Using device: cuda

=== Gemini V3: Advanced CNN Experiments for CIFAR-10 ===

=== Running Experiment: SGD_StepLR_Baseline ===
Config: {'model_name': 'AdvancedCNN', 'use_gap': True, 'dropout_rate': 0.3, 'initial_channels': 64, 'block_config': [(64, 2, 1), (128, 2, 2), (256, 2, 2), (512, 2, 2)], 'optimizer': 'sgd', 'learning_rate': 0.1, 'momentum': 0.9, 'weight_decay': 0.0005, 'use_nesterov': True, 'scheduler': 'step', 'scheduler_step_size': 15, 'scheduler_gamma': 0.1, 'epochs': 50, 'batch_size': 128, 'use_augmentation': True, 'use_rotation': True, 'use_colorjitter': True, 'use_cutmix': False, 'use_mixup': False, 'label_smoothing': 0.0, 'warmup_epochs': 0}
Model: AdvancedCNN
Total parameters: 11,172,170
Using SGD optimizer with LR=0.1, Momentum=0.9, WD=0.0005, Nesterov=True
Using StepLR scheduler with step_size=15, gamma=0.1.
[Epoch: 1/50] Loss: 1.5341, Acc: 44.02%, LR: 0.100000, Time: 12.67s
[Epoch: 2/50] Loss: 1.0871, Acc: 61.32%, LR: 0.100000, Time: 12.51s
[Epoch: 3/50] Loss: 0.9073, Acc: 67.85%, LR: 0.100000, Time: 12.53s
[Epoch: 4/50] Loss: 0.8061, Acc: 71.88%, LR: 0.100000, Time: 12.53s
[Epoch: 5/50] Loss: 0.7481, Acc: 73.81%, LR: 0.100000, Time: 12.47s
[Epoch: 6/50] Loss: 0.7110, Acc: 75.43%, LR: 0.100000, Time: 12.53s
[Epoch: 7/50] Loss: 0.6801, Acc: 76.27%, LR: 0.100000, Time: 12.47s
[Epoch: 8/50] Loss: 0.6608, Acc: 77.16%, LR: 0.100000, Time: 12.61s
[Epoch: 9/50] Loss: 0.6412, Acc: 77.71%, LR: 0.100000, Time: 12.68s
[Epoch: 10/50] Loss: 0.6352, Acc: 78.08%, LR: 0.100000, Time: 12.71s
[Epoch: 11/50] Loss: 0.6169, Acc: 78.56%, LR: 0.100000, Time: 12.43s
[Epoch: 12/50] Loss: 0.6057, Acc: 78.99%, LR: 0.100000, Time: 12.43s
[Epoch: 13/50] Loss: 0.5944, Acc: 79.68%, LR: 0.100000, Time: 12.49s
[Epoch: 14/50] Loss: 0.5880, Acc: 79.76%, LR: 0.100000, Time: 12.62s
[Epoch: 15/50] Loss: 0.5810, Acc: 80.00%, LR: 0.100000, Time: 12.55s
[Epoch: 16/50] Loss: 0.4317, Acc: 85.19%, LR: 0.010000, Time: 12.57s
[Epoch: 17/50] Loss: 0.3822, Acc: 86.78%, LR: 0.010000, Time: 12.72s
[Epoch: 18/50] Loss: 0.3621, Acc: 87.41%, LR: 0.010000, Time: 12.55s
[Epoch: 19/50] Loss: 0.3504, Acc: 87.88%, LR: 0.010000, Time: 12.46s
[Epoch: 20/50] Loss: 0.3325, Acc: 88.61%, LR: 0.010000, Time: 12.86s
[Epoch: 21/50] Loss: 0.3256, Acc: 88.60%, LR: 0.010000, Time: 12.75s
[Epoch: 22/50] Loss: 0.3166, Acc: 89.00%, LR: 0.010000, Time: 12.82s
[Epoch: 23/50] Loss: 0.3090, Acc: 89.25%, LR: 0.010000, Time: 12.85s
[Epoch: 24/50] Loss: 0.3035, Acc: 89.41%, LR: 0.010000, Time: 12.94s
[Epoch: 25/50] Loss: 0.2976, Acc: 89.73%, LR: 0.010000, Time: 13.05s
[Epoch: 26/50] Loss: 0.2887, Acc: 90.02%, LR: 0.010000, Time: 12.89s

```
[Epoch: 27/50] Loss: 0.2875, Acc: 90.04%, LR: 0.010000, Time: 13.19s
[Epoch: 28/50] Loss: 0.2810, Acc: 90.10%, LR: 0.010000, Time: 13.02s
[Epoch: 29/50] Loss: 0.2770, Acc: 90.49%, LR: 0.010000, Time: 12.47s
[Epoch: 30/50] Loss: 0.2762, Acc: 90.53%, LR: 0.010000, Time: 12.68s
[Epoch: 31/50] Loss: 0.2374, Acc: 91.72%, LR: 0.001000, Time: 12.53s
[Epoch: 32/50] Loss: 0.2249, Acc: 92.11%, LR: 0.001000, Time: 12.57s
[Epoch: 33/50] Loss: 0.2168, Acc: 92.42%, LR: 0.001000, Time: 12.79s
[Epoch: 34/50] Loss: 0.2127, Acc: 92.53%, LR: 0.001000, Time: 12.55s
[Epoch: 35/50] Loss: 0.2107, Acc: 92.78%, LR: 0.001000, Time: 12.58s
[Epoch: 36/50] Loss: 0.2057, Acc: 92.82%, LR: 0.001000, Time: 12.48s
[Epoch: 37/50] Loss: 0.2028, Acc: 92.88%, LR: 0.001000, Time: 12.46s
[Epoch: 38/50] Loss: 0.2028, Acc: 92.90%, LR: 0.001000, Time: 12.65s
[Epoch: 39/50] Loss: 0.2015, Acc: 92.90%, LR: 0.001000, Time: 12.60s
[Epoch: 40/50] Loss: 0.1992, Acc: 93.21%, LR: 0.001000, Time: 12.72s
[Epoch: 41/50] Loss: 0.1949, Acc: 93.22%, LR: 0.001000, Time: 12.59s
[Epoch: 42/50] Loss: 0.1948, Acc: 93.18%, LR: 0.001000, Time: 12.52s
[Epoch: 43/50] Loss: 0.1921, Acc: 93.19%, LR: 0.001000, Time: 12.75s
[Epoch: 44/50] Loss: 0.1868, Acc: 93.47%, LR: 0.001000, Time: 12.49s
[Epoch: 45/50] Loss: 0.1857, Acc: 93.60%, LR: 0.001000, Time: 12.54s
[Epoch: 46/50] Loss: 0.1813, Acc: 93.64%, LR: 0.000100, Time: 12.52s
[Epoch: 47/50] Loss: 0.1818, Acc: 93.71%, LR: 0.000100, Time: 12.65s
[Epoch: 48/50] Loss: 0.1857, Acc: 93.50%, LR: 0.000100, Time: 12.67s
[Epoch: 49/50] Loss: 0.1803, Acc: 93.72%, LR: 0.000100, Time: 12.48s
[Epoch: 50/50] Loss: 0.1830, Acc: 93.71%, LR: 0.000100, Time: 12.46s
Test Set Accuracy: 92.24%


=== Running Experiment: AdamW_Cosine_AdvAug ===
Config: {'model_name': 'AdvancedCNN', 'use_gap': True, 'dropout_rate': 0.4,
'initial_channels': 64, 'block_config': [(64, 2, 1), (128, 2, 2), (256, 2,
2), (512, 2, 2)], 'optimizer': 'adamw', 'learning_rate': 0.001,
'weight_decay': 0.01, 'scheduler': 'cosine', 'epochs': 75, 'batch_size':
128, 'use_augmentation': True, 'use_rotation': True, 'use_colorjitter':
True, 'use_randaugment': True, 'use_cutmix': True, 'cutmix_alpha': 1.0,
'use_mixup': False, 'label_smoothing': 0.1, 'warmup_epochs': 5}
Model: AdvancedCNN
Total parameters: 11,172,170
Using Label Smoothing: 0.1
Using AdamW optimizer with LR=0.001, WD=0.01
Using CosineAnnealingLR scheduler for 70 epochs.
Applying 5 epochs of linear warmup.
/usr/local/lib/python3.11/dist-packages/torch/optim/lr_scheduler.py:243:
UserWarning: The epoch parameter in `scheduler.step()` was not necessary
and is being deprecated where possible. Please use `scheduler.step()` to
```

step the scheduler. During the deprecation, if epoch is different from
None, the closed form is used instead of the new chainable form, where
available. Please open an issue if you are unable to replicate your use
case: https://github.com/pytorch/pytorch/issues/new/choose.
  warnings.warn(EPOCH_DEPRECATION_WARNING, UserWarning)
[Epoch: 1/75] Loss: 2.0059, Acc: 28.96%, LR: 0.000522, Time: 17.27s
[Epoch: 2/75] Loss: 1.8317, Acc: 39.57%, LR: 0.000024, Time: 17.41s
[Epoch: 3/75] Loss: 1.7177, Acc: 45.40%, LR: 0.000225, Time: 17.37s
[Epoch: 4/75] Loss: 1.6391, Acc: 50.26%, LR: 0.000829, Time: 17.60s
[Epoch: 5/75] Loss: 1.5952, Acc: 52.74%, LR: 0.000950, Time: 17.39s
[Epoch: 6/75] Loss: 1.5423, Acc: 55.35%, LR: 0.000411, Time: 17.74s
[Epoch: 7/75] Loss: 1.5068, Acc: 56.53%, LR: 0.000002, Time: 17.50s
[Epoch: 8/75] Loss: 1.4532, Acc: 59.50%, LR: 0.000324, Time: 17.46s
[Epoch: 9/75] Loss: 1.4414, Acc: 59.69%, LR: 0.000905, Time: 17.42s
[Epoch: 10/75] Loss: 1.4241, Acc: 60.21%, LR: 0.000891, Time: 17.56s
[Epoch: 11/75] Loss: 1.3918, Acc: 62.76%, LR: 0.000303, Time: 17.33s
[Epoch: 12/75] Loss: 1.3907, Acc: 62.93%, LR: 0.000005, Time: 17.47s
[Epoch: 13/75] Loss: 1.3681, Acc: 63.22%, LR: 0.000433, Time: 17.29s
[Epoch: 14/75] Loss: 1.3564, Acc: 63.73%, LR: 0.000960, Time: 17.55s
[Epoch: 15/75] Loss: 1.3341, Acc: 65.32%, LR: 0.000812, Time: 17.26s
[Epoch: 16/75] Loss: 1.3320, Acc: 64.82%, LR: 0.000206, Time: 17.35s
[Epoch: 17/75] Loss: 1.3156, Acc: 65.81%, LR: 0.000032, Time: 17.25s
[Epoch: 18/75] Loss: 1.2986, Acc: 66.80%, LR: 0.000545, Time: 17.42s
[Epoch: 19/75] Loss: 1.3123, Acc: 66.57%, LR: 0.000992, Time: 17.29s
[Epoch: 20/75] Loss: 1.2989, Acc: 66.81%, LR: 0.000717, Time: 17.58s
[Epoch: 21/75] Loss: 1.2903, Acc: 66.71%, LR: 0.000123, Time: 17.23s
[Epoch: 22/75] Loss: 1.2678, Acc: 68.45%, LR: 0.000083, Time: 17.42s
[Epoch: 23/75] Loss: 1.2521, Acc: 68.47%, LR: 0.000655, Time: 17.42s
[Epoch: 24/75] Loss: 1.2609, Acc: 68.53%, LR: 0.000999, Time: 17.36s
[Epoch: 25/75] Loss: 1.2611, Acc: 68.56%, LR: 0.000611, Time: 17.36s
[Epoch: 26/75] Loss: 1.2397, Acc: 69.22%, LR: 0.000060, Time: 17.50s
[Epoch: 27/75] Loss: 1.2358, Acc: 69.58%, LR: 0.000154, Time: 17.47s
[Epoch: 28/75] Loss: 1.2132, Acc: 70.42%, LR: 0.000756, Time: 17.44s
[Epoch: 29/75] Loss: 1.2341, Acc: 69.73%, LR: 0.000982, Time: 17.25s
[Epoch: 30/75] Loss: 1.2143, Acc: 70.46%, LR: 0.000500, Time: 17.50s
[Epoch: 31/75] Loss: 1.2284, Acc: 70.35%, LR: 0.000018, Time: 17.51s
[Epoch: 32/75] Loss: 1.1951, Acc: 70.93%, LR: 0.000244, Time: 17.46s
[Epoch: 33/75] Loss: 1.2026, Acc: 70.96%, LR: 0.000846, Time: 17.45s
[Epoch: 34/75] Loss: 1.1826, Acc: 71.47%, LR: 0.000940, Time: 17.49s
[Epoch: 35/75] Loss: 1.2134, Acc: 70.34%, LR: 0.000389, Time: 17.35s
[Epoch: 36/75] Loss: 1.1839, Acc: 72.47%, LR: 0.000001, Time: 17.49s
[Epoch: 37/75] Loss: 1.1946, Acc: 71.07%, LR: 0.000345, Time: 17.52s
[Epoch: 38/75] Loss: 1.1760, Acc: 72.26%, LR: 0.000917, Time: 17.39s

```
[Epoch: 39/75] Loss: 1.1877, Acc: 71.36%, LR: 0.000877, Time: 17.50s
[Epoch: 40/75] Loss: 1.2002, Acc: 71.48%, LR: 0.000283, Time: 17.31s
[Epoch: 41/75] Loss: 1.1771, Acc: 72.67%, LR: 0.000008, Time: 17.61s
[Epoch: 42/75] Loss: 1.1622, Acc: 72.74%, LR: 0.000455, Time: 17.22s
[Epoch: 43/75] Loss: 1.1668, Acc: 73.11%, LR: 0.000968, Time: 17.60s
[Epoch: 44/75] Loss: 1.1320, Acc: 74.33%, LR: 0.000794, Time: 17.58s
[Epoch: 45/75] Loss: 1.1599, Acc: 73.42%, LR: 0.000188, Time: 17.41s
[Epoch: 46/75] Loss: 1.1366, Acc: 74.57%, LR: 0.000040, Time: 17.29s
[Epoch: 47/75] Loss: 1.1389, Acc: 73.93%, LR: 0.000567, Time: 17.48s
[Epoch: 48/75] Loss: 1.1061, Acc: 75.73%, LR: 0.000995, Time: 17.20s
[Epoch: 49/75] Loss: 1.1308, Acc: 73.78%, LR: 0.000697, Time: 17.53s
[Epoch: 50/75] Loss: 1.1576, Acc: 72.79%, LR: 0.000109, Time: 17.35s
[Epoch: 51/75] Loss: 1.1263, Acc: 74.36%, LR: 0.000095, Time: 17.74s
[Epoch: 52/75] Loss: 1.1184, Acc: 74.97%, LR: 0.000676, Time: 17.42s
[Epoch: 53/75] Loss: 1.0888, Acc: 76.16%, LR: 0.000998, Time: 17.37s
[Epoch: 54/75] Loss: 1.1257, Acc: 74.57%, LR: 0.000589, Time: 17.37s
[Epoch: 55/75] Loss: 1.1301, Acc: 74.73%, LR: 0.000050, Time: 17.60s
[Epoch: 56/75] Loss: 1.1423, Acc: 73.46%, LR: 0.000171, Time: 17.58s
[Epoch: 57/75] Loss: 1.0927, Acc: 75.70%, LR: 0.000775, Time: 17.45s
[Epoch: 58/75] Loss: 1.0955, Acc: 75.96%, LR: 0.000976, Time: 17.37s
[Epoch: 59/75] Loss: 1.0951, Acc: 75.80%, LR: 0.000478, Time: 17.71s
[Epoch: 60/75] Loss: 1.1138, Acc: 75.34%, LR: 0.000013, Time: 17.78s
[Epoch: 61/75] Loss: 1.1020, Acc: 75.69%, LR: 0.000263, Time: 17.43s
[Epoch: 62/75] Loss: 1.0918, Acc: 76.21%, LR: 0.000861, Time: 17.45s
[Epoch: 63/75] Loss: 1.0935, Acc: 75.69%, LR: 0.000929, Time: 17.29s
[Epoch: 64/75] Loss: 1.1067, Acc: 75.30%, LR: 0.000367, Time: 17.54s
[Epoch: 65/75] Loss: 1.0963, Acc: 76.10%, LR: 0.000000, Time: 17.40s
[Epoch: 66/75] Loss: 1.1398, Acc: 73.38%, LR: 0.000367, Time: 17.56s
[Epoch: 67/75] Loss: 1.1058, Acc: 74.78%, LR: 0.000929, Time: 17.42s
[Epoch: 68/75] Loss: 1.0830, Acc: 76.20%, LR: 0.000861, Time: 17.68s
[Epoch: 69/75] Loss: 1.1113, Acc: 74.64%, LR: 0.000263, Time: 17.60s
[Epoch: 70/75] Loss: 1.0728, Acc: 76.45%, LR: 0.000013, Time: 17.53s
[Epoch: 71/75] Loss: 1.0898, Acc: 75.81%, LR: 0.000478, Time: 17.45s
[Epoch: 72/75] Loss: 1.1093, Acc: 75.61%, LR: 0.000976, Time: 17.40s
[Epoch: 73/75] Loss: 1.0677, Acc: 76.95%, LR: 0.000775, Time: 17.33s
[Epoch: 74/75] Loss: 1.1083, Acc: 75.10%, LR: 0.000171, Time: 17.41s
[Epoch: 75/75] Loss: 1.0807, Acc: 76.97%, LR: 0.000050, Time: 17.54s
Test Set Accuracy: 90.58%

=== Running Experiment: SGD_OneCycle_AdvAug ===
Config: {'model_name': 'AdvancedCNN', 'use_gap': True, 'dropout_rate': 0.4,
'initial_channels': 64, 'block_config': [(64, 2, 1), (128, 2, 2), (256, 2,
2), (512, 2, 2)], 'optimizer': 'sgd', 'learning_rate': 0.01, 'max_lr': 0.1,
```

'momentum': 0.9, 'weight_decay': 0.0005, 'use_nesterov': True, 'scheduler': 'onecycle', 'epochs': 50, 'batch_size': 128, 'use_augmentation': True, 'use_rotation': True, 'use_colorjitter': True, 'use_trivialaugment': True, 'use_cutmix': False, 'use_mixup': True, 'mixup_alpha': 0.8, 'label_smoothing': 0.1, 'warmup_epochs': 0}
Model: AdvancedCNN
Total parameters: 11,172,170
Using Label Smoothing: 0.1
Using SGD optimizer with LR=0.01, Momentum=0.9, WD=0.0005, Nesterov=True
Using OneCycleLR scheduler with max_lr=0.1.
[Epoch: 1/50] Loss: 2.1233, Acc: 23.34%, LR: 0.005049, Time: 16.43s
[Epoch: 2/50] Loss: 1.9429, Acc: 33.60%, LR: 0.008151, Time: 16.33s
[Epoch: 3/50] Loss: 1.8182, Acc: 41.10%, LR: 0.013170, Time: 16.36s
[Epoch: 4/50] Loss: 1.7384, Acc: 45.59%, LR: 0.019887, Time: 16.34s
[Epoch: 5/50] Loss: 1.6823, Acc: 48.81%, LR: 0.028007, Time: 16.20s
[Epoch: 6/50] Loss: 1.6518, Acc: 50.69%, LR: 0.037177, Time: 16.30s
[Epoch: 7/50] Loss: 1.6014, Acc: 52.92%, LR: 0.046995, Time: 16.16s
[Epoch: 8/50] Loss: 1.5745, Acc: 54.13%, LR: 0.057031, Time: 16.29s
[Epoch: 9/50] Loss: 1.5391, Acc: 55.81%, LR: 0.066847, Time: 16.31s
[Epoch: 10/50] Loss: 1.4939, Acc: 57.85%, LR: 0.076015, Time: 16.26s
[Epoch: 11/50] Loss: 1.4927, Acc: 58.24%, LR: 0.084132, Time: 16.18s
[Epoch: 12/50] Loss: 1.4633, Acc: 59.18%, LR: 0.090845, Time: 16.42s
[Epoch: 13/50] Loss: 1.4583, Acc: 60.15%, LR: 0.095859, Time: 16.27s
[Epoch: 14/50] Loss: 1.4367, Acc: 60.65%, LR: 0.098956, Time: 16.18s
[Epoch: 15/50] Loss: 1.4509, Acc: 59.69%, LR: 0.100000, Time: 16.24s
[Epoch: 16/50] Loss: 1.4227, Acc: 60.92%, LR: 0.099798, Time: 16.53s
[Epoch: 17/50] Loss: 1.4339, Acc: 60.97%, LR: 0.099194, Time: 16.18s
[Epoch: 18/50] Loss: 1.4087, Acc: 61.74%, LR: 0.098195, Time: 16.29s
[Epoch: 19/50] Loss: 1.4099, Acc: 61.32%, LR: 0.096808, Time: 16.37s
[Epoch: 20/50] Loss: 1.4128, Acc: 61.61%, LR: 0.095043, Time: 16.52s
[Epoch: 21/50] Loss: 1.3981, Acc: 61.99%, LR: 0.092917, Time: 16.28s
[Epoch: 22/50] Loss: 1.3865, Acc: 62.39%, LR: 0.090444, Time: 16.23s
[Epoch: 23/50] Loss: 1.3950, Acc: 62.84%, LR: 0.087646, Time: 16.35s
[Epoch: 24/50] Loss: 1.3561, Acc: 63.58%, LR: 0.084545, Time: 16.12s
[Epoch: 25/50] Loss: 1.3572, Acc: 63.71%, LR: 0.081166, Time: 16.40s
[Epoch: 26/50] Loss: 1.3582, Acc: 63.70%, LR: 0.077535, Time: 16.36s
[Epoch: 27/50] Loss: 1.3987, Acc: 62.76%, LR: 0.073683, Time: 16.36s
[Epoch: 28/50] Loss: 1.3716, Acc: 63.32%, LR: 0.069641, Time: 16.32s
[Epoch: 29/50] Loss: 1.3904, Acc: 62.81%, LR: 0.065440, Time: 16.30s
[Epoch: 30/50] Loss: 1.3649, Acc: 63.87%, LR: 0.061115, Time: 16.45s
[Epoch: 31/50] Loss: 1.3335, Acc: 65.27%, LR: 0.056700, Time: 16.18s
[Epoch: 32/50] Loss: 1.3575, Acc: 64.71%, LR: 0.052232, Time: 16.47s
[Epoch: 33/50] Loss: 1.3308, Acc: 65.16%, LR: 0.047746, Time: 16.19s

```
[Epoch: 34/50] Loss: 1.3160, Acc: 66.15%, LR: 0.043277, Time: 16.32s
[Epoch: 35/50] Loss: 1.3301, Acc: 65.60%, LR: 0.038863, Time: 16.32s
[Epoch: 36/50] Loss: 1.3335, Acc: 65.12%, LR: 0.034538, Time: 16.35s
[Epoch: 37/50] Loss: 1.3191, Acc: 65.79%, LR: 0.030338, Time: 16.23s
[Epoch: 38/50] Loss: 1.2869, Acc: 66.99%, LR: 0.026297, Time: 16.27s
[Epoch: 39/50] Loss: 1.2810, Acc: 67.45%, LR: 0.022446, Time: 16.35s
[Epoch: 40/50] Loss: 1.2919, Acc: 67.43%, LR: 0.018817, Time: 16.37s
[Epoch: 41/50] Loss: 1.2705, Acc: 67.82%, LR: 0.015439, Time: 16.40s
[Epoch: 42/50] Loss: 1.2359, Acc: 69.67%, LR: 0.012339, Time: 16.33s
[Epoch: 43/50] Loss: 1.2039, Acc: 70.67%, LR: 0.009543, Time: 16.13s
[Epoch: 44/50] Loss: 1.2541, Acc: 69.51%, LR: 0.007072, Time: 16.26s
[Epoch: 45/50] Loss: 1.2156, Acc: 71.09%, LR: 0.004947, Time: 16.21s
[Epoch: 46/50] Loss: 1.1684, Acc: 72.67%, LR: 0.003185, Time: 16.26s
[Epoch: 47/50] Loss: 1.1605, Acc: 73.19%, LR: 0.001799, Time: 16.52s
[Epoch: 48/50] Loss: 1.1782, Acc: 73.15%, LR: 0.000802, Time: 16.19s
[Epoch: 49/50] Loss: 1.1447, Acc: 73.78%, LR: 0.000201, Time: 16.44s
[Epoch: 50/50] Loss: 1.1591, Acc: 73.47%, LR: 0.000000, Time: 16.20s
Test Set Accuracy: 90.58%

=== Running Experiment: AdamW_Cosine_WiderNet ===
Config: {'model_name': 'AdvancedCNN', 'use_gap': True, 'dropout_rate': 0.5,
'initial_channels': 96, 'block_config': [(96, 2, 1), (192, 2, 2), (384, 2,
2), (768, 2, 2)], 'optimizer': 'adamw', 'learning_rate': 0.001,
'weight_decay': 0.01, 'scheduler': 'cosine', 'epochs': 75, 'batch_size':
128, 'use_augmentation': True, 'use_rotation': True, 'use_colorjitter':
True, 'use_randaugment': True, 'use_cutmix': True, 'cutmix_alpha': 1.0,
'label_smoothing': 0.1, 'warmup_epochs': 5}
Model: AdvancedCNN
Total parameters: 25,126,378
Using Label Smoothing: 0.1
Using AdamW optimizer with LR=0.001, WD=0.01
Using CosineAnnealingLR scheduler for 70 epochs.
Applying 5 epochs of linear warmup.
/usr/local/lib/python3.11/dist-packages/torch/optim/lr_scheduler.py:243:
UserWarning: The epoch parameter in `scheduler.step()` was not necessary
and is being deprecated where possible. Please use `scheduler.step()` to
step the scheduler. During the deprecation, if epoch is different from
None, the closed form is used instead of the new chainable form, where
available. Please open an issue if you are unable to replicate your use
case: https://github.com/pytorch/pytorch/issues/new/choose.
  warnings.warn(EPOCH_DEPRECATION_WARNING, UserWarning)
[Epoch: 1/75] Loss: 2.0364, Acc: 27.31%, LR: 0.000522, Time: 17.43s
[Epoch: 2/75] Loss: 1.8597, Acc: 37.74%, LR: 0.000024, Time: 17.22s
```
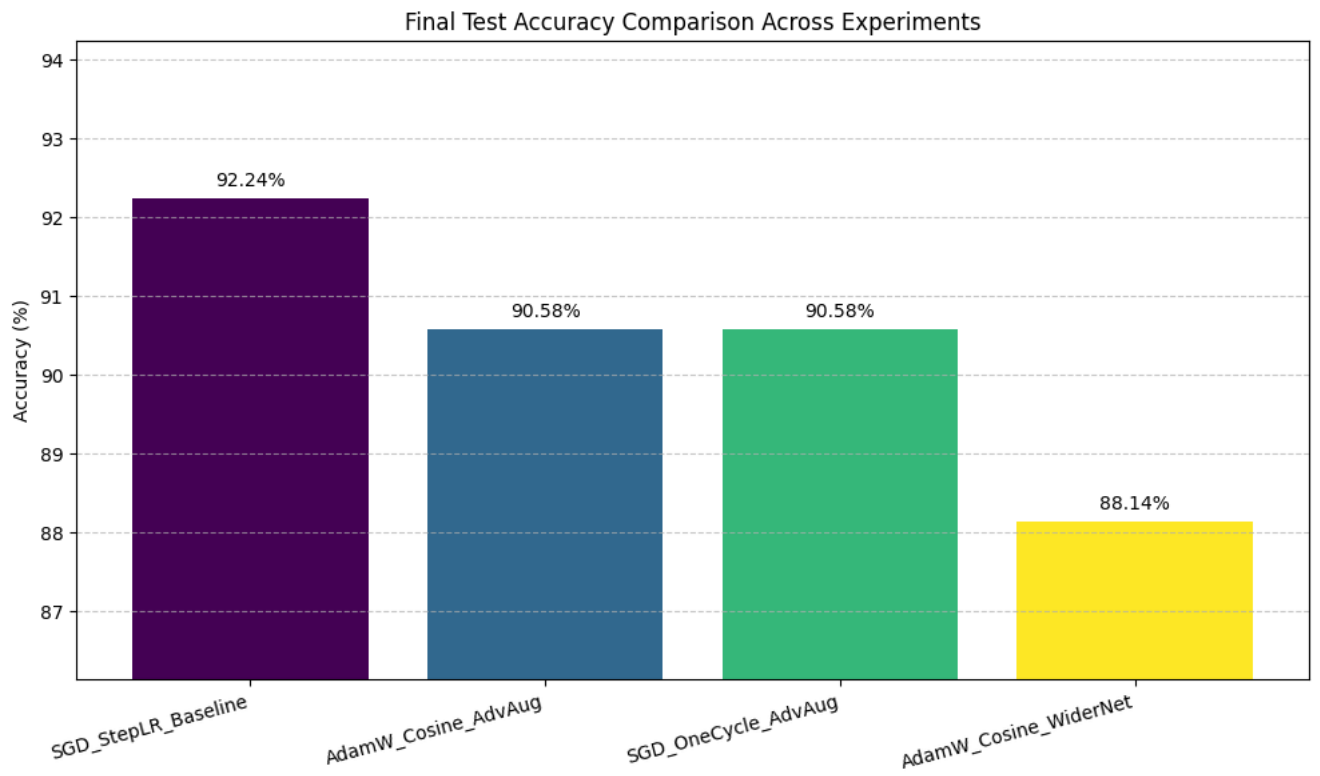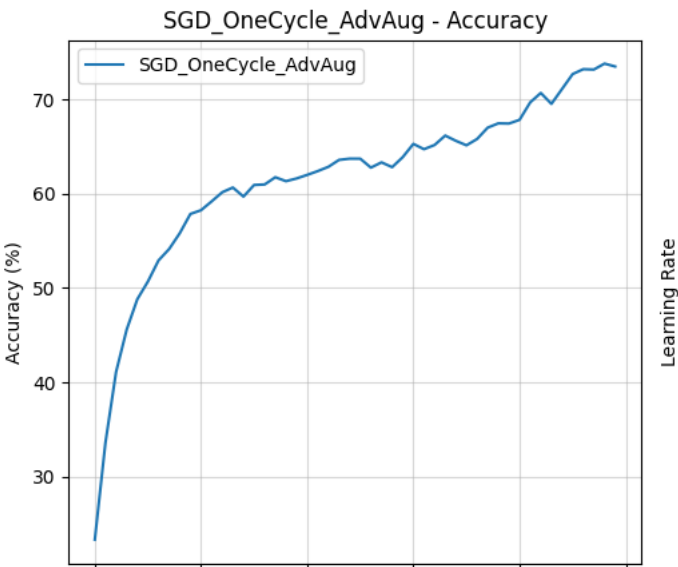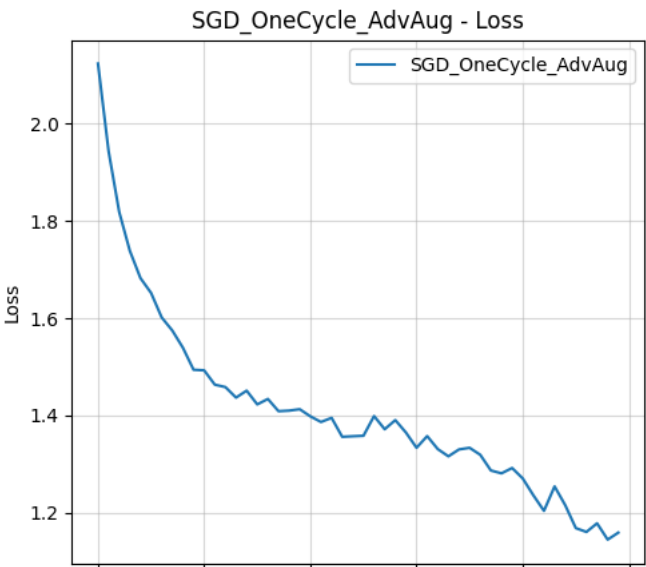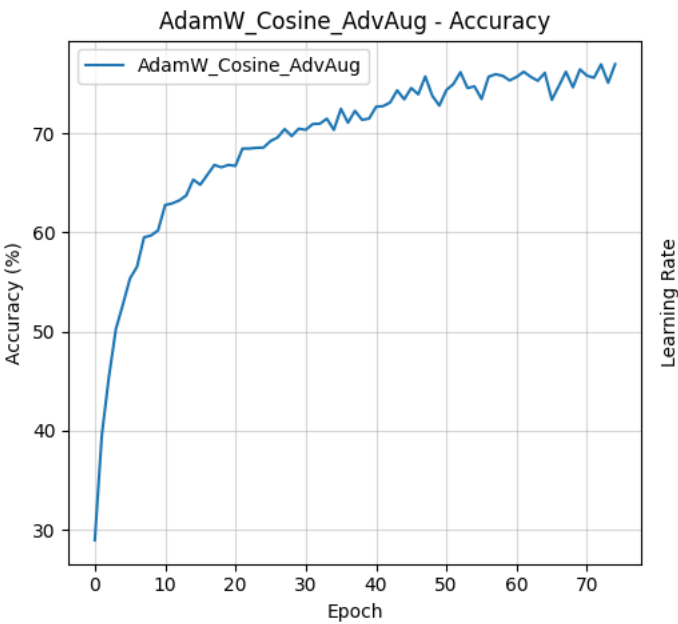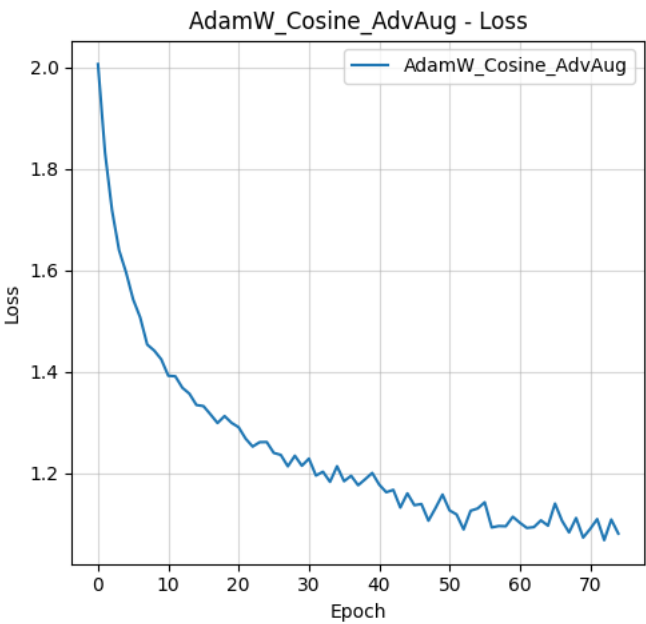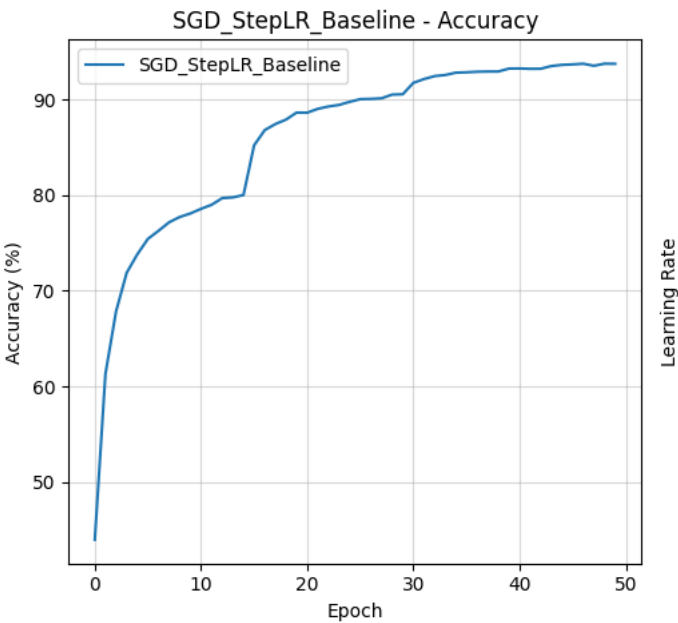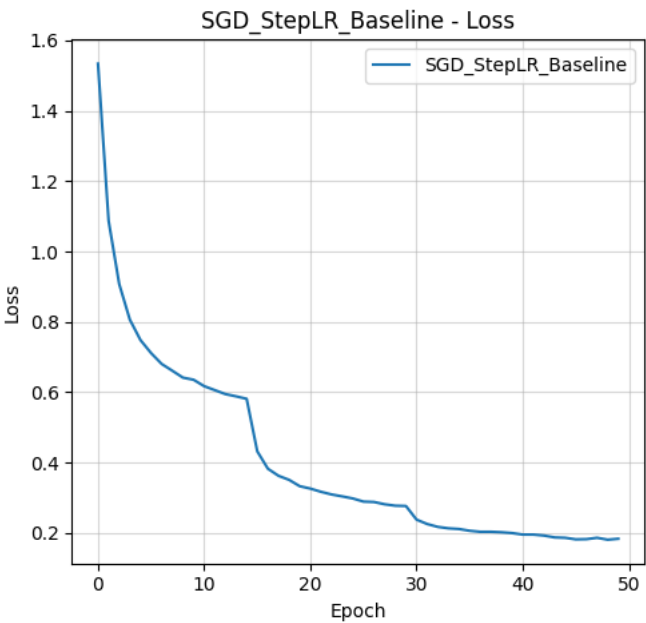
```
[Epoch: 3/75] Loss: 1.7796, Acc: 42.26%, LR: 0.000225, Time: 17.58s
[Epoch: 4/75] Loss: 1.6890, Acc: 47.47%, LR: 0.000829, Time: 17.52s
[Epoch: 5/75] Loss: 1.6416, Acc: 49.67%, LR: 0.000950, Time: 17.54s
[Epoch: 6/75] Loss: 1.6256, Acc: 49.98%, LR: 0.000411, Time: 17.46s
[Epoch: 7/75] Loss: 1.5609, Acc: 54.00%, LR: 0.000002, Time: 17.79s
[Epoch: 8/75] Loss: 1.5446, Acc: 55.06%, LR: 0.000324, Time: 17.49s
[Epoch: 9/75] Loss: 1.4834, Acc: 57.81%, LR: 0.000905, Time: 17.55s
[Epoch: 10/75] Loss: 1.4364, Acc: 60.00%, LR: 0.000891, Time: 17.64s
[Epoch: 11/75] Loss: 1.4370, Acc: 60.43%, LR: 0.000303, Time: 17.34s
[Epoch: 12/75] Loss: 1.4289, Acc: 61.11%, LR: 0.000005, Time: 17.66s
[Epoch: 13/75] Loss: 1.3934, Acc: 62.47%, LR: 0.000433, Time: 17.49s
[Epoch: 14/75] Loss: 1.3805, Acc: 62.95%, LR: 0.000960, Time: 17.32s
[Epoch: 15/75] Loss: 1.3611, Acc: 63.63%, LR: 0.000812, Time: 17.48s
[Epoch: 16/75] Loss: 1.3588, Acc: 64.41%, LR: 0.000206, Time: 17.42s
[Epoch: 17/75] Loss: 1.3336, Acc: 65.63%, LR: 0.000032, Time: 17.48s
[Epoch: 18/75] Loss: 1.3431, Acc: 64.86%, LR: 0.000545, Time: 17.44s
[Epoch: 19/75] Loss: 1.2966, Acc: 66.14%, LR: 0.000992, Time: 17.48s
[Epoch: 20/75] Loss: 1.2989, Acc: 66.39%, LR: 0.000717, Time: 17.43s
[Epoch: 21/75] Loss: 1.3078, Acc: 66.33%, LR: 0.000123, Time: 17.45s
[Epoch: 22/75] Loss: 1.2892, Acc: 66.76%, LR: 0.000083, Time: 17.52s
[Epoch: 23/75] Loss: 1.2747, Acc: 67.12%, LR: 0.000655, Time: 17.32s
[Epoch: 24/75] Loss: 1.2720, Acc: 67.88%, LR: 0.000999, Time: 17.40s
[Epoch: 25/75] Loss: 1.2836, Acc: 67.36%, LR: 0.000611, Time: 17.50s
[Epoch: 26/75] Loss: 1.2659, Acc: 68.58%, LR: 0.000060, Time: 17.47s
[Epoch: 27/75] Loss: 1.2539, Acc: 68.82%, LR: 0.000154, Time: 17.57s
[Epoch: 28/75] Loss: 1.2313, Acc: 69.54%, LR: 0.000756, Time: 17.47s
[Epoch: 29/75] Loss: 1.2401, Acc: 69.65%, LR: 0.000982, Time: 17.46s
[Epoch: 30/75] Loss: 1.2679, Acc: 68.04%, LR: 0.000500, Time: 17.35s
[Epoch: 31/75] Loss: 1.2251, Acc: 70.28%, LR: 0.000018, Time: 17.64s
[Epoch: 32/75] Loss: 1.2374, Acc: 68.78%, LR: 0.000244, Time: 17.50s
[Epoch: 33/75] Loss: 1.2193, Acc: 70.82%, LR: 0.000846, Time: 17.41s
[Epoch: 34/75] Loss: 1.2212, Acc: 70.36%, LR: 0.000940, Time: 17.61s
[Epoch: 35/75] Loss: 1.1994, Acc: 71.48%, LR: 0.000389, Time: 17.58s
[Epoch: 36/75] Loss: 1.1998, Acc: 71.75%, LR: 0.000001, Time: 17.64s
[Epoch: 37/75] Loss: 1.2061, Acc: 71.02%, LR: 0.000345, Time: 17.51s
[Epoch: 38/75] Loss: 1.1600, Acc: 72.97%, LR: 0.000917, Time: 17.54s
[Epoch: 39/75] Loss: 1.1598, Acc: 73.25%, LR: 0.000877, Time: 17.45s
[Epoch: 40/75] Loss: 1.2172, Acc: 70.63%, LR: 0.000283, Time: 17.51s
[Epoch: 41/75] Loss: 1.1971, Acc: 71.63%, LR: 0.000008, Time: 17.37s
[Epoch: 42/75] Loss: 1.1650, Acc: 72.94%, LR: 0.000455, Time: 17.57s
[Epoch: 43/75] Loss: 1.1760, Acc: 72.27%, LR: 0.000968, Time: 17.54s
[Epoch: 44/75] Loss: 1.1516, Acc: 73.19%, LR: 0.000794, Time: 17.57s
[Epoch: 45/75] Loss: 1.1488, Acc: 73.57%, LR: 0.000188, Time: 17.64s
```
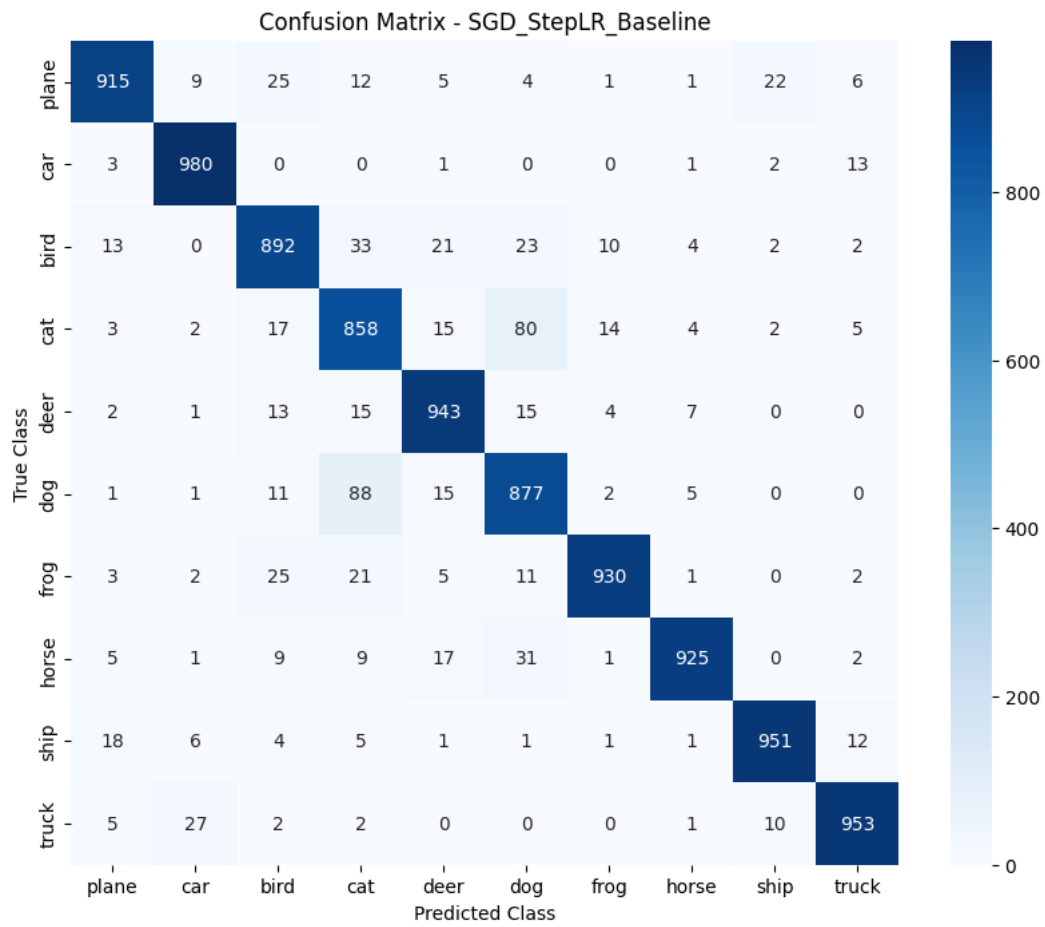
```
[Epoch: 46/75] Loss: 1.1608, Acc: 73.40%, LR: 0.000040, Time: 17.71s
[Epoch: 47/75] Loss: 1.1635, Acc: 73.02%, LR: 0.000567, Time: 17.44s
[Epoch: 48/75] Loss: 1.1488, Acc: 73.59%, LR: 0.000995, Time: 17.64s
[Epoch: 49/75] Loss: 1.1284, Acc: 74.45%, LR: 0.000697, Time: 17.53s
[Epoch: 50/75] Loss: 1.1510, Acc: 73.42%, LR: 0.000109, Time: 17.52s
[Epoch: 51/75] Loss: 1.1626, Acc: 72.59%, LR: 0.000095, Time: 17.47s
[Epoch: 52/75] Loss: 1.1247, Acc: 74.74%, LR: 0.000676, Time: 17.48s
[Epoch: 53/75] Loss: 1.1201, Acc: 74.70%, LR: 0.000998, Time: 17.73s
[Epoch: 54/75] Loss: 1.1202, Acc: 74.12%, LR: 0.000589, Time: 17.36s
[Epoch: 55/75] Loss: 1.1324, Acc: 73.80%, LR: 0.000050, Time: 17.88s
[Epoch: 56/75] Loss: 1.1440, Acc: 73.37%, LR: 0.000171, Time: 17.47s
[Epoch: 57/75] Loss: 1.1103, Acc: 75.22%, LR: 0.000775, Time: 17.99s
[Epoch: 58/75] Loss: 1.1442, Acc: 73.80%, LR: 0.000976, Time: 17.51s
[Epoch: 59/75] Loss: 1.0973, Acc: 76.03%, LR: 0.000478, Time: 17.62s
[Epoch: 60/75] Loss: 1.1215, Acc: 74.84%, LR: 0.000013, Time: 17.37s
[Epoch: 61/75] Loss: 1.0908, Acc: 76.58%, LR: 0.000263, Time: 17.49s
[Epoch: 62/75] Loss: 1.0869, Acc: 76.21%, LR: 0.000861, Time: 17.42s
[Epoch: 63/75] Loss: 1.1075, Acc: 75.16%, LR: 0.000929, Time: 17.77s
[Epoch: 64/75] Loss: 1.1083, Acc: 74.93%, LR: 0.000367, Time: 17.45s
[Epoch: 65/75] Loss: 1.0880, Acc: 76.09%, LR: 0.000000, Time: 17.74s
[Epoch: 66/75] Loss: 1.1239, Acc: 73.89%, LR: 0.000367, Time: 17.40s
[Epoch: 67/75] Loss: 1.0947, Acc: 75.90%, LR: 0.000929, Time: 17.46s
[Epoch: 68/75] Loss: 1.0751, Acc: 76.42%, LR: 0.000861, Time: 17.32s
[Epoch: 69/75] Loss: 1.1035, Acc: 75.33%, LR: 0.000263, Time: 17.56s
[Epoch: 70/75] Loss: 1.0645, Acc: 77.25%, LR: 0.000013, Time: 17.45s
[Epoch: 71/75] Loss: 1.0847, Acc: 76.33%, LR: 0.000478, Time: 17.81s
[Epoch: 72/75] Loss: 1.0479, Acc: 77.63%, LR: 0.000976, Time: 17.40s
[Epoch: 73/75] Loss: 1.0744, Acc: 76.11%, LR: 0.000775, Time: 17.35s
[Epoch: 74/75] Loss: 1.0800, Acc: 76.60%, LR: 0.000171, Time: 17.52s
[Epoch: 75/75] Loss: 1.0754, Acc: 76.70%, LR: 0.000050, Time: 17.65s
Test Set Accuracy: 88.14%

=== Experiment Results Summary ===
```
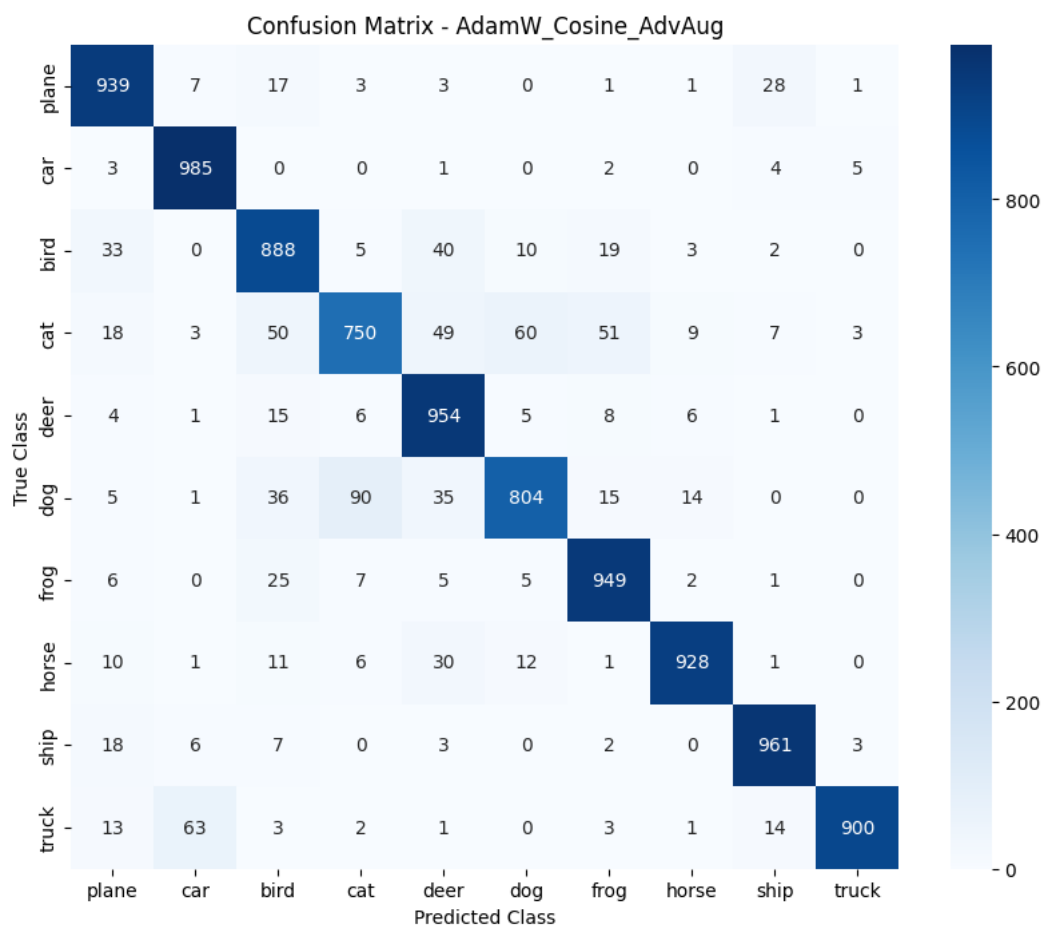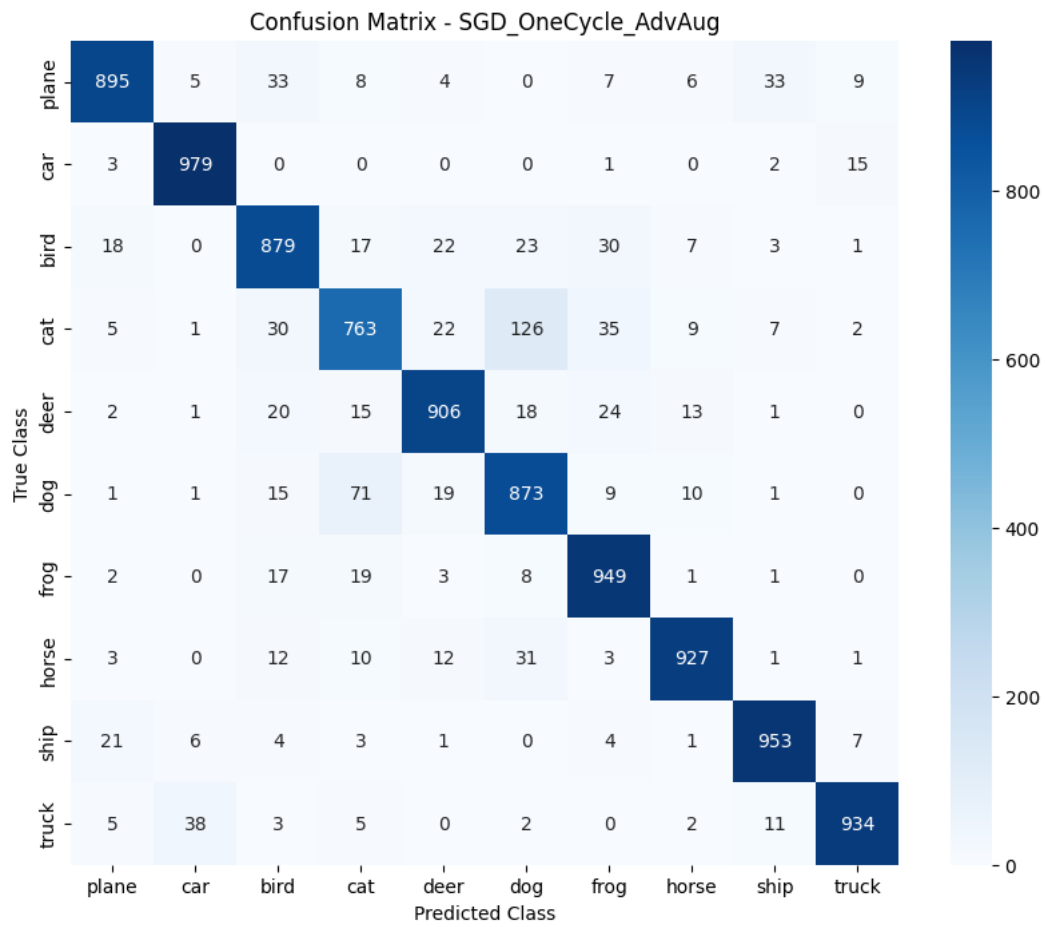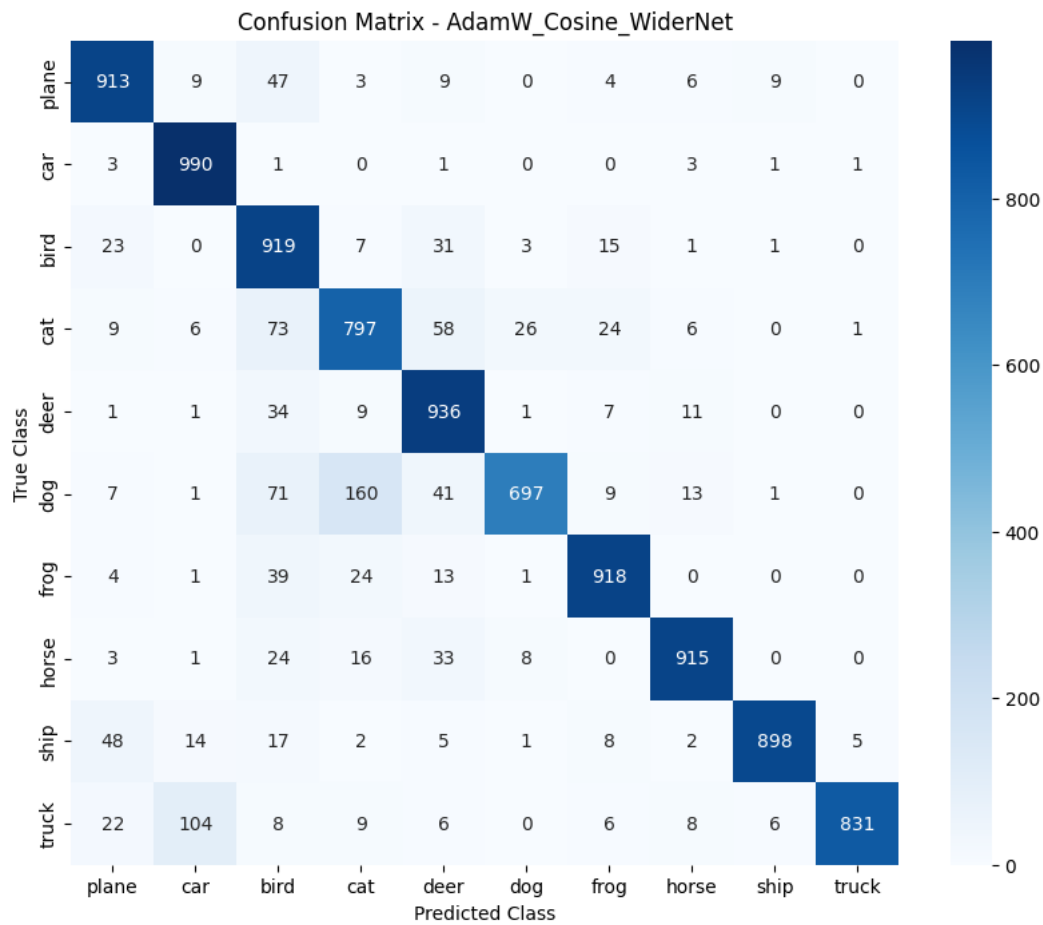
Final Test Accuracy Comparison Across Experiments

Training Process Comparison

Confusion Matrix - SGD_StepLR_Baseline

Confusion Matrix - AdamW_Cosine_AdvAug

Confusion Matrix - SGD_OneCycle_AdvAug

Confusion Matrix - AdamW_Cosine_WiderNet

=== All Experiments Completed ===
- SGD_StepLR_Baseline: 92.24%
- AdamW_Cosine_AdvAug: 90.58%
- SGD_OneCycle_AdvAug: 90.58%
- AdamW_Cosine_WiderNet: 88.14%