

DL Assignment 02

Mini VGG Implementation on CIFAR 10

Sakshi Manish Mukkirwar

04.23.2024

1. Summary

Variant Final Architecture	Data Augmentation	Batch Normalization	Dropouts	Regularization	Scheduler	Observation
Mini VGG (Relu)	Yes	Yes (after every conv layer)	Yes (0.3) before FC	Yes Weight decay = 0.2	Yes	BN, dropout and scheduler helped in reducing overfitting and reaching convergence.
Variant 1 (Selu)	Yes	Yes	No	Yes	Yes	Improved Test accuracy
Variant 1 Swish	Yes	Yes	Yes (0.3) before FC	Yes	Yes	Almost same accuracy as mini vgg, better convergence
Variant 2	Yes	Yes	Yes (0.3) before FC	Yes	Yes	Spikes observed, lower test accuracy
Variant 3 (Dropout after every conv layer)	Yes	Yes	Yes Dropout of 0.5 before the FC layer	Yes	Yes	Good test accuracy and better convergence
Variant 3 (Dropout before FC)	Yes	Yes	Yes Dropout of 0.2 after every conv block	Yes	Yes	Reduced test accuracy, increased gap in the training and validation plots
Variant 4	Yes	Yes	Yes dropout of 0.1 before FC layer	Yes	Yes	Lower test accuracy as compared to others

2. Trials Performed

a) Mini VGG

Initially, original architecture was followed and the network was alarmingly overfitting. To address this, introduced BN, dropout and scheduler to stabilize the training and reduce overfitting. Below are the trials performed:

- 1) LR=0.003, weight_decay=0.008, Dropout(0.3) - **Overfitting**
- 2) LR=0.0003, weight_decay=0.008, Dropout(0.4) - **Slightly underfitting suggesting that the dropout should be reduced**
- 3) LR=0.0003, weight_decay=0.005, Dropout(0.3) = **better convergence**
- 4) LR=0.0008, weight_decay=0.005, Dropout(0.3) - **good convergence**

b) Variant 1 (Using Selu and Swish Activation)

SELU had the best test accuracy. Both SELU and Swish had **desirable convergence** for the below parameters:

LR=0.0008, weight_decay=0.005, Dropout = 0.5 before the FC Layer

c) Variant 2 (Removing Maxpool Layers and adding Stride=2)

- 1) LR = 0.0008 **lower test accuracy but a good convergence.**
- 2) LR = 0.001 **better test accuracy and good convergence.**

d) Variant 3 (Adding few dropout Layers)

- 1) PART A: LR=0.0005, weight_decay=0.008, 0.5 dropout before the FC layer – better test accuracy and good convergence
- 2) PART B: LR=0.0005, weight_decay=0.008, dropout after every convolution layer was 0.2, 0.3, 0.4 respectively – **Overfitting**
- 3) LR=0.0005, weight_decay=0.008, dropout after every convolution layer was 0.2, 0.2, 0.2 respectively – **good convergence**

e) Variant 4

Initially the network was extremely underfitting, the only hyperparameter tuned here was the LR.

- 1) lr=0.0008, weight_decay=0.008 **Alarmingly Underfitting**
- 3) lr = 0.005 **good test accuracy but slightly overfitting**
- 4) lr == 0.001 **underfitting**
- 5) lr = 0.002, dropout = 0.1 – **good test accuracy and convergence**