretraining data (e.g., medical images vs. natural images), when resolution mismatches cause information loss (as potentially occurred with our 32×32 to 224×224 upsampling), or when the frozen feature extractors are suboptimal for the specific classification boundaries needed. The learning curves further illuminate the advantages, showing pretrained models beginning at higher accuracy levels and improving quickly, while the scratch model learns gradually. These results convincingly demonstrate that for modest datasets like CIFAR-100, leveraging pretrained transformer models through fine-tuning delivers superior performance with substantially reduced computational requirements.

Andrew Nicola 801136465 HW5

Github: https://github.com/Eskdagoat/4106/blob/main/NicolaAndrew_801136465_HW6.ipynb