ATNN H6 Model

Ivan Remus

November 2023

1 Remark

This model hasn't been compiled/traced/scripted to compare its performance nor has it achieved 94% validation accuracy on CIFAR-10.I did save the model and create the inference.py anyway.

2 Model Description

For my model,a block is composed of 2 convolutional layers,one for extracting features without reducing the image dimensions(via padding) and another 1x1 convolution layer and ELU activation function to introduce non-linearities on channels.

The model starts with 2 pairs containing 1 block and 1 max pooling layer in order to extract features from the convolution as well as add more channels and halve the dimensions of the image.After this,we apply another block as well as a 1x1 convolution layer to further increase the information on the channels,following these results to be passed trough 2 more linear layers.

I used ELU instead of ReLU in order to combat the dying ReLU problem and BatchNorm on both linear layers as well the convolution layers from the blocks to put values on a similar scale,as each next layer will expect a certain distribution.Dropout on the linear layers. was also used in order to avoid over-fitting.

3 Forward

We don't use any residual blocks or blocks with multi-branches ,so the forward function in our model is simply taking an input x and passes it from each layer sequentially.

4 Last layer activation

My model uses Softmax in the training pipeline and the CrossEntropy loss.