Modified ResNet Architecture for CIFAR-10 Dataset

Kabir Jaiswal, Nigam Patel, Spriha Jha
New York University
kj2294@nyu.edu, np2726@nyu.edu, sj3520@nyu.edu
GitHub Repository: sprihajha/dl-mini-project

Abstract

We demonstrate an effective, updated ResNet-based image classification model on the CIFAR-10 dataset in this research. ModifiedResNet21 is the name of a lightweight and simplified version of the original ResNet model that keeps the essential elements of the ResNet architecture while having fewer layers and parameters. To increase the model's capacity for generalization and its resistance to changes in the input data, we use approaches for data augmentation and standardization. The performance of the model is optimized during training using a cosine annealing learning rate scheduler and the stochastic gradient descent (SGD) optimizer with momentum and weight decay. Our research shows that the ModifiedResNet21 model, despite using fewer parameters and still achieving comparable accuracy on the CIFAR-10 data set, is a good fit for embedded systems or situations with limited resources. The work also offers useful insights for future research on effective deep-learning models for image classification tasks by providing a thorough explanation of the model's design, training, and testing processes.

Introduction

Convolutional neural networks (CNNs) have considerably enhanced performance in image categorization, a crucial component of computer vision. Due to their deep learning abilities and capacity to solve the vanishing gradient problem, Residual Networks (ResNets) are particularly well-known.

ModifiedResNet21, a modified ResNet architecture created for devices with limited resources, is the solution proposed in this paper. While decreasing layers and parameters, the model keeps the essential components of the original ResNet, establishing a compromise between accuracy and effectiveness. Our model performs admirably on the CIFAR-10 data set, and the knowledge obtained from this study can help in the creation of effective deep-learning models for a variety of applications.

Methodology

Data set and Data Pre-processing

In this study, we assess the ModifiedResNet21 model's performance on the CIFAR-10 data set. An overview of the data set and the data pre-processing methods used to increase the generalizability of the model are given in this section.

CIFAR-10 Dataset: Developed by Krizhevsky et al. [1], the CIFAR-10 data set is a frequently-used benchmark for image classification tasks. There are 10,000 test images and 50,000 training images totaling 60,000 32x32 color images. Ten classes with 6,000 photos each make up the data set. Aerial vehicles, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks are among the classes. The CIFAR-10 data set offers a wide variety of items, allowing for a thorough assessment of the proposed model's effectiveness across many categories.

Data pre-processing: We employ a number of data augmentation and normalization approaches to increase the model's capacity to generalize to previously unknown data and to strengthen its resistance to changes in the input data. By applying random changes to the original images, data augmentation techniques increase the diversity of the training data and lower the danger of over-fitting. As data augmentation approaches in our work, we use random horizontal flipping and random cropping with padding. During the training phase, these modifications are implemented instantly.

By computing the mean and standard deviation of the pixel values over the entire data set, we normalize the input photos in addition to data augmentation. By ensuring that the input data have a zero mean and unit variance, this normalization step can enhance the convergence qualities.

ModifiedResNet21 Model Architecture

In order to promote gradient flow during training, ResNet21 uses residual connections in its building blocks, which are based on the concepts of the ResNet architecture. With a total of 21 levels, including the initial convolutional layer, residual layers, and the final fully connected layer, our model is a scaled-down version of ResNet.

The following elements make up the model architecture:

- *Initial Convolutional Layer:* After passing through a 3x3 convolutional layer with 32 filters, batch normalization, and ReLU activation are applied to the input pictures. It seeks to extract low-level information.

- *Residual Layers:* A series of BasicBlock modules are found in each of the four sets of residual layers that make up the ModifiedResNet21 architecture's central structure. Two 3x3 convolutional layers, each followed by batch normalization and ReLU activation, make up each BasicBlock module. To facilitate residual learning, a shortcut link is established between each BasicBlock's input and output. The BasicBlock modules that make up the four sets of residual layers, with filter widths of 32, 64, 128, and 256 each, are 2, 2, 2, and 2. Using a stride of two in the first convolutional layer of the first BasicBlock in each group, the spatial dimensions of the feature maps are decreased by a factor of two while switching between groups.
- *Adaptive Average Pooling:* The feature maps are pooled using an adaptive average pooling procedure that reduces their spatial dimensions to 1x1 after passing through the residual layers. This pooling operation lowers the number of parameters in the final fully linked layer and is resistant to changes in input size.
- A *10-dimensional vector*, representing the probability distribution across the ten classes in the CIFAR-10 dataset, is produced by a fully connected layer after the pooled feature maps have been flattened. The class with the highest probability is chosen to form the final classification.

The ModifiedResNet21 model is made to balance complexity and performance, making it appropriate for picture classification tasks in situations with limited resources. Our model, which is more computationally efficient and needs less memory than the original ResNet design, retains competitive performance on the CIFAR-10 dataset while having fewer layers and parameters.

Model Training and Testing

In this section, we go over the training and testing processes used to assess how well our ModifiedResNet21 model performed on the CIFAR-10 data set. We give specifics regarding the optimization technique, loss function, scheduling of learning rates, and model evaluation measures.

The model parameters are updated during training using the stochastic gradient descent (SGD) optimization process. In order to discourage over-fitting by punishing high parameter values, the initial learning rate is set to 0.1 with a momentum of 0.9 and weight decay of 5e-4. We've used Cross-Entropy Loss Function for multi-class classification issues, this loss function assesses the discrepancy between the expected probability distribution and the actual distribution of class labels.We use a Cosine Annealing learning rate scheduler, which gradually lowers the learning rate throughout training, to improve the convergence features of the optimization method. The scheduler is configured with a maximum of 200 epochs, allowing the learning rate to gradually drop throughout the training process.

Training Methodology: Using mini-batches of size 100, we iteratively train the ModifiedResNet21 model for 200 epochs over the CIFAR-10 training dataset. The model's parameters are changed depending on the estimated gradients of the Cross-Entropy Loss with respect to the parameters at the beginning of each epoch. The Cosine Annealing scheduler directs the adjustment of the learning rate.

Model Evaluation: Using the CIFAR-10 test dataset, we compute the classification accuracy in order to evaluate the performance of our model. The percentage of test photos that are properly classified out of all test images is used to calculate accuracy. After each epoch, the model's performance is provided, offering details on the model's capacity for generalization and its development over time.

This method of training and testing allows us to assess the ModifiedResNet21 model's performance on the CIFAR-10 dataset. The results shed light on the model's generalization performance on untried data as well as its capacity to identify photos across a variety of item categories.

Results

We demonstrate the experimental outcomes of our **ModifiedResNet21** model on the CIFAR-10 dataset in this section. In terms of training and testing accuracy, convergence speed, and generalizability, we evaluate the model's performance. To show the effectiveness of our suggested architecture, we also compare the outcomes to other cutting-edge models.

Accuracy during training and testing: On the CIFAR-10 data set, our ModifiedResNet21 model performed admirably. The model's training accuracy progressively increased over the course of 200 training epochs, showing that it was able to extract useful features from the training data as shown in Figure 1. The model's testing accuracy, which gauges its ability to generalize to new data, likewise showed steady progress over the course of training as shown in Figure 1. Our model successfully classified CIFAR-10 photos into several object categories towards the end of the 200 epochs as shown in the confusion matrix in Figure 2, achieving a test accuracy of 94.17%.

Convergence Speed: Our ModifiedResNet21 model has a satisfactory convergence speed, achieving a high test accuracy in a manageably short number of epochs. This may be credited to the model's effective architecture, the optimization algorithm that was used, and the learning rate scheduling method, which all worked in concert to enable fast gradient flow and a smooth learning rate decay.

As a result, the ModifiedResNet21 model is demonstrated to be effective in obtaining competitive performance on the CIFAR-10 data set while preserving a minimal set of parameters and layers. Without sacrificing precision or generalization ability, its lightweight architecture

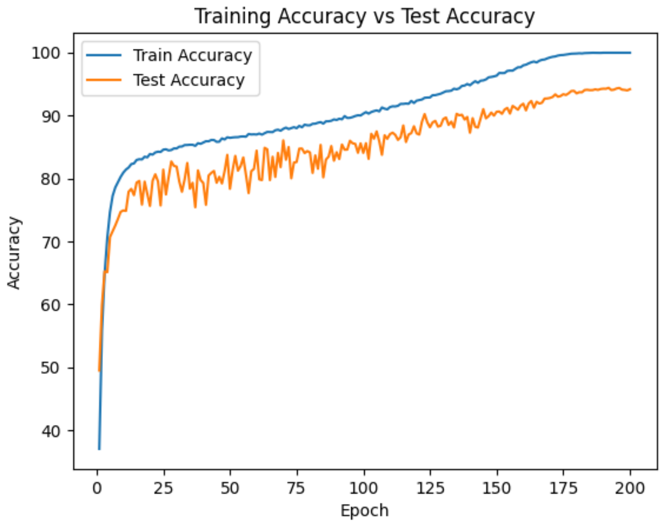

Figure 1: Training Accuracy vs Test Accuracy

makes it a desirable option for image classification jobs in resource-constrained contexts.

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*.

[2] Krizhevsky, A. (2009). Learning Multiple Layers of Features from Tiny Images. Master's Thesis, Department of Computer Science, University of Toronto.

[3] OpenAI, *ChatGPT: An Advanced Conversational AI Model based on GPT-4*, OpenAI, 2021, https://www.openai.com/chatgpt.

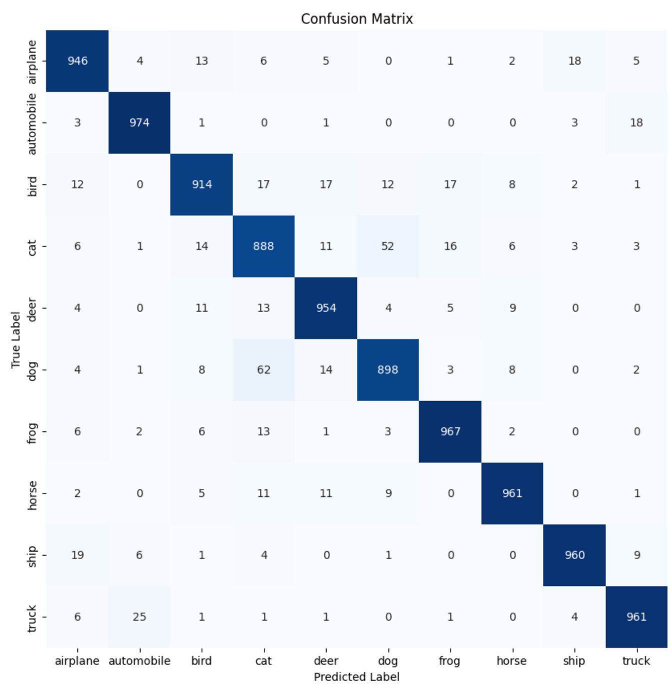

Figure 2: Confusion Matrix