A short detail on hyper-parameter choices for PneumoniaMNIST data

Pragya Kumari

June 8, 2025

1 Hyperparameter Choices (Detailed Explanation)

Hyperparameter	Value	Justification
Learning Rate (Phase 1)	0.001	Standard for training new layers; high enough to learn quickly but not cause instability.
Learning Rate (Phase 2)	1.00E-05	Very small to avoid distorting pre-trained ResNet-50 features during fine-tuning.
Batch Size	32	Balances GPU memory usage and gradient stability. Smaller batches add noise (regularization), but too small may slow convergence.
Epochs (Phase 1)	15	Early stopping typically halts training earlier ($\sim$5-10 epochs).
Epochs (Phase 2)	25	Fine-tuning requires more epochs due to the low learning rate.
Dropout Rate	0.5	Common default; empirically works well to prevent co-adaptation of neurons.
Patience (Early Stopping)	5 (Phase 1), 8 (Phase 2)	More patience in Phase 2 to allow slow fine-tuning progress.