How to use BP to optimize CNN

Daquan Lin 85610653
ShanghaiTech University
Shanghai, China
lindq@shanghaitech.edu.cn

From the loss, we calculate the parameters' gradient layer by layer, including weight matrix each layer(both Conv layer and dense layer) and the inputs' gradient(it use to calculate the above layer's weight), and update each layer's weight according to the layer's weight gradient.

Optimization methods: such as SGD, SGD+Momentum, Nesterov Momentum, AdaGrad, RMSProp and Adam.

You can see it in below link:

http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture7.pdf

In terms of how to calculate the gradient.

Please look at the below link:

https://github.com/Line290/cs231n_1017spring/blob/master/assignment2/cs231n/layers.py