DLSR\_lab3-2\_3

Introduction

The lab is going design a scalable residual network with three hyperparameters: depth, width, and resolution. Also, monitor its accuracy, execution time, and memory usage.

Experiment Setup

* Python 3.7.4
* Pytorch 1.2.0
* Pytorchvision 0.4.0
* Pillow 6.1.0
* Matplotlib 3.1.3

Workflow

1. Data Preprocessing
2. Test on Validation Data
3. [Loop] Train on Training Data
4. [Loop] Test on Validation Data
5. [Loop] If the Validation Accuracy is better, then save the weights
6. Test on Testing Data with the last saved weights

Data Preprocessing

1. Random Horizontal Flip with p = 0.5
2. Resize to (RESOLUTION \* 32, RESOLUTION \* 32)

Structure

1. CNN (floor (32 \* WIDTH)) + BN + ReLU
2. Type 1 Layer: Block (64 \* WIDTH) x DEPTH
3. Type 1 Layer: Block (128 \* WIDTH) x DEPTH
4. Type 1 Layer: Block (256 \* WIDTH) x DEPTH
5. Type 1 Layer: Block (512 \* WIDTH) x DEPTH
6. Fully-connected Layer

Monitor

1. Accuracy: accuracy on test data.
2. Execution Time: from forward to calculate the batch accuracy. (forward once only)
3. Memory Usage: memory usage in GPU.

Result

1. Independent Variable: DEPTH

(X: Epoch, Y: Accuracy)

(X: DEPTH, Y: Second)

(X: DEPTH, Y: 100MB)

1. Independent Variable (y-axis: epoch): WIDTH

(X: Epoch, Y: Accuracy)

(X: WIDTH, Y: Second)

(X: WIDTH, Y: 100MB)

1. Independent Variable (y-axis: epoch): RESOLUTION

(X: Epoch, Y: Accuracy)

(X: RESOLUTION, Y: Second)

(X: RESOLUTION, Y: 100MB)

1. Scale All Factors Uniformly

(X: Epoch, Y: Accuracy)

(X: Scalar, Y: Second)

(X: Scalar, Y: 100MB)

Discussion

1. Do you get better results by scaling up the model?
   1. Increase of resolution or width rises the accuracy.
   2. But the accuracy decreases when the depth increases.
      1. Unknown bugs.
      2. Overfitting.
      3. Because resolution and width are also low in the experiment, too many layers may over-explain the data, lowing the accuracy.
      4. It may need more epochs.
2. Does scaling up only 1 factor/uniformly scaling up 3 factors lead to a different result?
   1. Yes, scaling the resolution has the best result, followed by scaling all factors.
3. What should we adjust when a model is scaled up?
   1. Batch Size:
      1. Larger model has lower batch size.
      2. GPU (1050-Ti) memory limit is 4GB.
   2. Learning Rate
      1. Larger model should be trained with larger learning rate.
      2. If we train a large model with small learning rate, then it will improve slowly.
      3. If we train a small model with large learning rate, then it will not converge precisely.

Other

1. Memory usage sometimes drop? Why?
   1. We guess it’s due to internal optimization.