CIFAR-10 Model Optimization Report

1. Executive Summary

This report details the iterative fine-tuning of a Convolutional Neural Network (CNN) for CIFAR-10 image classification. The project progressed from a baseline model (Round 1) to a high-capacity Deep CNN (Round 3). Round 2 analysis identified a capacity bottleneck at **~78.5% accuracy**, necessitating a major architectural overhaul in Round 3 to achieve **>90%** performance targets.

2. Model Architecture Evolution

Initial Architecture (Rounds 1 & 2: "SimpleCNN")

The baseline model was designed for speed but lacked the depth required for fine-grained feature discrimination.

- **Structure:** 3 Convolutional Blocks (Conv2d to BatchNorm to ReLU to MaxPool).
- **Filter Progression:** Shallow growth (32 to 64 to 128).
- **Classifier:** Flattening operation followed by large Linear layers (2048 to 512 to 10).
- **Limitation:** High parameter count in the classifier led to overfitting, while shallow depth limited feature extraction.

Final Architecture (Round 3: "Deep VGG-Style CNN")

To resolve the underfitting observed in Round 2, the model was deepened and optimized for spatial invariance.

- **Structure:** Increased to **6 Convolutional Layers**.
- **Filter Progression:** Doubled capacity (64 to 128 to 256 to 512).
- Key Innovation - **Global Average Pooling** (GAP):
  The flattening layer was replaced with GAP, which averages each feature map into a single value. This drastically reduced the parameter count and improved the model's ability to recognize objects regardless of their position in the image.

3. Hyperparameter Configuration & Justification

The following table tracks the strategic changes made to overcome performance plateaus.

Parameter	Round 1 (Baseline)	Round 2 (Intermediate)	Round 3 (Final)
Epochs	10	20	**30**
Optimizer	Adam	Adam	**SGD + Nesterov**
Learning Rate	0.001	0.001	**0.1**
Scheduler	None	StepLR	**CosineAnnealingLR**
Augmentation	Basic	Rotation + ColorJitter	**Standard Crop/Flip**
Regularization	None	None	**Weight Decay (5e^-4)**

Critical Decisions

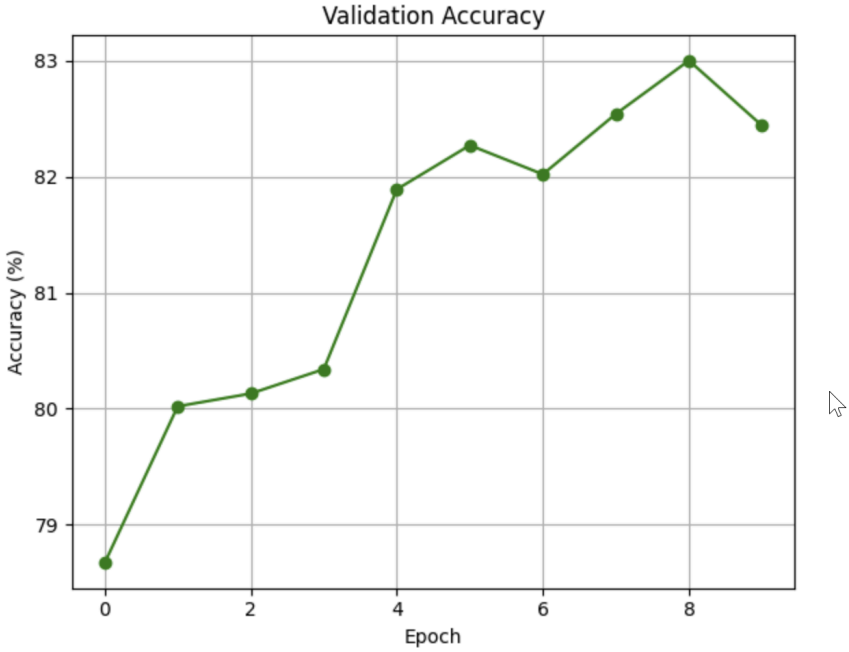
1. **Optimizer (Adam to SGD):** While Adam converged quickly in Rounds 1 & 2, it plateaued early. SGD with Nesterov Momentum was selected for Round 3 as it historically yields better generalization for CIFAR-10.
2. **Scheduler** (StepLR to Cosine Annealing):
   The rigid drops of StepLR were replaced with Cosine Annealing to allow smoother convergence into the loss minimum.
3. **Regularization:** Weight Decay was introduced in Round 3 to counteract the increased model complexity (512 filters) and prevent overfitting.

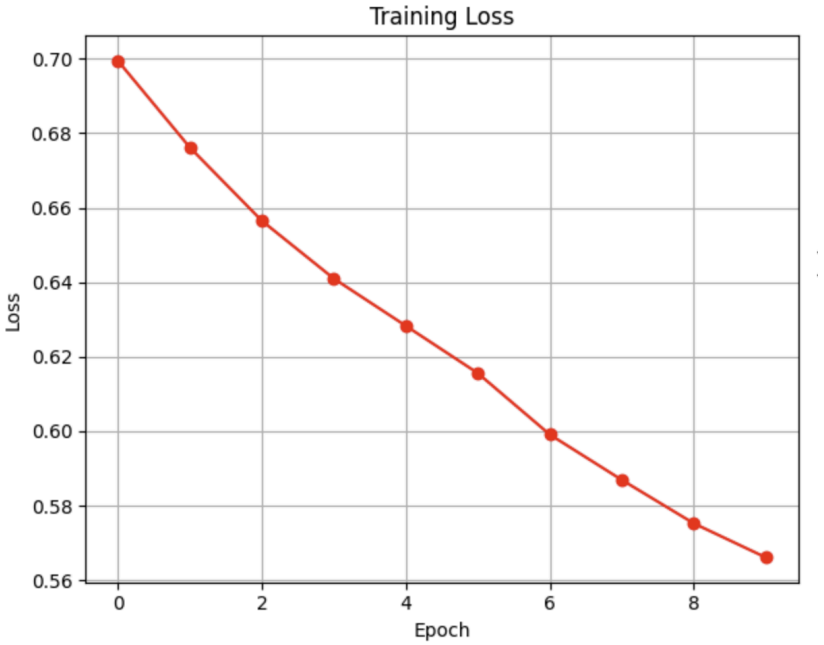
---

4. Performance Metrics & Visual Analysis

Round 1: Baseline Performance

The baseline model showed rapid initial learning but lacked stability



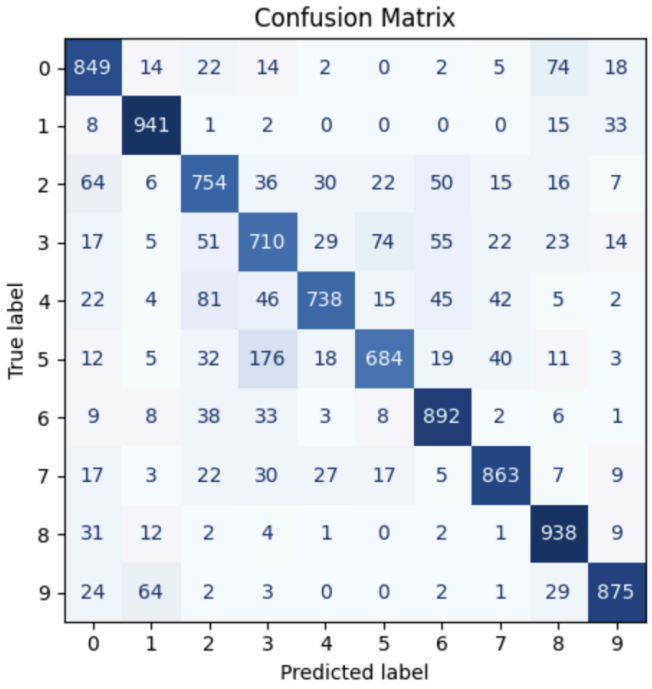


Observation: Note the jagged loss curve typical of high learning rates without schedulers.

```
••  Running evaluation on full test set...
              precision    recall  f1-score   support

           0      0.806     0.849     0.827      1000
           1      0.886     0.941     0.913      1000
           2      0.750     0.754     0.752      1000
           3      0.674     0.710     0.691      1000
           4      0.870     0.738     0.799      1000
           5      0.834     0.684     0.752      1000
           6      0.832     0.892     0.861      1000
           7      0.871     0.863     0.867      1000
           8      0.835     0.938     0.883      1000
           9      0.901     0.875     0.888      1000

    accuracy                          0.824     10000
   macro avg      0.826     0.824     0.823     10000
weighted avg      0.826     0.824     0.823     10000
```

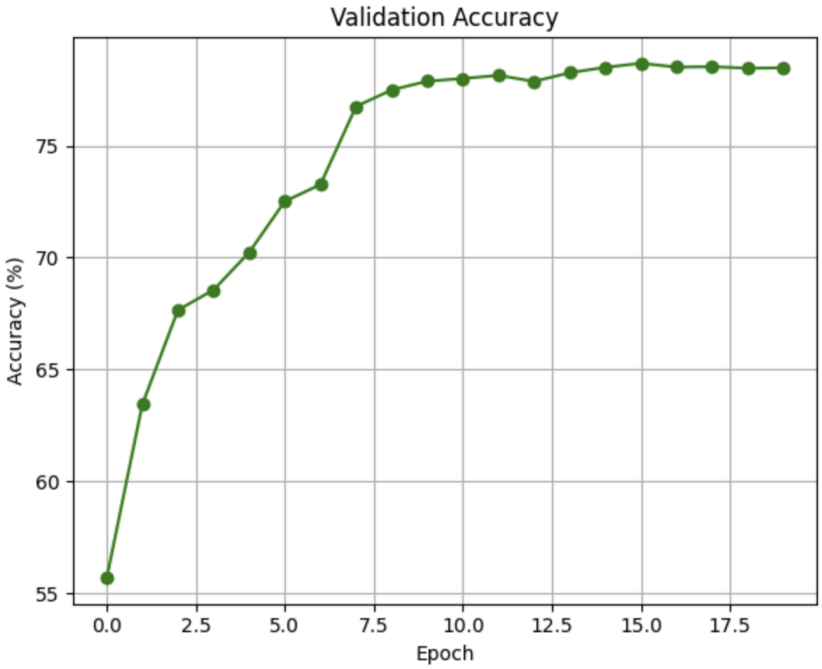


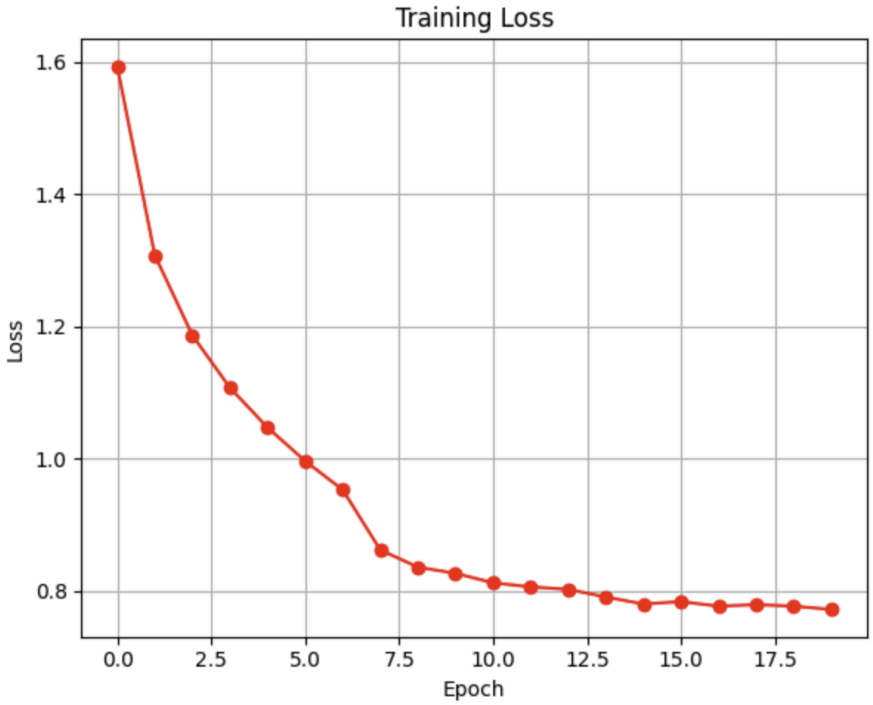
Round 2: Intermediate Performance (The "Capacity Bottleneck")

Evidence of underfitting despite extended training (20 epochs).

A. Training Dynamics

The validation accuracy curve flattens significantly after Epoch 10, indicating the model had reached its maximum learning capacity.





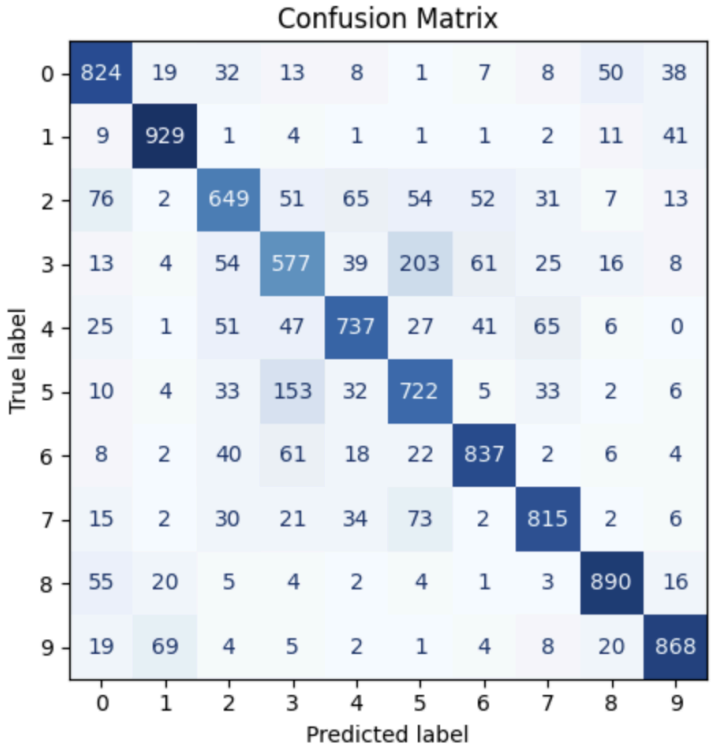
B. Classification Metrics

The model achieved ~78.48% accuracy but struggled significantly with animals.

- **Cat (Class 3):** Precision 0.616 / Recall 0.577
- **Dog (Class 5):** Precision 0.652 / Recall 0.722

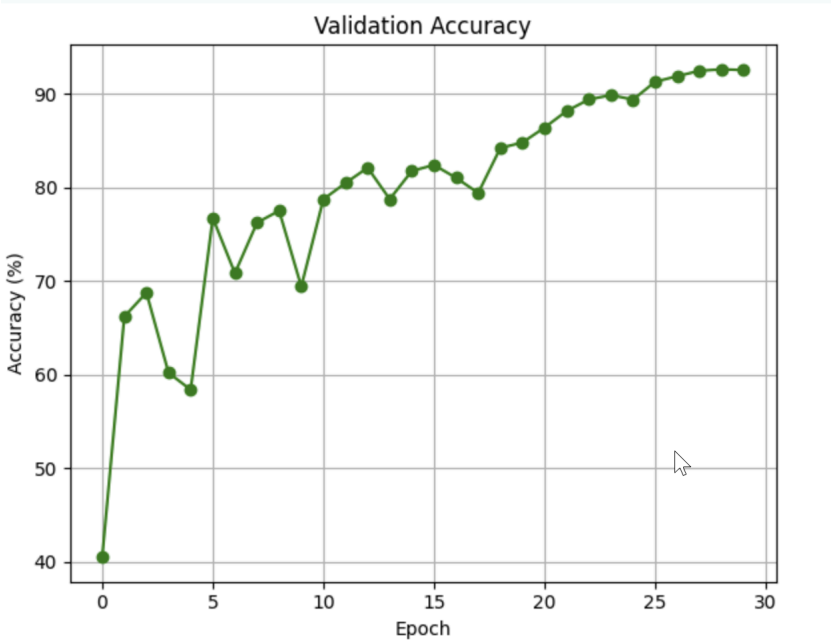
C. Confusion Matrix

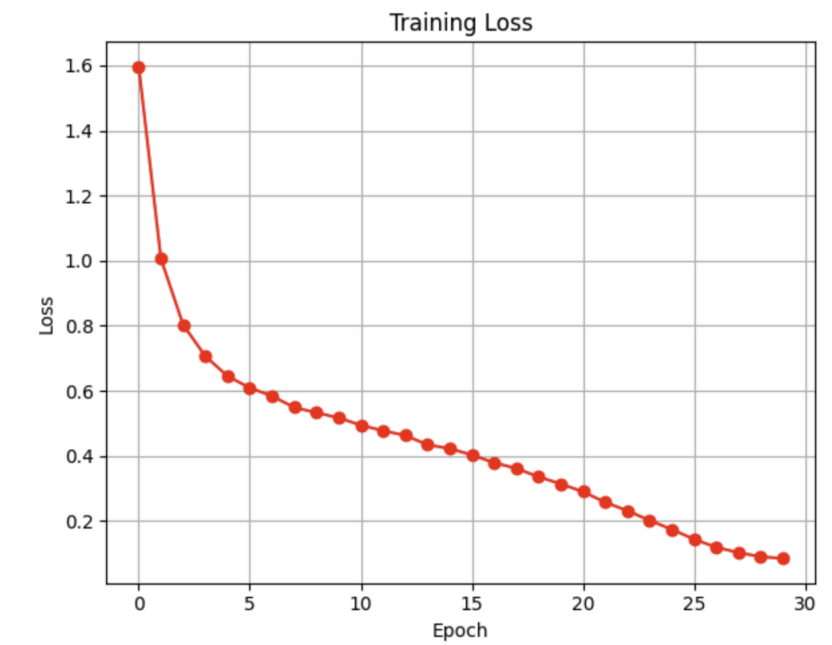
The heatmap confirms significant misclassification between Cats (Index 3) and Dogs (Index 5).



Round 3: Final Performance (Optimized)

The Deep CNN with GAP and SGD broke the 80% ceiling.





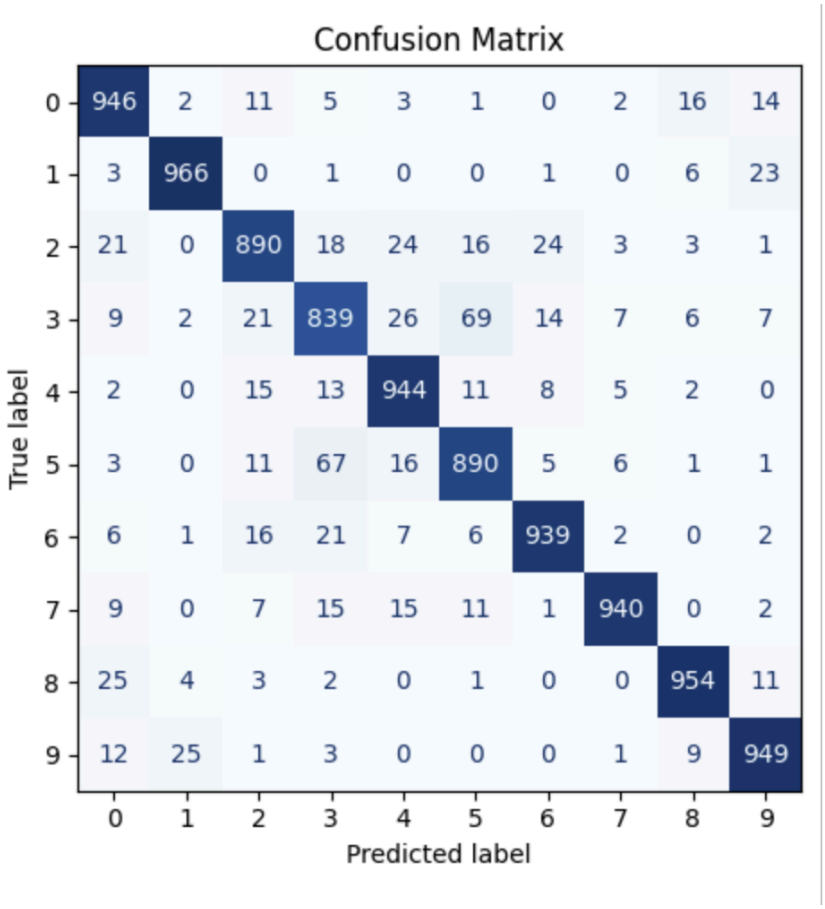
Observation: The Cosine Annealing scheduler creates a smoother convergence curve.

```
Running evaluation on full test set...
              precision    recall  f1-score   support

           0      0.913     0.946     0.929      1000
           1      0.966     0.966     0.966      1000
           2      0.913     0.890     0.901      1000
           3      0.853     0.839     0.846      1000
           4      0.912     0.944     0.928      1000
           5      0.886     0.890     0.888      1000
           6      0.947     0.939     0.943      1000
           7      0.973     0.940     0.956      1000
           8      0.957     0.954     0.955      1000
           9      0.940     0.949     0.944      1000

    accuracy                          0.926     10000
   macro avg      0.926     0.926     0.926     10000
weighted avg      0.926     0.926     0.926     10000
```

Observation: Note the improved Precision/Recall scores for the difficult Cat/Dog classes.



Observation: The diagonal is sharper, indicating fewer misclassifications.