VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

A RCHITECTURE

When training Convolutional Neural Networks (ConvNets), we feed them a fixed-size image in RGB format, which is 224 x 224 pixels. We apply one pre-processing step i.e. subtracting the mean RGB value, which we calculate from the training set, for each pixel to the image. The image is passed through a stack of convolutional layers, where we use filters with a very small receptive field: 3 3 .

C ONFIGURATIONS

Table 1 gives a list of ConvNet configurations evaluated in this paper, with each column assigning a name to each Net. These will be referred to by their assigned names throughout the article.

D ISCUSSION

Our ConvNet models have been designed differently compared to the top-performing entries in the ILSVRC-2012 and ILSVRC-2013 challenges.

Number of parameters

A,A-LRN such layers have a 7 7 effective receptive field. By using a stack of three 3x3 convolutions, we have been able to gain a multitude of advantages. This includes improved performance, faster training speeds, fewer parameters & computational resources needed, and better overall accuracy. layers instead of a single 7 7 layer? As a starting point, we have implemented three non-linear rectification layers instead of one. This enhances the distinction capability of the decision function.

T RAINING

The ConvNet is usually trained using mini-batch gradient descent, based on the multinomial logistic regression objective, via back-propagation with momentum. This process is in line with Krizhevsky et al's recommendations. We trained with a batch size of 256 and momentum of 0.9, and applied weight decay & dropout to the first two fully-connected layers for regularisation.

The original learning rate was set to 10−2, however, when the validation accuracy stopped increasing it was decreased by a factor of 10. This process went on for 3 times and the learning ceased at 370K iterations.

T ESTING

During testing, a ConvNet is applied to an input image and classified accordingly. The image is first changed to a universal size (Q) which may not match the training scale (S). This process is known as isotropic rescaling.