Supplementary Material

This is the supplementary material for “Hierarchical Model Compression via Shape-Edge Representation of Feature Maps — an Enlightenment from the Primate Visual System”. In Section I, we completely visualize the blurred images and corresponding edge images in Fig. 2(c) and (d) of the main paper. In Section II, we show the experimental configurations in detail. In Section III, we further display the pruning results of HPSE. In Section IV, we further demonstrate and analyze the visualization of HPSE.

I. Rethinking the importance of shape and edge

In Section III of the main paper, we adopt the blurred images to illustrate the influence of edge information (as shown in Fig. 2(b-e) of the main paper). Here, we supplement all the blurred images and their edge images in detail. As shown in Fig. 1, the labels of objects along the column are “Monarch butterfly”, “Junco”, “Castle”, “Tiger cat”, “Panda”, “Border collie”, “Pay-phone”, “Hartebeest” and “Conch”, respectively. The images in the 1-st column are original images. The images in the 2-nd to 5-th column are the blurred images, which is processed by the circular mean filter with radiuses of 5, 10, 15 and 20, respectively. The corresponding edge images are shown in the 6-th to 10-th column. The edge information is extract by the Sobel operator, and the thresholds of different images are set to 0.02, 0.015, 0.015, 0.015, 0.015, 0.02, 0.02, 0.015 and 0.02, respectively. Finally, Fig. 2(e) of the main paper shows the probability change of the ground truth of blurred images with respect to different radii. As shown in Fig. 2(e) of the main paper, the more blurred the image, the fewer the edge features, the lower the probability of the ground truth.

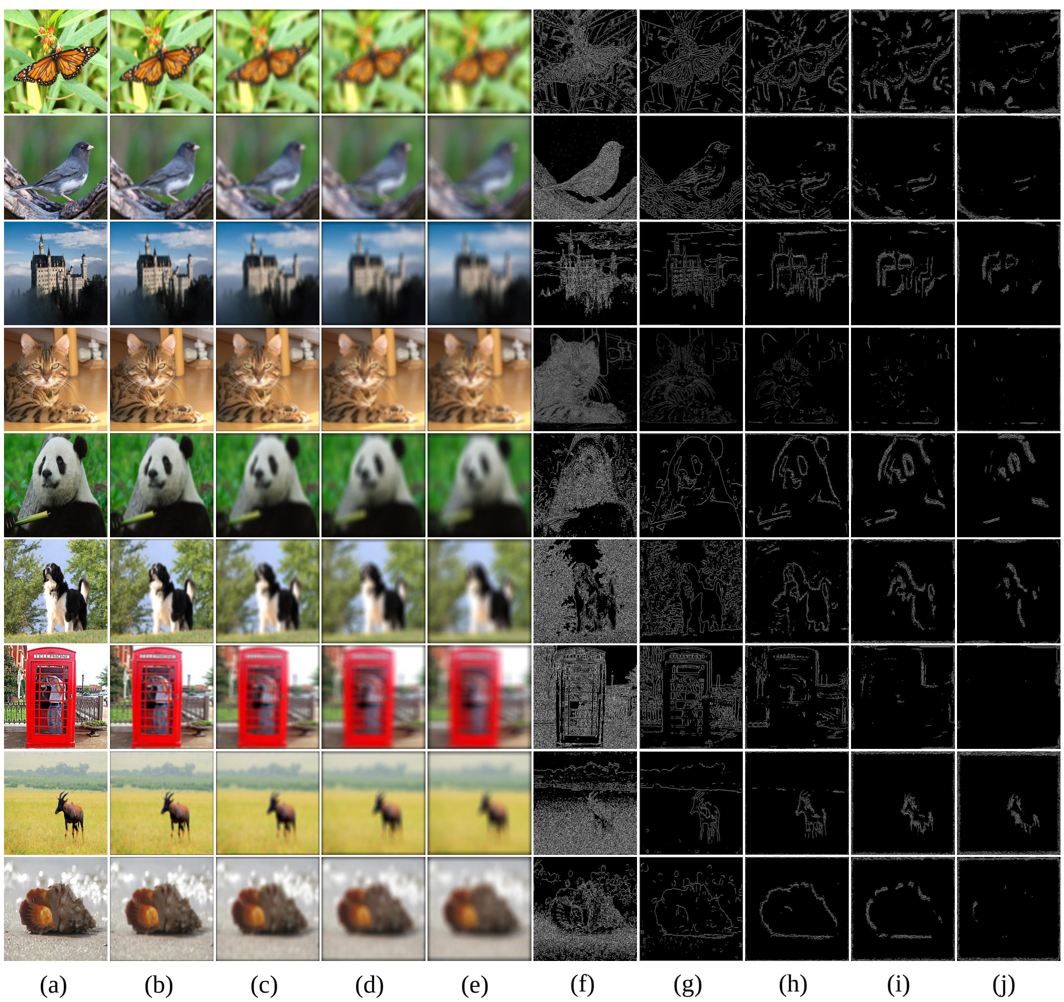


Fig. 1. Blurred images and edge images. (a) Original images. (b) Blurred images (radius is 5). (c) Blurred images (radius is 10). (d) Blurred images (radius is 15). (e) Blurred images (radius is 20). (f) Edge images of original images. (g) Edge images of Blurred images when the radius is 5. (h) Edge images of Blurred images when the radius is 10. (i) Edge images of Blurred images when the radius is 15. (j) Edge images of Blurred images when the radius is 20.

II. Experimental Configurations

In this section, the experiment setups for restoring the network accuracy are summarized in the table.

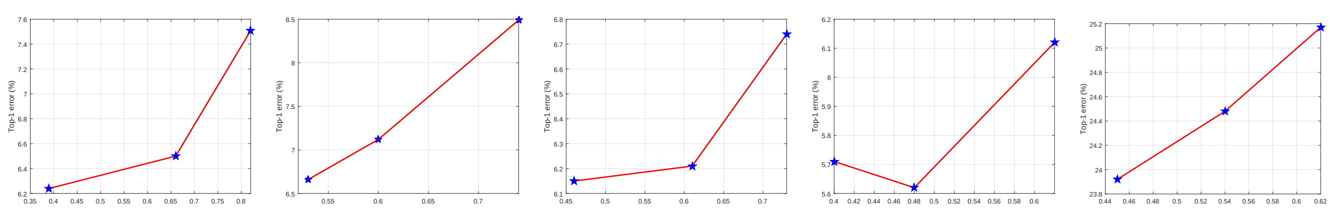
Network	Pre-trained	Epoch	Momentum	Weight decay	Batch size	Initial learning rate	Learning rate
VGG-16	✓	30	0.9	0.0005	128	0.01	$\begin{cases} 0.01 & if \quad 1 \leq epoch < 10 \\ 0.001 & if \quad 10 \leq epoch < 20 \\ 0.0001 & if \quad 20 \leq epoch \leq 30 \end{cases}$
ResNet-56	✓	100	0.9	0.0001	128	0.01	$\begin{cases} 0.01 & if \quad 1 \leq epoch < 30 \\ 0.001 & if \quad 30 \leq epoch < 80 \\ 0.0001 & if \quad 80 \leq epoch \leq 100 \end{cases}$
ResNet-110	✓	100	0.9	0.0001	128	0.01	$\begin{cases} 0.01 & if \quad 1 \leq epoch < 30 \\ 0.001 & if \quad 30 \leq epoch < 80 \\ 0.0001 & if \quad 80 \leq epoch \leq 100 \end{cases}$
DenseNet	✓	100	0.9	0.0001	128	0.01	$\begin{cases} 0.01 & if \quad 1 \leq epoch < 30 \\ 0.001 & if \quad 30 \leq epoch < 80 \\ 0.0001 & if \quad 80 \leq epoch \leq 100 \end{cases}$
ResNet-50	✓	50	0.9	0.0001	512	0.01	$\begin{cases} 0.01 & if \quad 1 \leq epoch < 25 \\ 0.001 & if \quad 25 \leq epoch < 45 \\ 0.0001 & if \quad 45 \leq epoch \leq 100 \end{cases}$
VGG-16	✗	200	0.9	0.0005	128	0.1	$\begin{cases} 0.1 & if \quad 1 \leq epoch < 100 \\ 0.01 & if \quad 100 \leq epoch < 150 \\ 0.001 & if \quad 150 \leq epoch \leq 200 \end{cases}$
ResNet-56	✗	200	0.9	0.0001	128	0.1	$\begin{cases} 0.1 & if \quad 1 \leq epoch < 100 \\ 0.01 & if \quad 100 \leq epoch < 150 \\ 0.001 & if \quad 150 \leq epoch \leq 200 \end{cases}$
ResNet-110	✗	200	0.9	0.0001	128	0.1	$\begin{cases} 0.1 & if \quad 1 \leq epoch < 100 \\ 0.01 & if \quad 100 \leq epoch < 150 \\ 0.001 & if \quad 150 \leq epoch \leq 200 \end{cases}$
DenseNet	✗	200	0.9	0.0001	128	0.1	$\begin{cases} 0.1 & if \quad 1 \leq epoch < 100 \\ 0.01 & if \quad 100 \leq epoch < 150 \\ 0.001 & if \quad 150 \leq epoch \leq 200 \end{cases}$
ResNet-50	✗	150	0.9	0.0001	512	0.1	$\begin{cases} 0.1 & if \quad 1 \leq epoch < 80 \\ 0.01 & if \quad 80 \leq epoch < 120 \\ 0.001 & if \quad 120 \leq epoch \leq 150 \end{cases}$

TABLE I
EXPERIMENT SETUPS. "✗" AND "✓" REPRESENT WHETHER TO PRUNE THE DNN FROM SCRATCH OR NOT, RESPECTIVELY.

In Table I, we show the settings for restoring the network accuracy in detail, including, the configurations of retraining with pre-trained models and retraining from scratch, respectively.

III. Pruning Results

We demonstrate the accuracy change with respect to the different compression rates in this section.



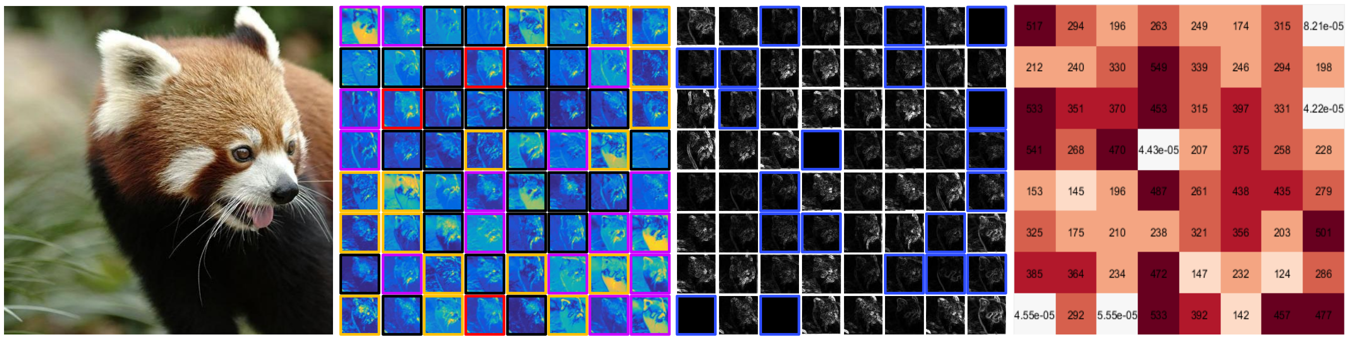
(a) VGG-16 (b) ResNet-56 (c) ResNet-110 (d) DenseNet (e) ResNet-50

Fig. 2. Top-1 error change *w.r.t.* compression rate. x- and y-axis represent FLOPs drop and Top-1 error, respectively. (a) to (e) represent the result of VGG-16, ResNet-56, ResNet-110, DenseNet and ResNet-50, respectively.

For each network, we compress the network with three compression ratios, and the other settings are the same as those in Section 2. For VGG-16, we set δ to 20%, 45% and 60%, respectively; For ResNet-56, we set δ to 35%, 45% and 53%, respectively; For ResNet-110, we set δ to 30%, 40% and 50%, respectively; For DenseNet, we set δ to 35%, 40% and 50%, respectively; For ResNet-50, we set δ to 30%, 37% and 43%, respectively. The relationship between the Top-1 error change and FLOPs drop is shown in Fig. 2. Except DenseNet, the errors of other networks increase with the increase of FLOPs drop. When δ are 30% and 40%, the accuracies of pruned DenseNet are similar.

IV. VISUALIZATION

In Section 6.2 of the main paper, we visualize the HPSE. Here, we use the heat map to show the edge content and further explain the visualization results.



(a) Input Image (b) Shape Feature Map (c) Edge Feature Map (d) Heat Map

Fig. 3. Visualization. (a), (b), (c), and (d) represent input, SFM, EFM, and heat map, respectively. FMs are grouped into 4 groups, and marked by pink, black, yellow, red boxes in (b). The filters corresponding to EFM marked by the blue boxes in (c) are pruned. (d) records the edge-feature content in the EFM.

As shown in Fig. 3, we use the first layer of the first basic block of ResNet-50 to visualize HPSE. We cluster the SFMs corresponding to 64 filters of this layer into 4 groups (marked by 4 different colors in Fig. 3(b)), the SFMs in different group highlight different features. For example, The SFMs in the red boxes emphasize the tongue, beard, ears, and eyes of the lesser panda, while the SFMs at the 1-st row and the 1-st column (denoted as (1,1)) and (6,8) in the purple boxes emphasize the body of the lesser panda and ignore the top of the head. The EFMs are shown in Fig. 3(c), and the pruned EFMs are marked by blue boxes. All the EFMs without edge features are removed (EFMs corresponding to the white heat maps in Fig. 3(d)). In addition, the EFM at (3,2) is pruned, while the EFMs with less edge content are retained (e.g., (2,2), (4,2), etc.). To explain, EFM at (3,2) contains the least edge information when compared with other EFMs in its group. Hence, the grouping stage of HPSE is necessary and important.