A1 Supplementary Materials

Visualization of network internal representations To gain additional qualitative understanding of the pooling methods we are considering, we use the popular t-SNE [39] algorithm to visualize embeddings of some internal feature responses from pooling operations. Specifically, we again use four networks (one utilizing each of the selected types of pooling) trained on the CIFAR10 training set (see Sec. 5 for architecture details used across each network). We extract feature responses for a randomly chosen 800-image subset of the CIFAR10 test set at the first (i.e., earliest) and second pooling layers of each network. These feature response vectors are then embedded into 2-d using t-SNE; see Figure A1.

The first row shows the embeddings of the internal activations immediately after the first pooling operation; the second row shows embeddings of activations immediately after the second pooling operation. From left to right we plot the t-SNE embeddings of the pooling activations within networks that are trained with average, max, gated max-avg, and (2 level) tree pooling. We can see that certain classes such as "0" (airplane), "2" (bird), and "9" (truck) are more separated with the proposed methods than they are with the conventional average and max pooling functions. We can also see that the embeddings of the second-pooling-layer activations are generally more separable than the embeddings of first-pooling-layer activations.

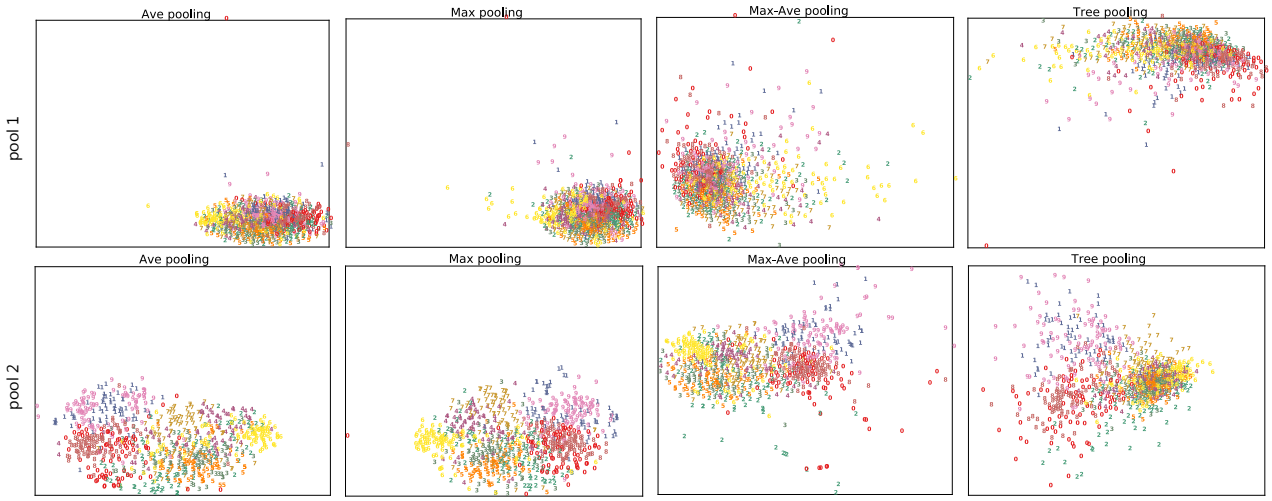

Figure A1: t-SNE embeddings of the output responses from different pooling operations on the CIFAR10 test set (with classes indicated). From left to right: average, max, gated max-avg, and (2 level) tree pooling. The first and the second rows show the first and the second pooling layers, respectively. Best viewed in color.

Table A1: Here we provide explicit statement of the experimental conditions (specifically, network layer configurations) explored in Tables 1, 2, and 4. We list all conv-like layers and pool-like layers, but ReLUs are suppressed to lighten the amount of text; these follow each standard conv layer. Also, all network configurations incorporate deep supervision after each standard convolution layer; this is also suppressed for clarity. We bold the changes made to the baseline DSN layer configuration. We now describe the meaning of entries in the table. Each column in the table lists the sequence of layer types used in that network configuration. When a row cell spans multiple columns (i.e. configurations), this indicates that the layer type listed in that cell is kept the same across the corresponding network configurations. Thus, every network in our experiments begins with a stacked pair of 3x3 (standard) conv layers followed by a 1x1 mlpconv layer. For a specific example, let us consider the network configuration in the column headed "mixed max-avg" - the sequence of layers in this configuration is: 3x3 (standard) conv, 3x3 (standard) conv, 1x1 mlpconv, **3x3 mixed max-avg pool**, 3x3 (standard) conv, 3x3 (standard) conv, 1x1 mlpconv, **3x3 mixed max-avg pool**, 3x3 (standard) conv, 3x3 (standard) conv, 1x1 mlpconv, 1x1 mlpconv, 8x8 global vote (cf. [24]) (we again omit mention of ReLUs and deep supervision). CIFAR100 uses (2 level) tree+max-avg; CIFAR10 uses (3 level) tree+max-avg. As a final note: for the MNIST experiments only, the second pooling operation uses 2x2 regions instead of the 3x3 regions used on the other datasets.

Network layer configurations reported in Tables 1, 2, and 4 of the main paper.					
DSN (baseline)	mixed max-avg	gated max-avg	2 level tree pool	3 level tree pool	tree+gated max-avg pool
3x3 (standard) conv					
3x3 (standard) conv					
1x1 mlpconv					
3x3 maxpool	**3x3 mixed max-avg**	**3x3 gated max-avg**	**3x3 2 level tree pool**	**3x3 3 level tree pool**	**3x3 2/3 level tree pool**
3x3 (standard) conv					
3x3 (standard) conv					
1x1 mlpconv					
3x3 maxpool	**3x3 mixed max-avg**	**3x3 gated max-avg**	3x3 maxpool	3x3 maxpool	**3x3 gated max-avg**
3x3 (standard) conv					
3x3 (standard) conv					
1x1 mlpconv					
1x1 mlpconv					
8x8 global vote					