MATH6380P Project-2:

Experiments on DCFNet

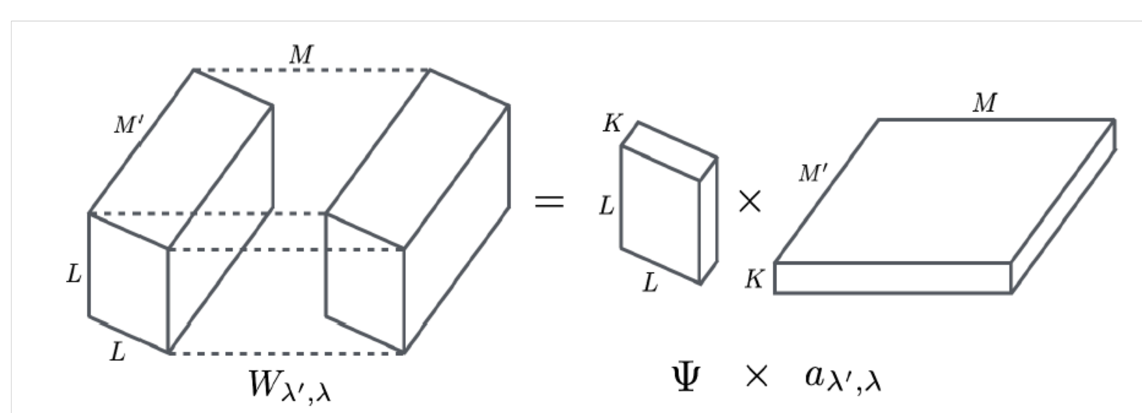
LIANG Zhicong, HUANG Zhichao, WEN Ruixue
Department of Mathematics, Hong Kong University of Science and Technology

1. Introduction

We conduct experiments on CIFAR10 with neural networks combining decomposed convolutional filters (DCF)[1] and successful network structures, such as VGG and ResNet, and evaluate the capacity of these two newworks according to their classification accuracy. As for the decomposition of the filters, three kinds of bases are considered, including Fourier Bessel bases (FB), random bases and PCA bases. We found that in both networks, even with a significant drop of the number of parameters, the accuracy is still comparable or even better than the ones reported in the paper.

2. DCFNet

The key idea of DCFNet is to decompose each convolutional filter into a truncated expansion with pre-fixed bases in the spatial domain, where the expansion coefficients remain data dependent. In this case, a convolutional layer with weight size $L * L * M' * M$ is transformed into a layer with weight size $1 * 1 * M'K * M$, so the number of parameters of this layer is a factor K/L^2 smaller. This model compression method can reduce the redundancy of parameters of each layer without loss of too much accuracy or even with better performance.



$$W_{\lambda',\lambda}(u) = \sum_{k=1}^{K} (a_{\lambda',\lambda})_k \psi_k(u),$$

Illustration of a DCFNet layer, K denotes the number of selected bases

7. References

[1]*Qiang Qiu, Xiuyuan Cheng, Robert Calderbank, Guillermo Sapiro. DCFNet: Deep Neural Network with Decomposed Convolutional Filters. arXiv: 1802.04145, 2018.*

3. Implementation Stetting

	epoch	batch size	weight decay	Optimizer	learning rate
VGG	300	256	5e-5	Adam	0.001 (halved every 30 epochs)
ResNet	300	64	5e-4	SGD	0.1 (halved every 30 epochs)

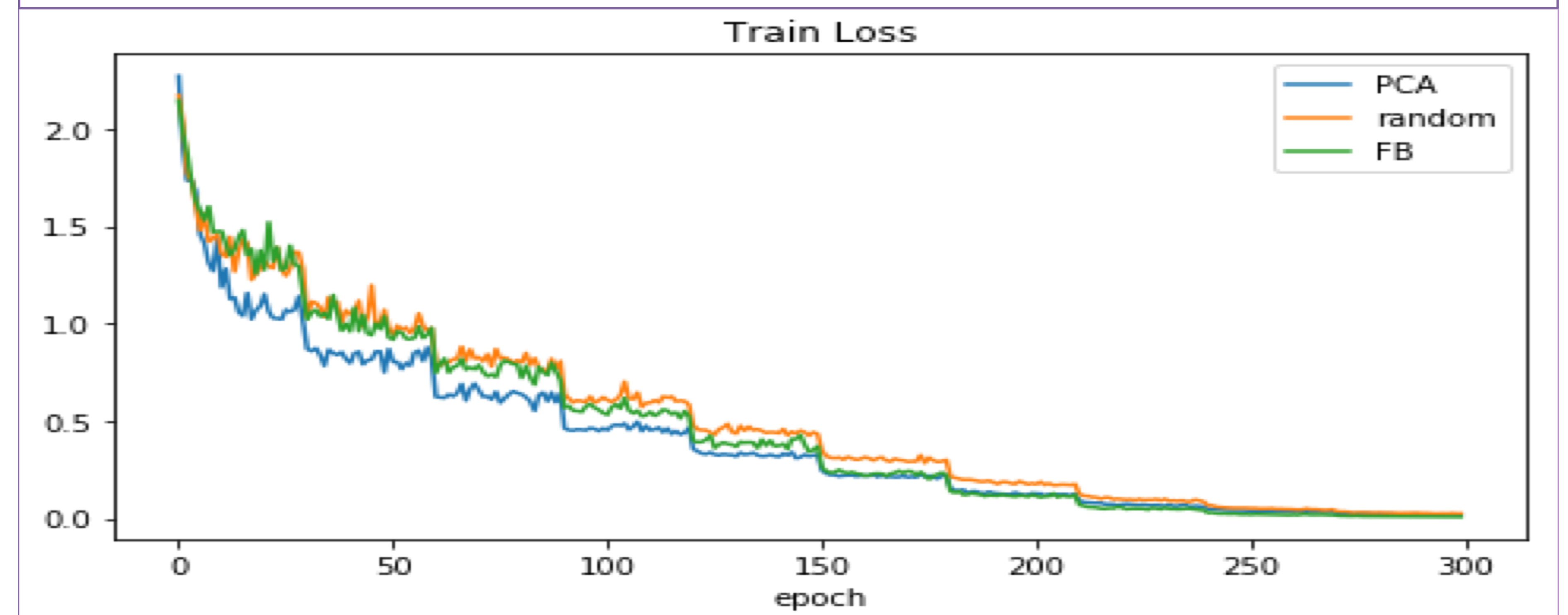
We implement the experiments with K=3,5,9. As for DCF-PCA, bases are constructed layer by layer, by choosing the principle components with the biggest K eigenvalues, from the same layer of corresponding pre-trained CNN model.

4. Results

		K=3	K=5	K=9
VGG16-CNN acc.		0.8702		
#param.		1.47e+7		
VGG16-DCF (results in paper) acc.	FB	0.8821	0.8779	/
	RB	0.7846	0.8416	/
	PCA	0.8754	0.8760	/
VGG16-DCF acc.	FB	0.9013	0.9127	/
	RB	0.8911	0.9048	0.9159
	PCA	0.8677	0.8865	0.8912
#param.		4.91e+6	8.18e+6	1.47e+7
ResNet50-CNN acc.		0.9362		
#param.		2.46e+7		
ResNet50-DCF acc.	FB	0.8341	0.8408	/
	RB	0.8029	0.8315	0.8468
	PCA	0.8602	0.8802	0.8956
#param.		8.20e+6	1.37e+7	2.46e+7

By fine-tuning the hyper parameters (i.e. batch size, weight decay, learning rate, etc.), we achieve better results than the ones shown in the paper. In each bases case, the classification accuracy of DCFNet increases as K becomes larger. One noteworthy, generally, results of ResNet50-DCF are not as goophenomenon is thatd as that of VGG16-DCF, we leave the concrete analysis in discussion part.

In ResNet, as shown below(k=5), network with PCA bases outperforms the other two. However, in VGG case, we found that network with FB base converges fastest. One thing common about two networks is that, DCFNet with random bases converges slowest, which shows that structured bases advanced the random ones in training.



5. Discussion

As mentioned before, compared with VGG16-DCF, ResNet50-DCF performs worse. One hypothesis is that in the first decomposed convolution layer of ResNet50, kernel size is 7x7 and K=3, 5, which means that the number of parameters is a factor of $\frac{K}{L^2} \approx 0.06, 0.10$ smaller, losing too much information from original input.

As for VGG, DCFNet shows a comparable accuracy with appropriate parameters' reduction order as K/L^2. When tuning hyper parameters in this part, loss function decays slow as learning rate is set larger than 0.01.

When arguing the benefit of DCFNet, the paper mentions that high-frequency nuance details are often irrelevant to classification tasks. However, this may not be true with respect to deeper feature maps in the middle or near the end of a neural network, since their distribution is totally different from the input image.

This also provides a reasonable explanation for our experiment outcome that, in ResNet50, FB bases seems inferior to PCA bases. Since the residual parts of ResNet act like perturbation sometimess, it may be mis-ingored by the low-frequency capturing DCFNet filters.

6. Conclusion

We implement the DCFNet, who offers a new method to constructed convolutional filters in neural networks, maintaining capacity while reducing the number of parameter. Our result outperform the ones in the original paper. And we analyze roughly the dynamic behind different combination of network and DCFNet bases.