

DL-HW2

Shai Vaknin (034658492), Ev Zisselman (200479483), Yochai Zur (03050991)

January 1, 2017

Github link: <https://github.com/yochaiz/DL-hw>

Branch: hw2

Architecture description

We tried 3 architectures with reduced to 50K parameters, based on Wide-resnet, VGG, NiN. When reducing the parameters, we tried to maintain the basic principles of the architectures.

Best architecture in terms of speed of convergence and final accuracy was a variation of Wide-resnet.

In Wide-resnet there is a widen factor that controls the width (and thus size) of each layer, so we simply tuned it until we got to 50K parameters. We also had to modify the dropout layer probability.

Training procedure

For regularization, we used dropout (0.3) and batch normalization layers. No weight decay.

First we choose optimization method.

Optimization method

We have tried to use Adagrad, Adadelata and Adam. Figures 1,2 shows the results.

Table 1: Summery of optimization method results. No data augmentation.

Method	Accuracy	Epochs to best accuracy
AdaGrad	51.88%	243
AdaDelta	82.37%	188
Adam	83.03%	247

Figure 1: Test and train errors for different optimization methods. As we can see, AdaGrad is very bad and Adam is slightly better than AdaDelta. Thus we decided to stick with Adam. Another interesting point is that AdaGrad still decreases.

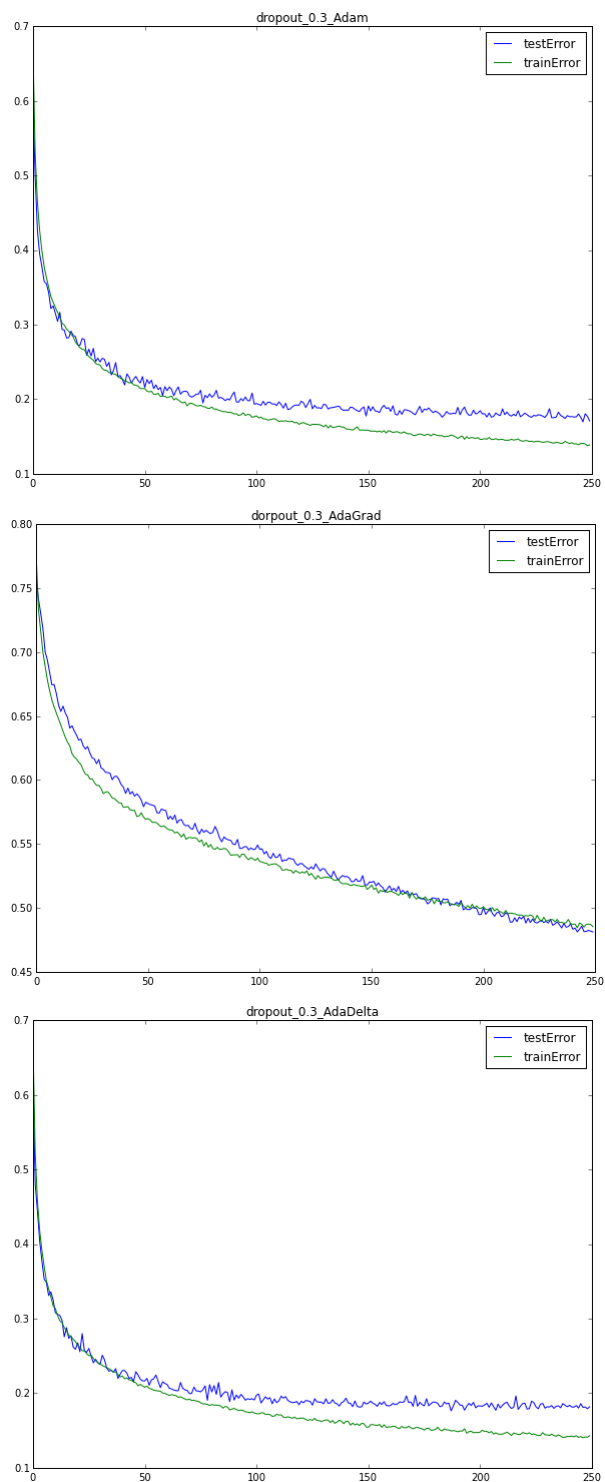
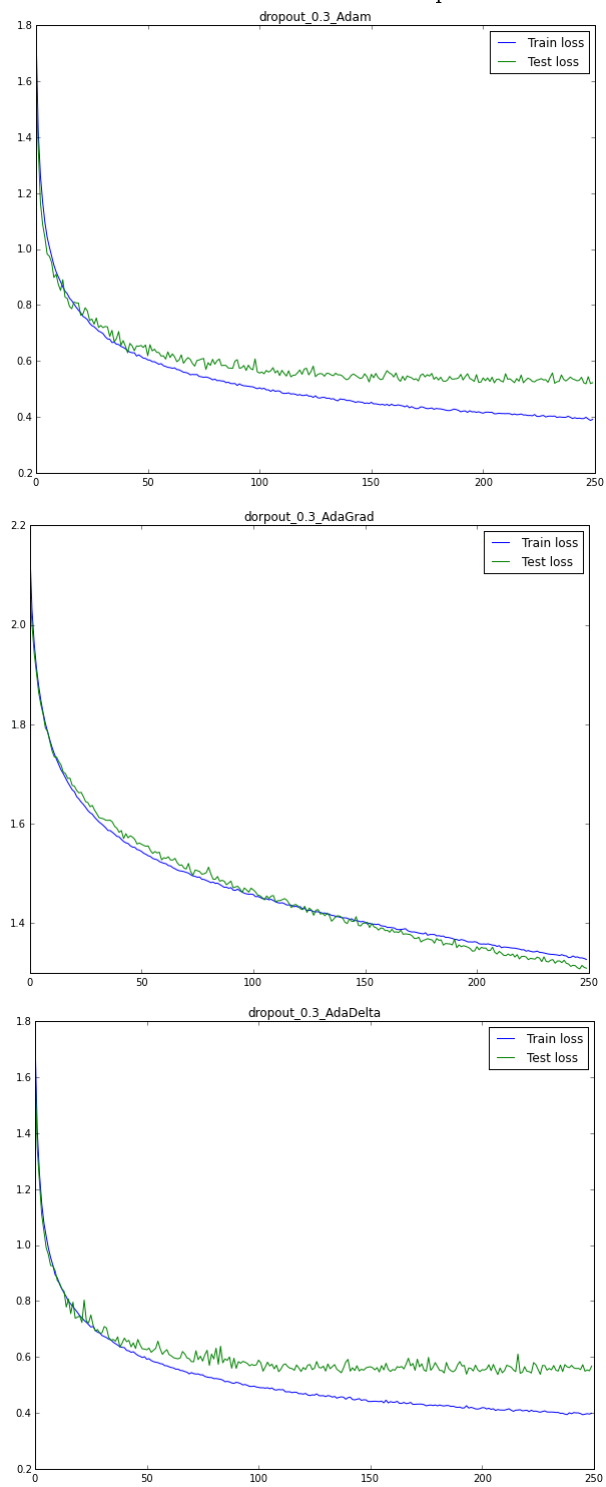


Figure 2: Test and train loss for different optimization methods.



Data augmentation

We tried many combinations of data augmentation, out of crop, horizontal flip and vertical flip.

Best results were achieved with Wide-resnet using only horizontal flip. The rest of the data augmentations performed poorer than without data augmentation at all.

Figures 3,4 show the results for these experiments.

Table 2: Summary of data augmentation results, in all of these experiments we used Adam.

Method	Accuracy	Epochs to best accuracy
vflip	77.17%	299
hflip	84.73%	277
crop + vflip or hflip	55.85%	10

We tried few more data augmentation combinations, but as shown in table, none of the combinations, except horizontal flip only, gave better results.

Figure 3: Test and train errors for different data augmentation methods. Thus we choose hflip.

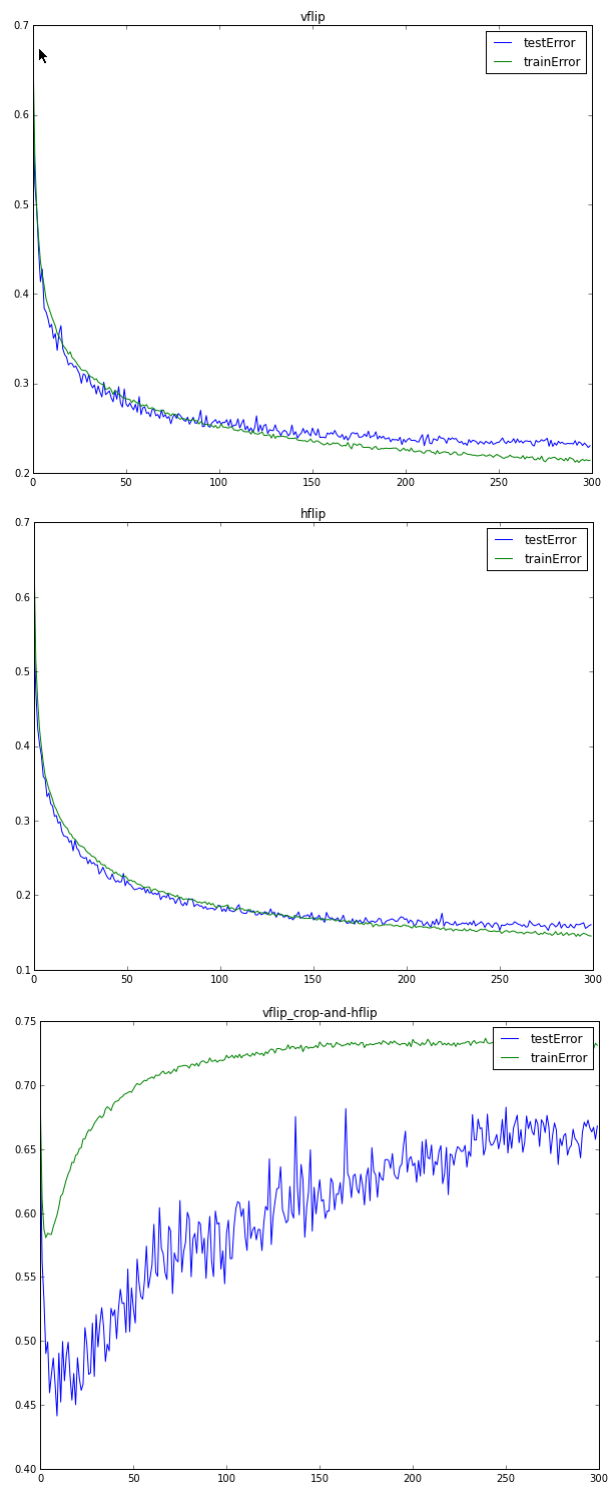
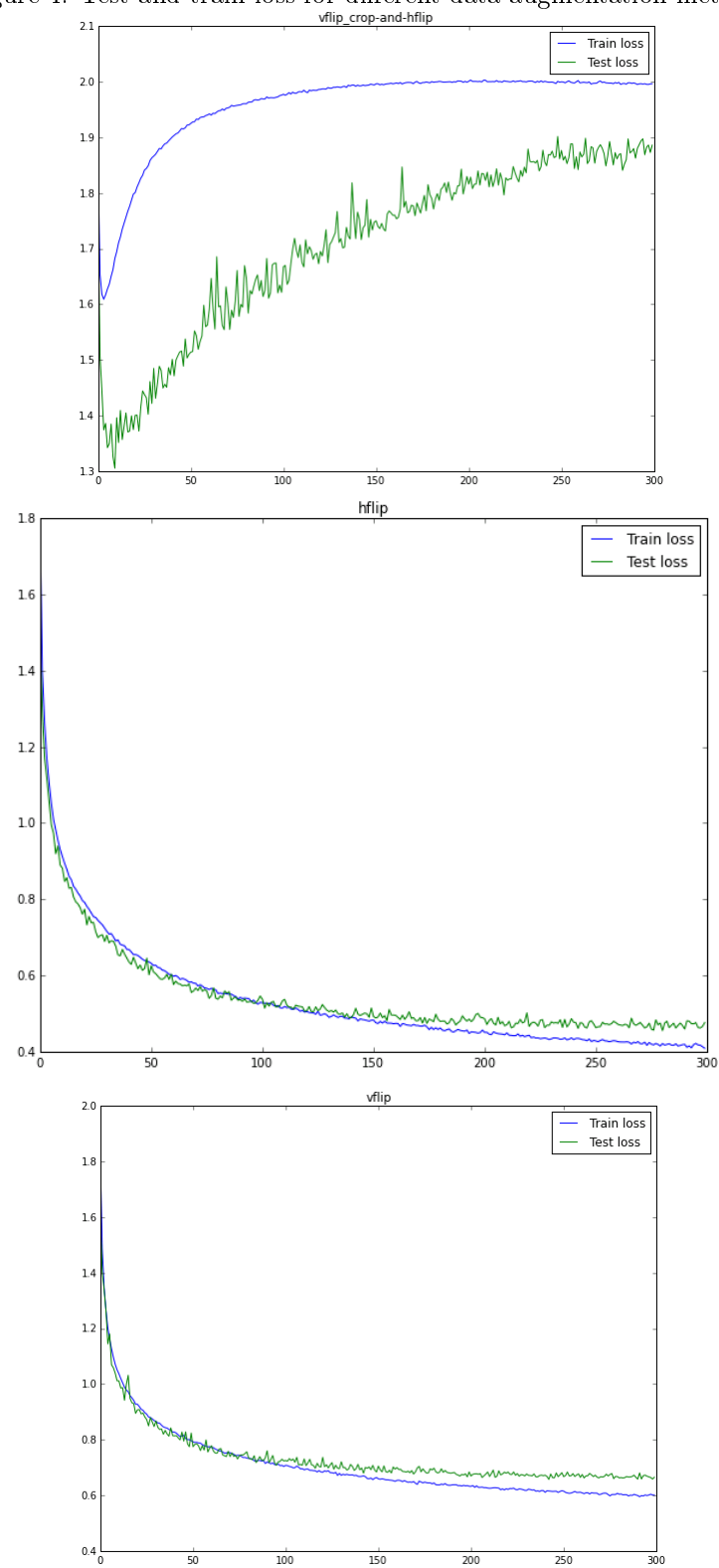


Figure 4: Test and train loss for different data augmentation methods.



Summary:

Adam achieve best results with hflip data augmentation.

Best accuracy we got was ~84.7%.