Squeeze-and-Excitation Networks

SUPPLEMENTARY MATERIAL

1. Training Curves on ImageNet

The training curves for the four plain architectures, i.e., ResNet-152, ResNeXt-50, BN-Inception and Inception-ResNet-v2, and their SE counterparts are respectively depicted in Fig. 2, illustrating the consistency of the improvement yielded by SE blocks throughout the training process.

2. Details of SENet-154

SENet-154 is constructed by integrating SE blocks to a modified version of the 64×4d ResNeXt-152 that extends the original ResNeXt-101 [3] by following the block stacking of ResNet-152 [1]. More differences to the design and training (beyond the use of SE blocks) were as follows: (a) The number of first 1×1 convolutional channels for each bottleneck building block was halved to reduce the computation cost of the network with a minimal decrease in performance. (b) The first 7×7 convolutional layer was replaced with three consecutive 3×3 convolutional layers. (c) The down-sampling projection 1×1 with stride-2 convolution was replaced with a 3×3 stride-2 convolution to preserve information. (d) A dropout layer (with a drop ratio of 0.2) was inserted before the classifier layer to prevent overfitting. (e) Label-smoothing regularisation (as introduced in [2]) was used during training. (f) The parameters of all BN layers were frozen for the last few training epochs to ensure consistency between training and testing. (g) Training was performed with 8 servers (64 GPUs) in parallelism to enable a large batch size (2048) and initial learning rate of 1.0.

3. Four Class Examples

To understand how the self-gating *excitation* mechanism operates in practice, we study example activations from the SE-ResNet-50 model and examine their distribution with respect to different classes at different blocks in Section "The role of Excitation". We sample four classes from the ImageNet dataset that exhibit semantic and appearance diversity, namely *goldfish*, *pug*, *plane* and *cliff*. The example images from the four classes are shown in Fig. 1. See that section for detailed analysis.

(a) goldfish (b) pug (c) plane (d) cliff

Figure 1: Example images from the four classes of ImageNet used in Section "The role of Excitation".

References

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016.

[3] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *CVPR*, 2017.

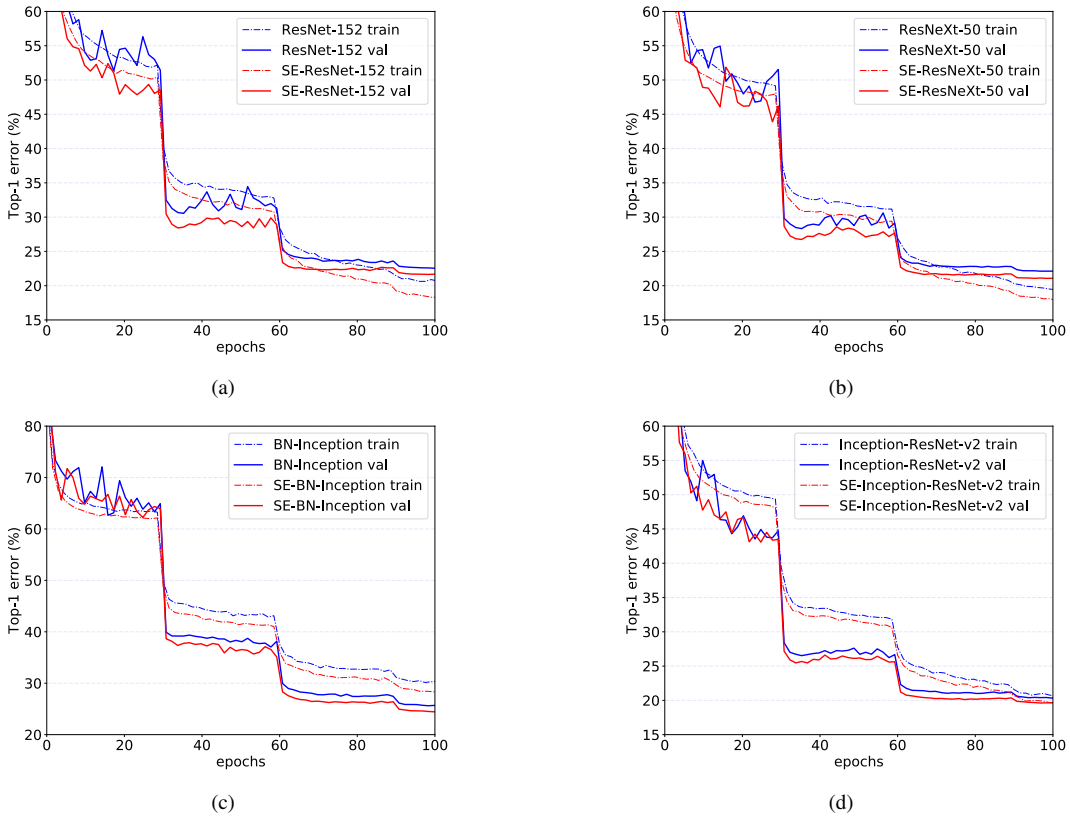

Figure 2: Training curves on ImageNet. (a): ResNet-152 and SE-ResNet-152; (b): ResNeXt-50 and SE-ResNeXt-50; (c): BN-Inception and SE-BN-Inception; (d): Inception-ResNet-v2 and SE-Inception-ResNet-v2.