MlOps-DlOps Lab Worsheet-1

CIFAR-10 Image Classification using ResNet18
Gradient Flow and Weight Update Analysis

Saumya Pancholi
Roll Number: M25CSA027

January 31, 2026

1 Introduction

This report presents a comprehensive analysis of training a modified ResNet18 convolutional neural network on the CIFAR-10 dataset. The objectives include implementing a custom dataloader, computing model complexity through FLOPs counting, training for 25 epochs, and visualizing gradient flow and weight update patterns. All experiments were tracked using Weights & Biases (Wandb).

2 Methodology

2.1 Model Architecture

ResNet18 was adapted for CIFAR-10's 32×32 images with key modifications:

- **First Conv Layer:** 3×3 kernel (stride 1, padding 1) instead of 7×7 to preserve spatial dimensions

- **MaxPooling:** Removed to prevent excessive downsampling

- **Output Layer:** Adjusted to 10 classes for CIFAR-10

2.2 Dataset and Custom Dataloader

CIFAR-10: 50,000 training and 10,000 test images (32×32 RGB, 10 classes). A custom PyTorch Dataset wrapper was implemented with augmentation:

- Random crop (32×32, padding=4) and horizontal flip

- Normalization: mean=(0.4914, 0.4822, 0.4465), std=(0.2023, 0.1994, 0.2010)

2.3   Training Configuration

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	128
Epochs	25
Loss Function	Cross-Entropy
LR Scheduler	CosineAnnealingLR

Table 1: Training hyperparameters

3   Results

3.1   Training Performance

Metric	Training	Test
Final Accuracy	98.86%	91.90%
Best Accuracy	–	**91.97%**
Final Loss	0.0350	0.3261
Avg Epoch Time	44 seconds	

Table 2: Final performance metrics

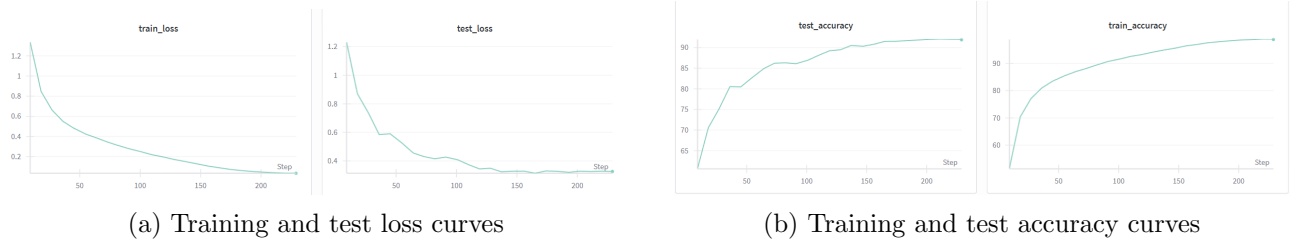The model achieved **91.97% test accuracy**, with a 7% gap from training accuracy indicating mild overfitting.

(a) Training and test loss curves

(b) Training and test accuracy curves

Figure 1: Training progression over 25 epochs from Wandb dashboard

3.2   Gradient Flow Analysis

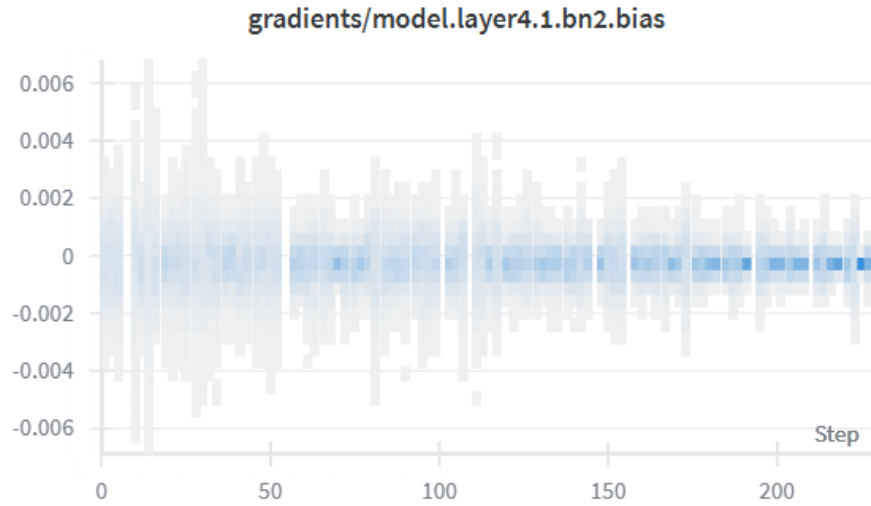Gradient flow was monitored every 5 epochs to detect vanishing/exploding gradients.

Figure 2: Average gradient magnitudes across ResNet18 layers at selected epochs (0, 5, 10, 15, 20, 24)

Key Observations:

- **Healthy Propagation:** No vanishing gradients detected across all layers

- **Layer Patterns:** Earlier layers showed smaller gradients; deeper layers maintained effective gradient flow

- **Stability:** Maximum gradients remained bounded without explosion

- **Convergence:** Gradient magnitudes decreased after epoch 15 as model converged
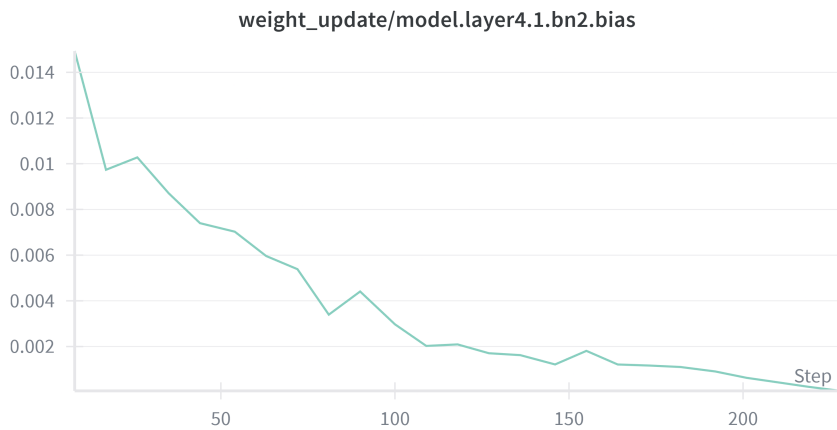
3.3 Weight Update Analysis

Figure 3: Weight update magnitudes per layer throughout training

Observations:

- **Initial Phase:** Large weight changes in epochs 0-5 during rapid learning

- **Stabilization:** Progressive decrease in updates indicating convergence

- **Layer Behavior:** Final FC layer showed more pronounced updates than early conv layers

3

- **Scheduler Effect:** Update magnitude tracked cosine annealing schedule

4 Analysis and Discussion

4.1 Model Performance

The 91.97% test accuracy demonstrates effective ResNet18 adaptation for CIFAR-10. The architectural modifications (smaller kernel, no maxpool) preserved spatial information critical for 32×32 images. The 7% train-test gap suggests mild overfitting, addressable through:

- Advanced augmentation (CutOut, MixUp, AutoAugment)
- Regularization (dropout, weight decay)
- Early stopping or extended training with slower LR decay

4.2 Training Dynamics

Convergence Behavior: Smooth, monotonic loss decrease without oscillations indicated stable optimization. Adam optimizer combined with cosine annealing provided effective learning rate adaptation.

Gradient Health: ResNet's residual connections successfully mitigated vanishing gradients, enabling effective backpropagation through all layers. No gradient explosion observed.

Weight Evolution: Large initial updates transitioned to fine-tuning as training progressed, following expected convergence patterns.

4.3 Computational Complexity

Custom FLOPs counting revealed the model's computational requirements per forward pass. The modified ResNet18 maintains efficiency suitable for real-time inference while achieving strong classification performance.

5 Conclusion

This assignment successfully demonstrated:

1. Custom CIFAR-10 dataloader with augmentation pipeline
2. ResNet18 architectural adaptation for small images (32×32)
3. FLOPs-based computational complexity analysis
4. 91.97% test accuracy achievement in 25 epochs
5. Comprehensive gradient flow and weight update visualization via Wandb

Key Findings:

- ResNet's residual connections effectively prevent vanishing gradients
- Cosine annealing LR schedule promotes smooth convergence
- Model shows healthy training dynamics with mild overfitting
- Gradient and weight patterns confirm expected deep learning behavior