# DL homework №2 report

alexander.telepov

April 2021

Initially I start experiments with Resnet-like architecture with 8 basic blocks(linearly increasing number of channels from 64 after first convolution to 512 and decreasing feature map sizes) and leakyrelu activation ; I didn't add any augmentations beside normalization(default normalization for Imagenet with subtracting mean and division by std per channel) and horizontal flip. Network had poor performance on validation set and results fluctuate significantly. To overcame overfitting I change default convolutions to combination of depthwise and pointwise convolutions, add dropout layer before linear and add several augmentations(random gaussian noise, random rotation, random shift, random crop and rescaling, random color/contrast jitter etc). These modifications boost performance significantly and I start getting meaningfull accuracy scores. Due to success of EfficientNet paper I tried to use several setting from it : I try to add squeze-and-excitation blocks, but it didn't give any score boost (however I left this in architecture, because for me idea to combine global feature maps information with local information meaningfull); also I try to replace Adam to RMSProp optimizer but it also didn't give any benefit. Rest of experiments were aimed to find best performing learning rate and weight decay. Some other setting of experiments: I keeped batch size 128, I apply kaiming gaussian initialization of linear and convolution layers, I use cosine annealing scheduller, also I use early stopping based on validation performance.

Deep learning insights I gained : 1) I saw that separable convolutions gives some regularization effect while require less computation power; 2) I found that randomly select subsect of augmentation transforms performance score can be increased.

Sources of code are my precious courses homeworks/projects(mostly based on torch documentation).

On plots dependency between train/test loss from epoch number for 3 experiments visualized: distinct line - best performed experiment with random augmentation and bigger weight decay, 2 close lines - experiments with random augmentations applied simultaneously and less weight decay(one of lines with squeeze-and-excitation blocks, other - without). Plots drawn in log-scale.

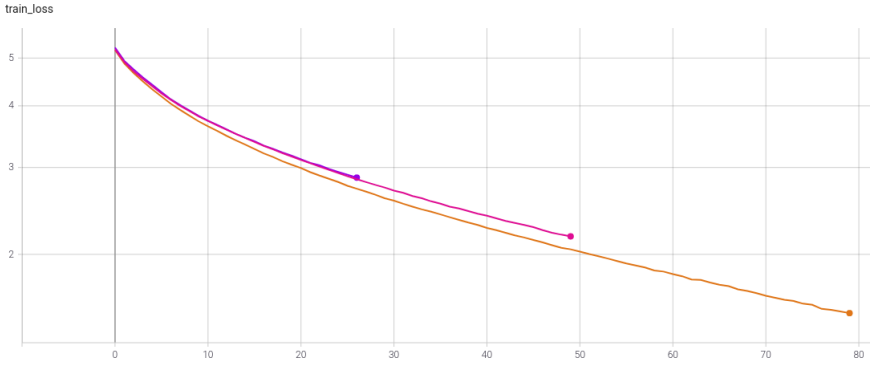


Figure 1: Train loss per epoch

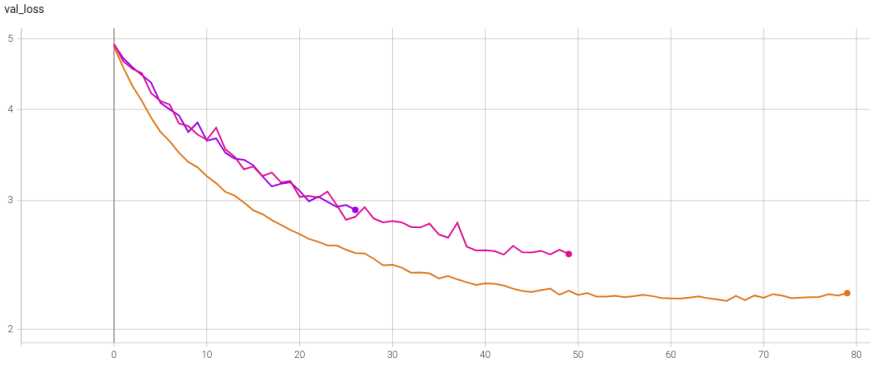


Figure 2: Validation loss per epoch