Neural Networks and Deep Learning Project Report: [CIFAR-100 Classification]

Ambar Roy, 12021002016011(12)

November 26, 2023

1 Introduction

Computer vision, a critical field within computer science, empowers machines to comprehend and interpret visual information akin to human perception. Its significance is profound and spans diverse domains, including image and video analysis, automation in robotics, medical imaging for diagnostics, and the development of autonomous vehicles. In security, computer vision facilitates facial recognition and object tracking, while in industries like agriculture, it aids in crop monitoring and disease detection. The technology's role extends to retail, where it enhances inventory management and customer tracking, and to human-computer interaction, enabling natural interfaces through gesture and facial expression interpretation. As a transformative force, computer vision continues to advance industries and daily life, contributing to the development of sophisticated systems and applications.

In this project, CIFAR-100 dataset is used to demonstrate how a neural network functions. CIFAR-100 is a dataset in computer vision, consisting of 60,000 32x32 color images across 100 diverse classes. The dataset is divided into 50,000 training images and 10,000 testing images, each belonging to one of the 100 classes. CIFAR-100 is an extension of CIFAR-10, providing a more complex and diverse set of images for research and evaluation of algorithms in image recognition, deep learning, and other areas of computer vision.

2 Architecture Details

2.1 Model Overview

- The model has 5 hidden layers with its heart as ResNet50.
- ResNet-50 can be adapted for evaluating CIFAR-100, but it's important to consider certain modifications and challenges associated with the architecture and dataset characteristics. CIFAR-100 consists of 32x32 pixel images, and ResNet-50 is designed for larger images.

2.2 Layers

- Upsampling2D
  The first layer in the model is an Upsampling2D layer, which increases the spatial dimensions of the input by a factor of (7, 7) using bilinear interpolation.

- resnet
  Assuming resnet is a pre-trained ResNet50 model, this layer adds the entire ResNet50 model to the Sequential model. ResNet50 is a deep neural network architecture known for its residual blocks, which helps address the vanishing gradient problem in training deep networks.

- GlobalAveragePooling2D
  Global Average Pooling reduces the spatial dimensions of the input to a single value for each feature map. It's a form of spatial compression that helps reduce the number of parameters in the model.

- Dropout
  Applies dropout regularization with a rate of 0.25, randomly setting a fraction of input units to 0 during training to prevent overfitting.

- Fully Connected Layer
  Fully connected dense layer with 256 units and ReLU activation function. It captures non-linear relationships in the data.

- BatchNormalization
  Normalizes and scales the activations of the previous layer, improving training stability and accelerating convergence.

- Output Layer
  The final layer with 100 units and softmax activation, representing the 100 classes in CIFAR-100. It outputs probability scores for each class.

2.3 Model Output

Layer (Type)	**Input Shape**	**Output Shape**	**Parameters**
UpSampling2D	(32, 32, 3)	(1568, 1568, 3)	0
ResNet50	(1568, 1568, 3)	(7, 7, 2048)	23587712
GlobalAveragePooling2D	(7, 7, 2048)	(2048)	0
Dropout	(2048)	(2048)	0
Dense	(2048)	(256)	524544
BatchNormalization	(256)	(256)	1024
Dense	(256)	(100)	25700

3 Results

3.1 Evaluation Metrics

In assessing the performance of ResNet50, the primary metric employed for evaluation was accuracy. Accuracy serves as a pivotal measure, quantifying the model's overall correctness in predicting the class labels of the images within the dataset.

3.2 Performance on Test Set

The model demonstrated a commendable accuracy of 76.83% on the test set, showcasing its proficiency in correctly classifying a significant portion of previously unseen images.

3.3 Performance Metrics Overview

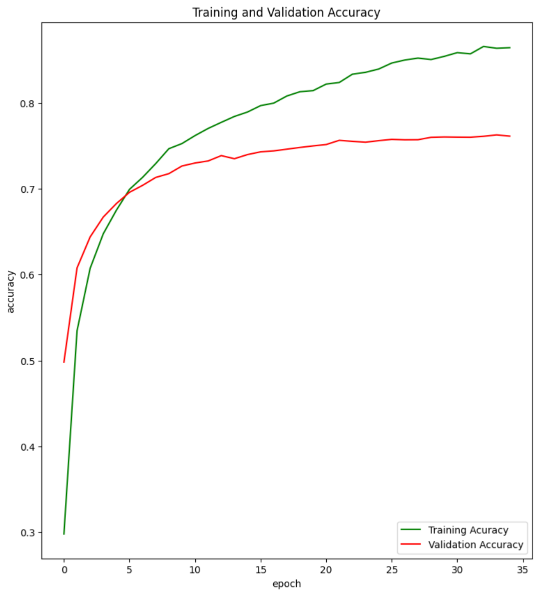

Figure 1: Sample Output

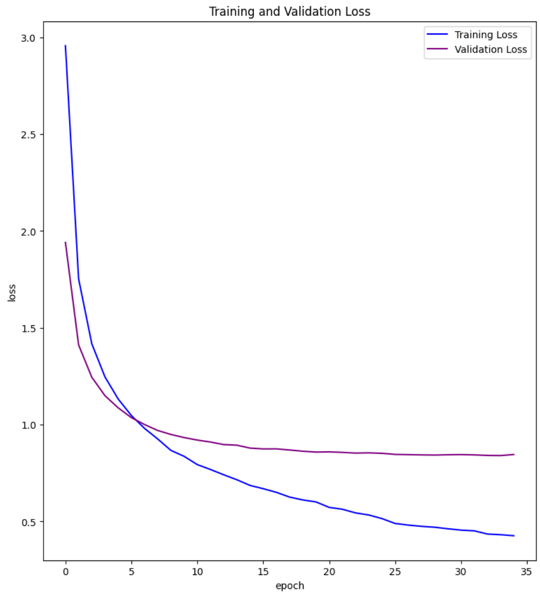

Figure 2: Sample Output

4 Conclusion

In Conclusion, the outcomes provide compelling evidence that approximating the expected optimal sparse structure using readily available dense building blocks is a viable strategy for enhancing neural networks in computer vision. This method offers a substantial improvement in quality with only a modest increase in computational requirements compared to shallower and less wide networks. While achieving a similar quality outcome might be possible with more computationally expensive networks of comparable depth and width, this approach firmly establishes that transitioning to sparser architectures is a feasible and beneficial concept in general.