Attn-HybridNet: Improving Discriminability of Hybrid Features with Attention Fusion (Supplementary Material)

Sunny Verma, Chen Wang, Liming Zhu, *Member, IEEE* and Wei Liu, *Senior Member, IEEE*

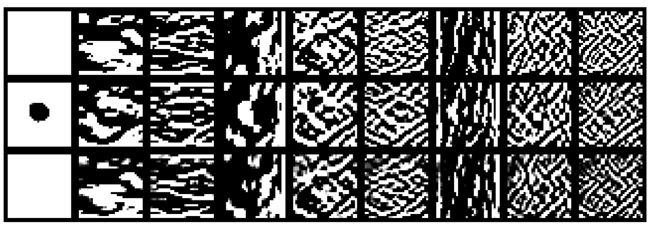
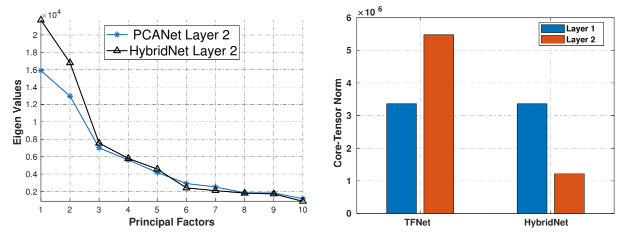


Fig. S1: Convolution responses from the MPCANet. The output is less diversified than TFNet (collapsed in the first column), but the visual resemblance is partially observable.



(a) Comparison of eigenvalues between networks

(b) Comparison of core-tensor strength between networks

Fig. S2: Comparison of factorization strength in Layer 2 of the PCANet, TFNet and *HybridNet* on CIFAR-10 dataset

I. Supplementary Comparisons between different networks.

a) Comparison of convolution outputs from MPCANet: We present convolution responses of MPCANet [1] in Fig. S1 compared to convolution outputs from the PCANet and the TFNet presented in the main draft. It is visible that the convolution responses in MPCANet are less diversified than those of TFNet, although the visible resemblance is still observable. We believe that this is because of how the convolution filters and convolution operations are designed in the MPCANet. The MPCANet also employs tensorized-convolution but differs from the TFNet in two ways: 1) design of kernels and 2) convolution operation. Technically, the convolution kernel and convolution operation in MPCANet are merged as conventional tensor operations by performing an n-mode product of the patch-tensor and factor matrices. We speculate that MPCANet might be more beneficial on datasets where the $m >> r$, where m is the dimensions of the original data, and r is the reduction required.

b) Comparison of Learning Schemes between PCANet, TFNet, and HybridNet: The plot in Fig. S2(a) compares the eigenvalues obtained in the second layer of the PCANet and the *HybridNet*. We exclude eigenvalues from the first layer of these networks as they completely overlap, which is their expected behavior. The leading eigenvalues obtained in the second layer of the *HybridNet* has a much higher magnitude than the corresponding eigenvalues in PCANet. This demonstrates that in the second layer of the *HybridNet*, the data has much diversity than that in PCANet.

Similarly, we compare the core-tensor strength in different layers of the *HybridNet* and the TFNet in Fig S2(b). In this regard, we plot the norm of the core-tensors for both the networks as they are analogous to eigenvalues, and the norm signifies the compression strength of the factorization [2]. Evidently, the norm of the core-tensor in the second layer of *HybridNet* is much lower than that of the TFNet, suggesting relatively higher factorization strength in *HybridNet*.

II. Supplementary Classification Comparisons on CIFAR-10 and CuReT Datatsets.

a) Experimental Settings:

1 CUReT texture dataset [3], consists images with dimensions 200×200 for 61 texture categories, where each category has images of the same material with different poses, illumination conditions, specularity, shadowing, and surface normals. Following the standard procedure in [3], [4] a subset of 92 cropped images were taken from each category and randomly partitioned into train and test sets with a split ratio of 50%. The classification results are averaged over 10 different trails. We set, $L_1, L2$ = [9, 8]; k_1, k_2 = [5, 5]. The overlapping block size region (B) was kept as [50, 50] during feature pooling.

2 CIFAR-100 [5] dataset closely follows CIFAR-10 dataset and consists 50k training and 10k testing images roughly distributed among 100 categories. We use the same experimental setup as in CIFAR-10 on this dataset.

A. Performance Comparisons

a) Performance on the CUReT texture classification: The *Attn-HybridNet* achieves the lowest classification error among all the networks i.e., the PCANet, the TFNet, and the *HybridNet*¡ However, it achieves a marginally higher classification error compared to state of the art in [6] which is only 0.5%.

TABLE SI: Classification performance on the CUReT dataset.

Methods	Error (%)
Textons [7]	1.50
BIF [8]	1.40
Histogram [9]	1.00
ScatNet [6]	**0.20**
PCANet [4]	0.84
TFNet [10]	0.96
HybridNet (proposed)	0.81
Attn-HybridNet (proposed)	0.72

Methods	**#Depth**	**#Params**	**Error**
ResNet reported in-[11]	110	1.7M	37.80
DenseNet-BC reported in-[12]	250	15.3M	**19.64**
PCANet [4]	3	7	53.00
TFNet [10]	3	7	53.63
HybridNet (proposed)	3	7	49.87
Attn-HybridNet (proposed)	4	42.2k	47.44

TABLE SII: Classification Error on CIFAR-100 dataset with no data augmentation.

b) Quantitative Performance on CIFAR-100 Dataset: The proposed *Attn-HybridNet* reduces the error by 5.91% on the CIFAR-100 dataset compared to the PCANet. Additionally, it further reduces the error by 4.87% in comparison to the *HybridNet*.

REFERENCES

[1] J. Wu, S. Qiu, R. Zeng, Y. Kong, L. Senhadji, and H. Shu, "Multilinear principal component analysis network for tensor object classification," *IEEE Access*, vol. 5, pp. 3322–3331, 2017.

[2] H. Lu, K. N. Plataniotis, and A. Venetsanopoulos, *Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data*. Chapman and Hall/CRC, 2013.

[3] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, 2009.

[4] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: a simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[5] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Master's thesis, Department of Computer Science, University of Toronto, 2009.

[6] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[7] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," in *European Conference on Computer Vision*. Springer, 2004, pp. 253–266.

[8] M. Crosier and L. D. Griffin, "Using basic image features for texture classification," *International Journal of Computer Vision*, vol. 88, no. 3, pp. 447–460, 2010.

[9] R. E. Broadhurst, "Statistical estimation of histogram variation for texture classification," in *Proc. Intl. Workshop on Texture Analysis and Synthesis*, 2005, pp. 25–30.

[10] S. Verma, W. Liu, C. Wang, and L. Zhu, "Hybrid networks: Improving deep learning networks via integrating two views of images," in *Neural Information Processing - 25th International Conference, ICONIP , Proceedings, Part I*, 2018, pp. 46–58.

[11] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *European conference on computer vision*. Springer, 2016, pp. 646–661.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.