Title: Deep Learning Experiments on CIFAR-100 Using ResNet, VGG, GoogLeNet, DenseNet, and ResNeXt

Table of Contents:

1. Introduction

This report explores the performance of various convolutional neural network (CNN) architectures on the CIFAR-100 dataset, including ResNet-18, VGG-16, GoogLeNet, DenseNet-121, and ResNeXt-50. The aim was to identify the architecture best suited for achieving high accuracy on this challenging dataset.

2. Dataset Overview

- **Dataset**: CIFAR-100
- **Structure**:
    - **Training**: 50,000 images (500 per class)
    - **Testing**: 10,000 images (100 per class)
    - **Validation**: 20% of training data
- **Labels**: "Fine" labels used (100 classes).
- **Normalization**: Mean: (0.5071, 0.4867, 0.4408), Std: (0.2675, 0.2565, 0.2761).
- **Data Augmentation**: Random horizontal flip and cropping.

3. Experimental Setup

- **Hardware**: Apple Silicon GPU (MPS backend) and CPU.
- **Frameworks and Libraries**: PyTorch, torchvision, NumPy.
- **Batch Size**: 128.

- **Optimization**: SGD/Adam with appropriate learning rate schedulers.
- **Loss Function**: Cross-Entropy Loss.

4. Model Architectures and Modifications

- **ResNet-18**: Modified first convolution and removed max-pooling.
- **VGG-16**: Updated the fully connected classifier layer.
- **GoogLeNet**: Adjusted auxiliary logits and the first convolution layer.
- **DenseNet-121**: Customized classifier and initial convolution.
- **ResNeXt-50**: Implemented a standard ResNeXt block with appropriate changes for CIFAR-100.

5. Training and Evaluation

- **Initial Training**:
    - Conducted on a subset of the training data.
    - Top-performing models identified based on validation accuracy.

Results from Initial Training:

Model	Validation Accuracy (LR = 0.1)	Validation Accuracy (LR = 0.001)
ResNet-18	50.30%	22.76%
VGG-16	0.84%	0.97%
GoogLeNet	51.76%	49.20%
DenseNet-121	59.20%	60.60%
ResNeXt-50	38.79%	22.00%

Full Training: Top models (ResNet-18, GoogLeNet, DenseNet-121) retrained on the full training set for 10 epochs.

Results from Full Training:

Model	**Test Accuracy**	**Test Accuracy (LR = 0.001)**
ResNet-18	53.11%	37.77%
GoogLeNet	43.26%	61.20%
DenseNet-121	**63.61%**	**66.78%**

6. Comprehensive Report on Top Performing Models: DenseNet-121, ResNet-18, and GoogLeNet

1. DenseNet-121

Architecture

DenseNet-121 is renowned for its densely connected architecture, where each layer receives direct inputs from all preceding layers. This design promotes feature reuse, mitigates the vanishing gradient problem, and reduces the number of parameters compared to traditional CNNs.

- **Modifications for CIFAR-100:**
    - **First Convolutional Layer:** Adjusted to a kernel size of 3, stride of 1, and padding of 1 to better accommodate the 32x32 image size of CIFAR-100.
    - **Initial Pooling Layer:** Removed to preserve spatial dimensions, enhancing feature extraction for smaller images.
    - **Classifier:** Modified to output 100 classes instead of the default 1000.

Activation Functions

- **ReLU (Rectified Linear Unit):** Employed throughout the network to introduce non-linearity and facilitate faster training.

Optimizer and Hyper-parameters

- **Optimizer:** AdamW
- **Learning Rate (LR):** 0.001
- **Weight Decay:** 1e-4
- **Batch Size:** 128
- **Number of Epochs:** 10
- **Learning Rate Scheduler:** StepLR with a step size of 30 epochs and a gamma of 0.1. (Note: Scheduler steps were not reached within the 10-epoch training period.)

Total Number of Parameters

- **Approximately 8 Million Parameters**

2. GoogleNet

Architecture

GoogleNet (Inception v1) is distinguished by its inception modules, which allow for simultaneous multi-scale feature extraction. This architecture efficiently balances computational cost and network depth.

- **Modifications for CIFAR-100:**
  - **First Convolutional Layer:** Adjusted to a kernel size of 3, stride of 1, and padding of 1 to suit the CIFAR-100 image dimensions.
  - **Initial Pooling Layer:** Removed to maintain higher resolution feature maps.
  - **Classifier:** Modified to output 100 classes instead of the default 1000.

Activation Functions

- **ReLU (Rectified Linear Unit):** Utilised across all layers to introduce non-linearity and improve training dynamics.

Optimizer and Hyper-parameters

- **Optimizer:** Stochastic Gradient Descent (SGD)
- **Learning Rate (LR):** 0.001
- **Momentum:** 0.9
- **Weight Decay:** 5e-4

- **Batch Size:** 128
- **Number of Epochs:** 10
- **Learning Rate Scheduler:** MultiStepLR with milestones at epochs 50 and 75, and a gamma of 0.1. (Scheduler steps were not activated within the 10-epoch training duration.)

Total Number of Parameters

- **Approximately 6.8 Million Parameters**

3. ResNet-18

Architecture

ResNet-18 is a residual network that introduces shortcut connections to bypass one or more layers. This design helps in training deeper networks by addressing the vanishing gradient problem and promoting better feature propagation.

- **Modifications for CIFAR-100:**
    - **First Convolutional Layer:** Adjusted to a kernel size of 3, stride of 1, and padding of 1 to better fit the CIFAR-100 image size.
    - **Initial Pooling Layer:** Removed to retain spatial information.
    - **Classifier:** Modified to output 100 classes instead of the default 1000.

Activation Functions

- **ReLU (Rectified Linear Unit):** Applied throughout the network to introduce non-linearity and facilitate learning.

Optimizer and Hyper-parameters

- **Optimizer:** Stochastic Gradient Descent (SGD)
- **Learning Rate (LR):** 0.001
- **Momentum:** 0.9
- **Weight Decay:** 5e-4
- **Batch Size:** 128
- **Number of Epochs:** 10
- **Learning Rate Scheduler:** MultiStepLR with milestones at epochs 50 and 75, and a gamma of 0.1. (Scheduler steps were not triggered within the 10-epoch training period.).

Total Number of Parameters

- **Approximately 11.7 Million Parameters**

7. Observations and Results

- DenseNet-121 consistently outperformed other models in both validation and test phases.
- VGG-16 underperformed significantly, suggesting incompatibility with the dataset or hyperparameter tuning issues.
- GoogLeNet and ResNet-18 showed potential but required further fine-tuning.

8. Additional Experiment: Impact of Learning Rate Adjustment

To evaluate the impact of learning rate adjustments on model performance, I conducted experiments with two learning rates: **0.1** (default setting) and **0.001** (adjusted setting). Validation accuracy for each model under both learning rate conditions is summarised below:

Observations:

- The **default learning rate (0.1)** yielded better validation accuracy for most models except DenseNet-121, which showed slight improvement at a lower learning rate (0.001).
- **GoogLeNet** maintained competitive performance across both learning rates, indicating robust architecture stability.
- Lower learning rates generally resulted in slower convergence, which negatively impacted validation accuracy for models such as ResNet-18 and ResNeXt-50.

Full Training Results with Top Models

Following the initial experiments, the top-performing models (**ResNet-18**, **GoogLeNet**, **DenseNet-121**) were retrained on the full training dataset for **10 epochs**. The test accuracy comparisons for the two learning rate settings are as follows:

Observations:

- **DenseNet-121** achieved the highest test accuracy (66.78%) with a learning rate of 0.001, demonstrating its strong generalisation capability.
- **GoogLeNet** performed significantly better with a reduced learning rate, achieving a test accuracy of 61.20%.
- **ResNet-18** showed better results with the default learning rate, indicating that higher rates may be more suitable for this architecture within the given training setup.

9. Ranking on CIFAR-100 Benchmark

20	**ResNet50 Without Transfer Learning**	67.060	×	ResNet50_on_Cifar_100_Without_Transfer_Learning
21	**AlexNet** (KP)	66.78	×	Learning the Connections in Direct Feedback Alignment
22	**ACN**	66.3	×	Striving for Simplicity: The All Convolutional Net
23	**DLME** (ResNet-18, linear)	66.1	×	DLME: Deep Local-flatness Manifold Embedding

My best model using DenseNet-121 architecture ranks 21st position in the leaderboard when the filter of "model not using extra data" is applied.

10. Conclusion

DenseNet-121 emerged as the most effective architecture for CIFAR-100 under the experimental conditions. VGG-16's architecture struggled with this dataset, possibly due to the absence of batch normalisation and sensitivity to hyperparameters.