

Tuning Summary

The following table summarizes the best configuration found at each stage of the sequential hyperparameter tuning process for the Differential Vision Transformer (Diff-ViT) on the CIFAR-10 dataset. Each tuning run within a stage used 20 epochs.

Table 1: Best Results per Diff-ViT Tuning Stage (20 Epochs/Run)

Stage	Goal	Best Config ID	Selected Parameter	Best Val Acc	Test Acc
1	Patch Size	S1_Patch_4	patch_size=4	0.7072	0.6993
2	Model Params	S2_Param_wider	arch=wider*	0.7900	0.7900
3	Data Augmentation	S3_Aug_Mild	aug=Mild**	0.7920	0.7906
4	Pos. Embedding	S4_PE_sinusoidal	PE=sinusoidal	0.8100	0.8084
Overall Best Config from Tuning (Stage 4)				0.8100	0.8084

*Best architecture from Stage 2 (wider): embed_dim=384, depth=8, num_heads=12.

**Best augmentation from Stage 3 (Mild): ['random_crop', 'horizontal_flip'].

Note: Test accuracy corresponds to the model weights achieving the best validation accuracy during that specific run.

Final Model

The tuning process identified configuration `S4_PE_sinusoidal` as the best overall. The script saved the final model's state dictionary and configuration corresponding to this run. No additional final training stage was performed in this script version. The best test accuracy achieved by this configuration was 0.8084.