Deep Learning for Image Classification on 256 Object Categories

Abhyudaya Singh

June 28, 2025

1 Introduction

This report documents the implementation and evaluation of a deep learning model for image classification on the 256 Object Categories dataset. The project uses PyTorch to build and train a convolutional neural network (CNN) to classify images into 257 different categories.

2 Dataset

The dataset used is the 256 Object Categories dataset, containing images divided into 257 classes. The dataset was split into:

- Training set: 60% of the data
- Validation set: 20% of the data
- Test set: 20% of the data

Data augmentation techniques were applied to the training set, including random horizontal flips. All images were resized to 128x128 pixels and normalized using ImageNet mean and standard deviation values.

3 Model Architecture

The implemented model, named `ImprovedCNN`, consists of:

3.1 Feature Extraction

- Conv2d (3→64) + BatchNorm + ReLU + MaxPool
- Conv2d (64→128) + BatchNorm + ReLU + MaxPool
- Conv2d (128→256) + BatchNorm + ReLU + MaxPool
- Conv2d (256→512) + BatchNorm + ReLU + AdaptiveAvgPool

3.2 Classifier

- Linear (512*4*4→1024) + ReLU + Dropout(0.5)

- Linear (1024→512) + ReLU + Dropout(0.3)

- Linear (512→num_classes)

The model was trained on an NVIDIA GeForce RTX 4050 Laptop GPU when available, falling back to CPU otherwise.

4 Training

The model was trained with the following parameters:

- Loss Function: CrossEntropyLoss

- Optimizer: Adam (learning rate=0.0001, weight decay=1e-4)

- Learning Rate Scheduler: ReduceLROnPlateau (patience=3, factor=0.5)

- Early Stopping: Patience of 7 epochs

- Batch Size: 64

- Maximum Epochs: 30

The training process achieved the following best results:

- Training Accuracy: 51.52%

- Validation Accuracy: 43.76%
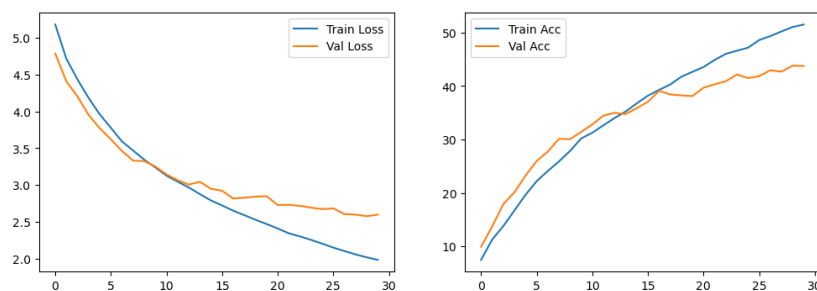
- Training Loss: 1.9830

- Validation Loss: 2.5976

Figure 1: Training and validation metrics over epochs (Model has overfitted a bit. Ideal epochs should've been around 12-13)

5　Evaluation

The model achieved the following performance on the test set:

Metric	Value
Accuracy	0.4417
Precision	0.4513
Recall	0.4417
F1-Score	0.4259

Table 1: Test set performance metrics

6　Conclusion

The implemented CNN model achieved moderate performance on the 256 Object Categories dataset, with a test accuracy of 44.17%. The model showed signs of overfitting, as evidenced by the higher training accuracy compared to validation accuracy. Potential improvements would include:

- More aggressive data augmentation

- Model architecture modifications (deeper networks, different regularization techniques)

- Hyperparameter tuning

- Longer training with different learning rate schedules

That's pretty much the end of it. But the data I downloaded was named archive, inside which there was another folder named 256_ObjectCategories. I got confused at first. But it contained all the empty photos that its parent folder didn't have.