

Assignment3 - DD2424

Silpa Soni Nallacheruvu

Gradient Check

To verify the correctness of my analytical gradient computations for my three-layer ConvNet, I implemented a comparison against the provided debug data for the gradients and other available variables. The relative mean and max errors were computed for each layer's weights, biases and more:

Comparison	Max Difference	Mean Difference
X_conv Vs conv_outputs	0.0	0.0
conv_outputs_mat Vs conv_outputs_flat	1.70×10^{-13}	6.83×10^{-16}
conv_flat	0.0	0.0
h	0.0	0.0
p	6.76×10^{-16}	5.01×10^{-17}
Fs_flat	1.64×10^{-14}	8.45×10^{-16}
W1	4.50×10^{-14}	2.03×10^{-16}
W2	4.51×10^{-15}	1.25×10^{-16}
b1	4.22×10^{-16}	4.22×10^{-17}
b2	1.31×10^{-15}	2.34×10^{-16}

Table 1: Numerical comparison of intermediate computations and gradients.

As the mean and max difference between provided debug values and my computed values are insignificant, I conclude that my implementation is bug free.

Initial Three layer ConvNet

The initial three-layer ConvNet was implemented with the following architecture:

- Convolutional layer with 32 filters of size 3x3, stride 1, and padding 1.
- ReLU activation function.

- Max pooling layer with a pool size of 2x2 and stride 2.
- Fully connected layer with 128 units.
- ReLU activation function.
- Output layer with softmax activation for classification.

The network was trained on the CIFAR-10 dataset with training size of 49,000 images and validation size of 1,000 images. The short training runs with cyclic learning rates from Assignment2 was applied with a minimum learning rate of 1e-5 and a maximum learning rate of 1e-1 for three cycles with constant steps of 800.

The model with Network architecture 2 ($f = 4, nf = 10, nh = 50$) achieved a final validation accuracy of 56.1% and test accuracy of 55.88%. The training time was reported as 13.07 seconds.

Next, we compare the final test accuracy and training time of all four network architectures with varying f and nf values with short training runs:

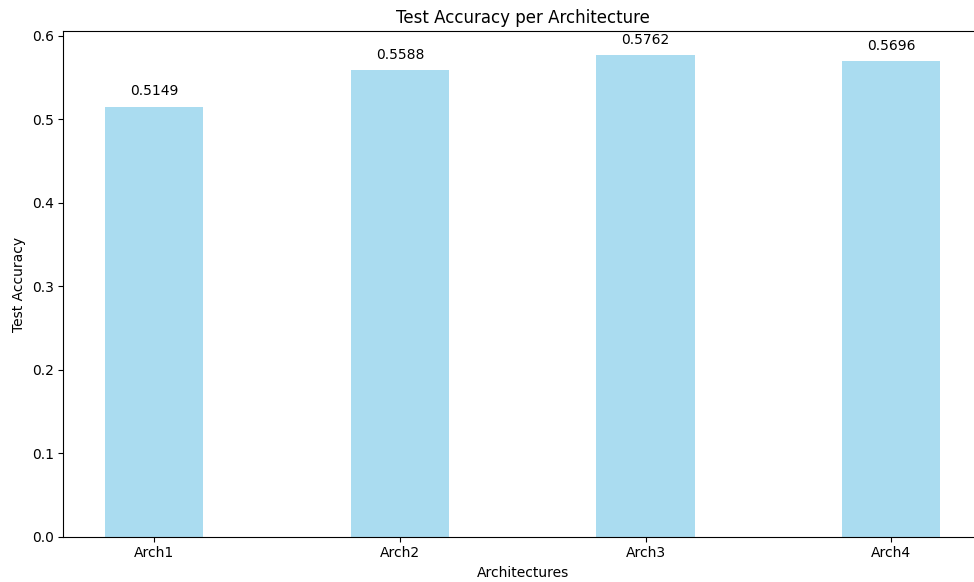


Figure 1: Comparison of final test accuracy for different network architectures.

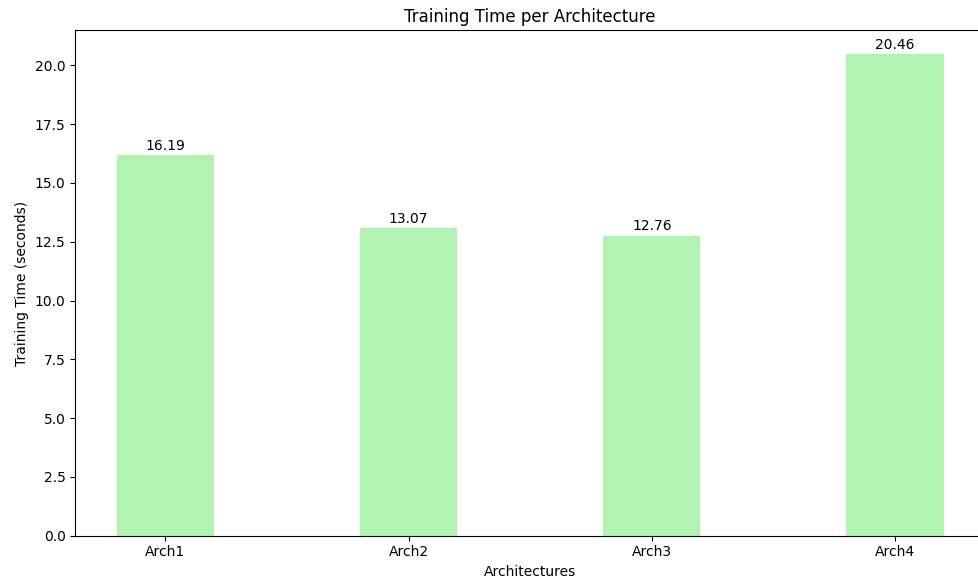


Figure 2: Comparison of total training time for different network architectures.