Deep Texture Recognition via Exploiting Cross-Layer Statistical Self-Similarity [Supplementary Materials]

1. Parameter influence analysis

The experiments on how the hyper-parameter setting influences the performance of the proposed method are conducted as follows. We consider two hyper-parameters. (i) The channel number of $\mathcal{V}_t$, *i.e.* Z. In our classification experiment, we set $Z = 16$. In the influence test, we try another two values: $Z = 8$ and $Z = 32$, and their corresponding results are listed in Tab. 1 for comparison. It can be seen that increasing Z will lead to improvement on FMD and KTH. (ii) The number of $\boldsymbol{Y}_\ell$s, *i.e.* L. It will affect the final descriptor length. In the paper, we set $L = 1$ for ResNet18 backbone. In the influence test, we also try other values: $L = 3, 5, 7, 9$, and their results are listed in Tab. 2 for comparison. It can be seen that enlarging L will lead to slight performance improvement. Note that when increasing Z or L, the model complexity will increase accordingly. We did not tuneup these hyper-parameters but just set them to common values which suffice to lead to good results.

Table 1. Classification accuracy (%) using different values of Z, with the ResNet18 backbone used.

Dataset	$Z = 8$	$Z = 16$	$Z = 32$
FMD	82.3 ± 0.6	82.5 ± 0.7	82.9 ± 0.7
KTH	85.4 ± 1.1	85.4 ± 1.1	86.1 ± 1.2

Table 2. Classification accuracy (%) using different values of L, with the ResNet18 backbone used.

Dataset	$L = 1$	$L = 3$	$L = 5$	$L = 7$	$L = 9$
FMD	82.5 ± 0.7	82.4 ± 0.8	82.6 ± 0.7	82.7 ± 0.7	82.7 ± 0.8
KTH	85.4 ± 1.1	85.5 ± 1.1	85.7 ± 1.3	85.8 ± 1.2	85.9 ± 1.2

2. Comparison to a deeper ResNet w/o CLASS

We construct a ResNet50 baseline with very close size as our mode, by adding a basic residual block to the front of ResNet50, which is denoted by ResNet50+. Its number of parameters is slightly larger than our CLASS-Net, *i.e.* 24.7M vs. 23.7M. See Table 3 for the results and comparison. Our CLASS-Net noticeably outperforms ResNet50+. Such results demonstrate that, the performance gain of CLASS-Net is not from the increased module size but from the mechanism of the CLASS module.

Table 3. Performance comparison of CLASS-Net and ResNet50+ in terms of classification accuracy (%).

	DTD	KTH	FMD	MINC	GTOS
CLASS-Net	74.0±0.5	87.7±1.3	86.2±0.9	84.0±0.6	85.6±2.2
Resnet50+	68.8±0.4	81.9±1.8	72.6±1.5	80.9±0.3	81.4±2.5

3. Visualizing cross-layer SSS

See Fig. 1 for an illustration on the log-log fitting done on certain feature tensors in DBC pooling on four texture images of two classes. For better illustration with more points, we set S to a larger value and retrained the model. Each red/blue square denotes the receptive region related to the feature points whose log-log behaviors are shown. As the points lie well on a line in Fig. 1, it indicates that the cross-layer SSS holds well and is captured by our model.

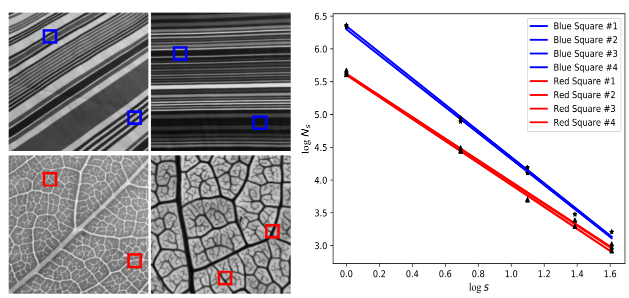

Figure 1. Illustration of log-log fitting in our DBC pooling.

4. Results of removing GAP

See Table 4 for the comparison of our method to a baseline constructed via removing the GAP while flattening the input features directly. Without GAP, there is certain performance decrease which varies on different datasets. Indeed, the CLASS module generates the description by examining the variations of feature maps, which can be roughly viewed as ‘high-pass’, while GAP characterizes feature maps via average, which can be viewed as ‘low-pass’. These results show that the descriptions generated by CLASS module encode aspects that are different from GAP, providing complementary information.

Table 4. Performance comparison of CLASS-Net w/ and w/o GAP on ResNet18 backbone, in terms of classification accuracy (%).

	DTD	KTH	FMD	MINC	GTOS
CLASS-Net	71.5±0.4	85.4±1.1	82.5±0.7	80.5±0.6	84.3±2.2
w/o GAP	66.0±0.5	84.8±1.2	79.3±1.0	79.5±0.8	83.4±2.0

5. Layer Contribution in DBC Pooling

Recall that our DBC pooling uses 5-layer feature maps for CLASS module with the RseNet18 backbone. It is interesting to check the performance change without one-layer feature map. As shown in Table 5, the performance w/o the k^{th} layer in DBC pooling decreases for all k. The decrease amount is similar for different k.

Table 5. Performance of CLASS-Net w/o the kth layer, in terms of classification accuracy (%), with the ResNet18 backbone used.

	w/o 1$^{\text{st}}$	w/o 2$^{\text{nd}}$	w/o 3$^{\text{rd}}$	w/o 4$^{\text{th}}$	w/o 5$^{\text{th}}$	Full
FMD	81.4±0.9	81.5±1.0	81.7±1.1	81.4±0.9	81.7±1.0	82.5±0.7
DTD	70.8±1.0	70.8±0.9	70.6±1.0	70.6±0.9	70.8±0.8	71.5±0.4