

CNN depth and optimisation on CIFAR-10

Portfolio research summary | University Deep Learning coursework, 2026

Question

Does adding convolutional depth improve image classification, and which training changes affect a deep model most?

Method

PyTorch and torchvision; CIFAR-10; seeded training/validation split; data augmentation on training only; softmax regression baseline; CNN depth variants (2, 8, 16 and 32 convolutional layers); 80-epoch CNN-16 improvement experiments. Metrics shown are from the executed notebook supplied for this portfolio.

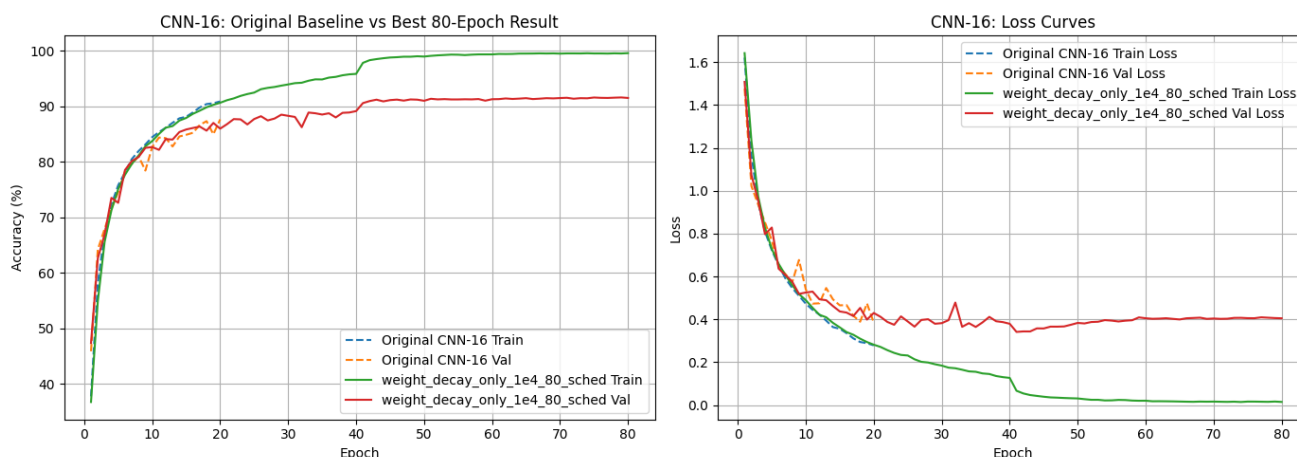
Finding 1. Depth alone did not help

CNN depth	Validation accuracy	Test accuracy
2 layers	75.68%	75.42%
8 layers	87.64%	87.43%
16 layers	87.63%	87.83%
32 layers	37.21%	37.15%

The 32-layer model had poor training and test accuracy, pointing to optimisation difficulty in this setup rather than simply generalisation failure.

Finding 2. Improving the 16-layer model

The original CNN-16 achieved 87.83% test accuracy. With stronger training control and light weight decay, the validation-selected checkpoint reached 91.58% validation accuracy and 91.35% test accuracy.



Original notebook figure: initial CNN-16 versus the selected improved 80-epoch experiment.

What changed

Longer training; step-down learning-rate schedule; selection of best validation checkpoint; evaluation of dropout, weight decay and their combination. The selected weight-decay model had the best validation score; the combined model had a marginally higher test score but was not selected on that basis.

Learning rate and batch size

In separate 20-epoch experiments, learning rate 0.01 returned 86.97% final test accuracy; a learning rate of 1.0 collapsed to chance accuracy. Batch size 64 returned 87.01% final test accuracy, compared with 84.66% for batch size 256.

Limitations

The notebook evaluated the test set during training. A stricter evaluation would reserve it for a final single use and repeat experiments over multiple seeds. These are results from one coursework configuration, not a generalisable benchmark.

Evidence provenance

Figures and metrics were transcribed or exported from the executed assignment notebook. The original report and notebook are retained privately for checking; this summary omits student identifiers and assignment submission details.

Research summary prepared for a personal portfolio. Do not construe as a peer-reviewed study.