VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

ABSTRACT

Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3 × 3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers.

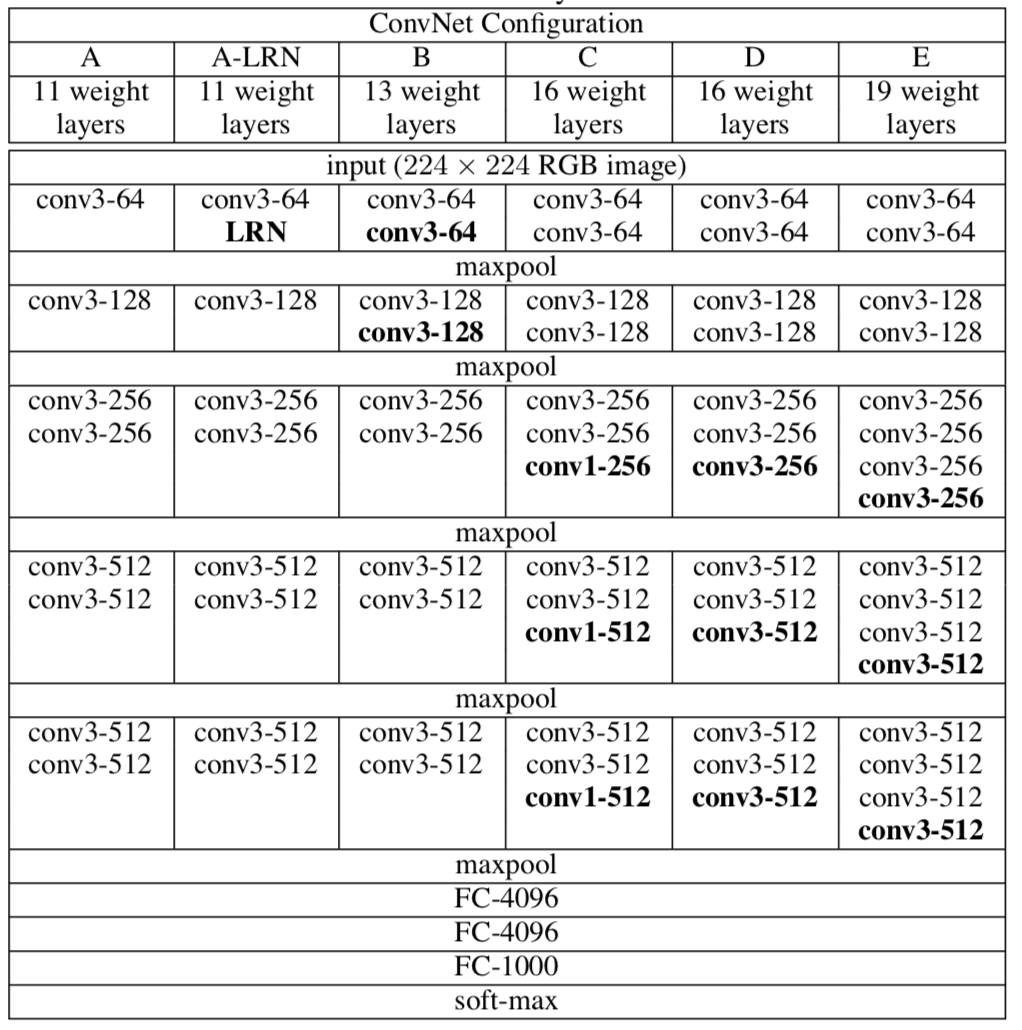
1 INTRODUCTION

In this paper, we address another important aspect of ConvNet architecture design – its depth. To this end, we fix other parameters of the architecture, and steadily increase the depth of the network by adding more convolutional layers, which is feasible due to the use of very small (3 × 3) convolution filters in all layers.

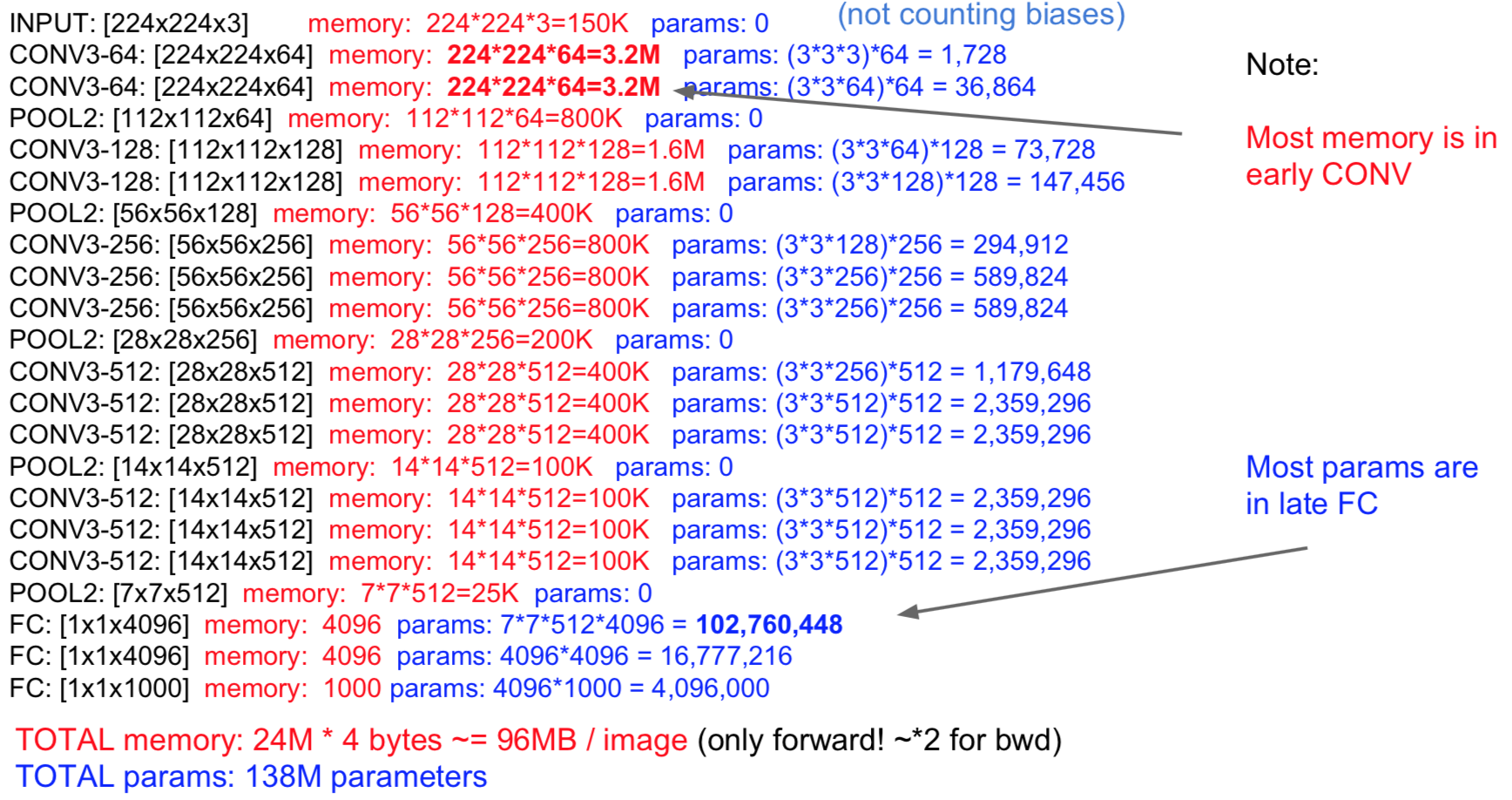
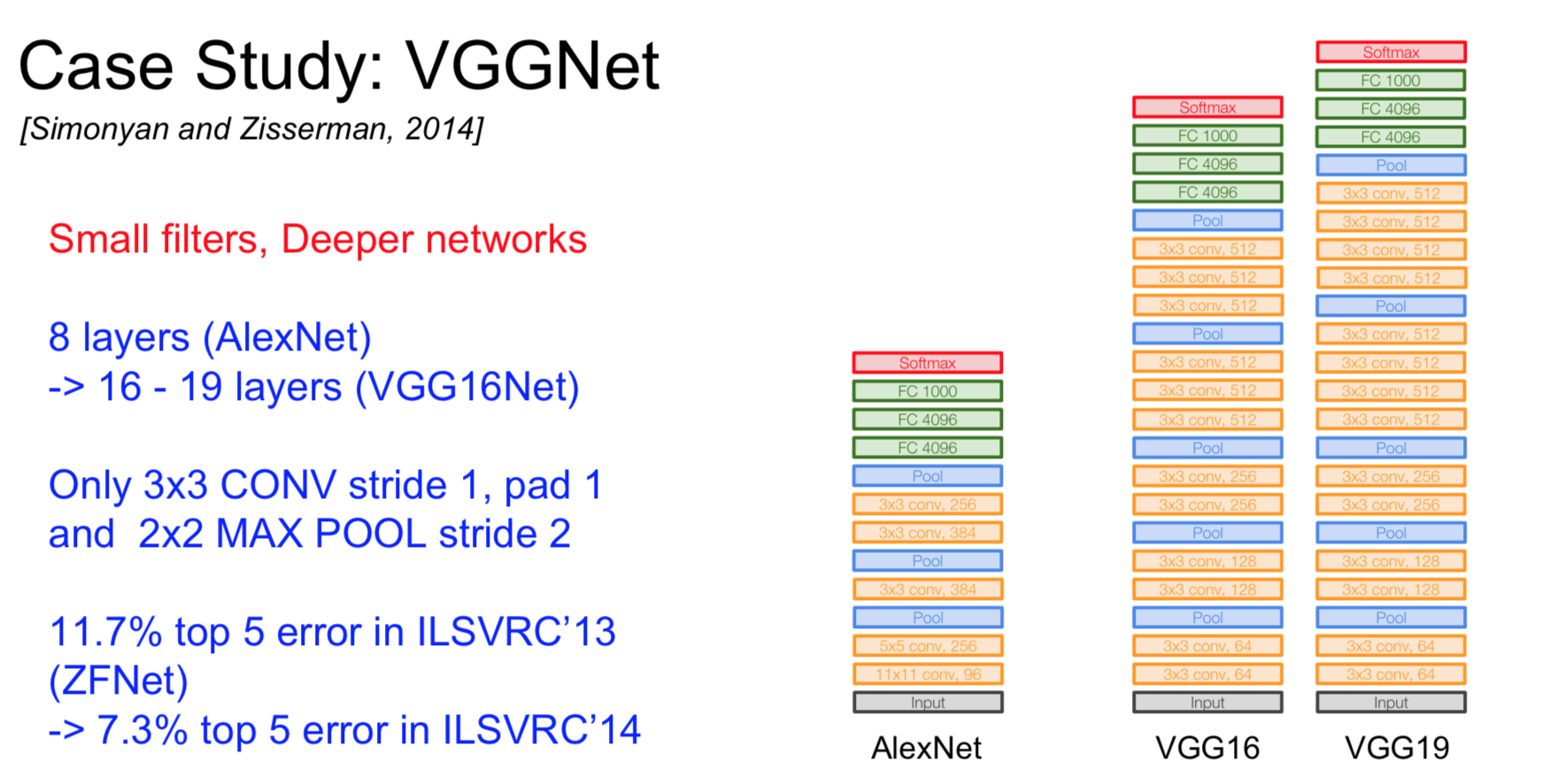
2 CONVNET CONFIGURATIONS

2.1 ARCHITECTURE

2.2 CONFIGURATIONS



2.3 DISCUSSION



3 CLASSIFICATION FRAMEWORK

3.1 TRAINING

We consider two approaches for setting the training scale S. The first is to fix S, which corresponds to single-scale training (note that image content within the sampled crops can still represent multi- scale image statistics).

The second approach to setting S is multi-scale training, where each training image is individually rescaled by randomly sampling S from a certain range [Smin,Smax] (we used Smin = 256 and Smax = 512).

3.2 TESTING

3.3 IMPLEMENTATION DETAILS

4 CLASSIFICATION EXPERIMENTS

**Dataset.**

4.1 SINGLE SCALE EVALUATION

4.2 MULTI-SCALE EVALUATION

4.3 MULTI-CROP EVALUATION

4.4 CONVNET FUSION

4.5 COMPARISON WITH THE STATE OF THE ART

5 CONCLUSION