Reading Report

2019.4.15

1 Network in Network(1*1's convolution)

1.1 mlpconv(MLP)

change Relu to

$$f_{i,j,k_1}^1 = max({w_{k1}^1}^T x_{i,j} + b_{k_1}, 0)$$

$$...$$

$$f_{i,j,k_n}^n = max({w_{kn}^n}^T f_{i,j}^{n-1} + b_{k_n}, 0)$$

advantages

1.fusion of different channels
2.the number of channels can be reduced or raised

1.2 global average pooling

Instead of adding fully connected layers on top of the feature maps, The paper take the average of each feature map, and the resulting vector is fed directly into the softmax layer.
1. average pooling is more native to the convolution structure by enforcing correspondences between feature maps and categories. Thus the feature maps can be easily interpreted as categories confidence maps.
2.there is no parameter to optimize in the global average pooling thus overfitting is avoided at this layer. Futhermore, global average pooling sums out the spatial information, thus it is more robust to spatial translations of the input.

2 AlexNet

2.1 max pooling

2.2 Local Response Normalization(LRN)

LRN by Hinton

$$b^i_{x,y} = a^i_{x,y}/(k + \alpha\Sigma^{min(N-1,i+n/2)}_{j=max(0,i-n/2)}(a^j_{x,y})^2)^\beta$$

a=conv1, means [batch,height,width,channel] think input as d of three-dimensional matrices

advantages

The local response normalization layer performs a kind of "lateral inhibition"by normalizing over local input regions

2.3 droupout

prevent overfitting

compare

Network without dropout

$$z^{(l+1)}_i = w^{l+1}_i y^l + b^{l+1}_i$$

$$y^{(l+1)}_i = f(z^{(l+1)}_i)$$

Network with dropout

$$r^{(l)}_j \sim Bernoulli(p)$$

$$y'^{(l)} = r^{(l)} * y'^l$$

$$z^{(l+1)}_i = w^{l+1}_i y'^l + b^{l+1}_i$$

$$y^{(l+1)}_i = f(z^{(l+1)}_i)$$

dropout can be looked as standard regularizers, and compare with vert(2), LASSO, max norm regularization.
while droupout need a lot of time to train.