**Back propagation for convolution**

Input is of dimension (C, H, W), each filter is of size (C, HH, WW), output responding to the filter is of size (H\_out, W\_out). X\_out refers to the output of convo layer, input of next layer. X\_in is the input of convo layer.

For each filter, the weight affects all the value in the output layer. According to chain rule,

Each filter can be calculated independently and with batch processing, it means aggregation of gradient.

**Back propagation for max pooling**

If the input is not maximum in the receptive field, the gradient is 0 since it does not contribute to the output. If the node input is the maximum, its gradient is equal to the responding dout.

**Back propagation for dropout**

If the node is dropped, its gradient is 0 since it does not contribute to the output. Else the gradient is equal to 1/(1-p) \* dout.

**Back propagation with regularization**

The gradient of regularization can be calculated independently.

First part can be calculated using the formula above.