VGG

The ability to increase depth came only due to the use of small 3x3 filters in all layers.

Architecture

- Preprocessing step is subtracting the mean RGB value, computed on the training set for each example.
- Spatial Pooling - An object is the same when it is zoomed ,tilted, etc. This technique makes our model more robust to minor deformations in features. (Here, it is carried out by 5 Max pooling layers).
- After the stack of convolutional layers, it is followed by 3 FC layers. Last layer being a softmax layer.
- Activation of all layers, except the last is ReLU.

The main difference between VGG and other previous state-of-the-art models was that VGG uses small receptive fields(filter size) throughout the model.

Advantages

- We can say that the stack of two 3x3 conv layers is the same as one 5x5 conv layer ; three 3x3 layers is the same as one 7x7.
- Addition of more layers gives rise to more complex decision boundaries.
- No.of parameters required for three 3x3 conv layers are very less than a single 7x7 layer.
- 1x1 conv layers are used to add complexity to the model.

Training

- Multinomial Logistic Regression; mini-batch of 256; Gradient Descent with momentum.
- Regularisation - L2 for weight decay, dropout for the first 2 FC layers.
- VGG required less epoch for convergence due to (1) implicit regularisation (2) Pre-initialisation of certain layers.
- Weight initialization for some layers were done by using the weights of an already trained shallow model. Biases were initialised with zero.
- Training image size: (1) Use a fixed training scale to crop out 224x224 image. (2) We randomly sample a training scale from a certain range for each training image. It takes into account the occurence of different sizes of objects. This method can be seen as scale jittering.

Local Response Normalisation is used to encourage lateral inhibition.