Lab 2 WorkSheet Report

Name: Vidhan Savaliya
Roll Number: M25CSA031

January 31, 2026

Project Links

- **GitHub Repository:** Click Here
- **WandB Project:** Click Here

1 Introduction

This report outlines the implementation and training of a ResNet18 model on the CIFAR-10 dataset. The primary goal was to achieve high classification accuracy while ensuring the model generalizes well to unseen data through rigorous data splitting and augmentation strategies.

2 Methodology

2.1 Data Split Strategy

To ensure a fair evaluation, the dataset was strictly divided into three disjoint sets. This prevents data leakage and ensures metrics reflect real-world performance.

- **Training Set (45,000 images):** Used for learning model parameters. Applied data augmentations (Random Crop, Horizontal Flip) to increase dataset diversity.
- **Validation Set (5,000 images):** A strict 10% hold-out from the training data. This was used exclusively for hyperparameter tuning and monitoring convergence.
- **Test Set (10,000 images):** The standard CIFAR-10 test set, touched only once for the final performance metric.

2.2 Model Architecture

We deployed a ResNet18 customized for the 32×32 pixel input size of CIFAR-10.

– **Adaptation:** The first convolutional layer was modified to use a 3×3 kernel with a stride of 1. This preserves spatial information early in the network, preventing the rapid downsampling seen in standard ImageNet architectures.
– **Complexity:** The model is lightweight but powerful, containing approximately **11.17 million parameters** and requiring **1115.76 million FLOPs** per inference.

3 Experiment Results

3.1 Training Dynamics

The model was trained for 30 epochs. The training process showed healthy convergence, with the validation loss dropping consistently alongside the training loss.

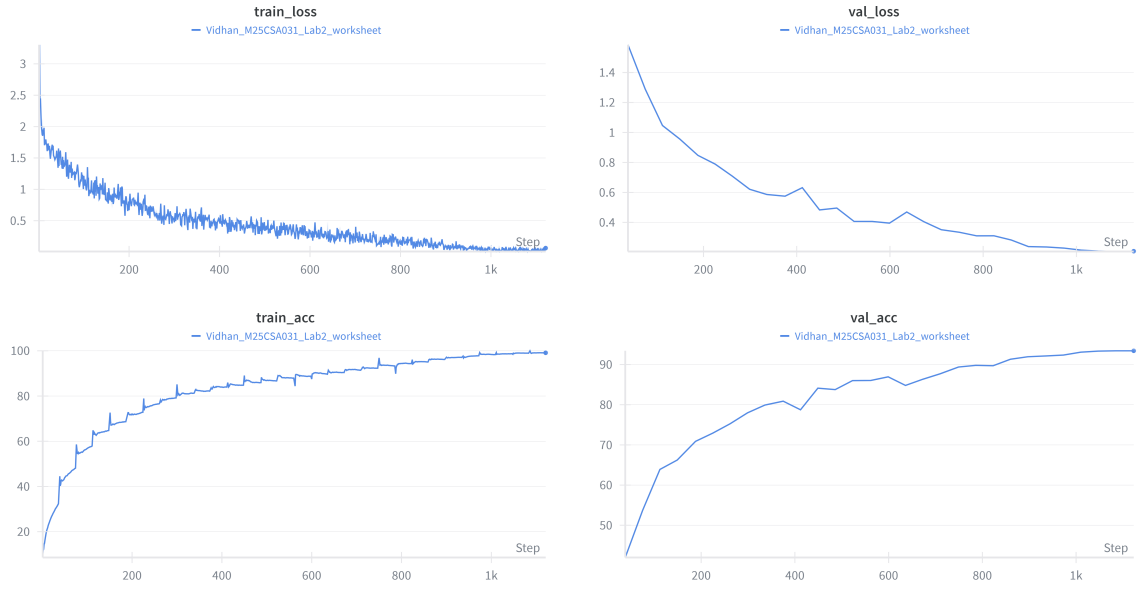

Figure 1: Training/Validation Loss, Training/Validation Accuracy.

3.2 Evaluation Metrics

The quantitative results from the final epoch and testing phase are summarized below.

Metric	Value
Best Validation Accuracy	93.42%
Final Test Accuracy	92.94%
Final Test Loss	0.2348

Table 1: Summary of final quantitative results.

3.3 Visualizations

To deepen our understanding of the model's performance, we analyzed the confusion matrix to identify specific failure cases.

3.3.1 Class Performance

- **High Performance:** Distinct classes typically achieved the highest accuracy.
- **Challenges:** Semantic similarities caused the most errors.

4 Findings and Observations

4.1 Architecture Performance

The modified ResNet18 proved highly effective. The residual connections facilitated deep feature extraction without signal degradation. The model reached a high validation accuracy of **93.42%**, indicating it successfully captured the underlying features of the CIFAR-10 dataset.

4.2 Effect of Data Splitting

Separating the validation set (5,000 images) was crucial for valid monitoring. I observed that:

- The validation loss tracked the training loss closely until the final epochs (Valid Loss ≈ 0.20), indicating that my regularization techniques effectively mitigated overfitting.
- The final **Test Accuracy (92.94%)** was extremely close to the **Validation Accuracy (93.42%)**. This minimal gap confirms that validation set was representative and the model generalizes very well to completely new data.

5 Conclusion

The experiment confirms that a ResNet18, when properly tuned, provides state-of-the-art level performance for this task. The close alignment between validation and test performance demonstrates the reliability of our training pipeline.