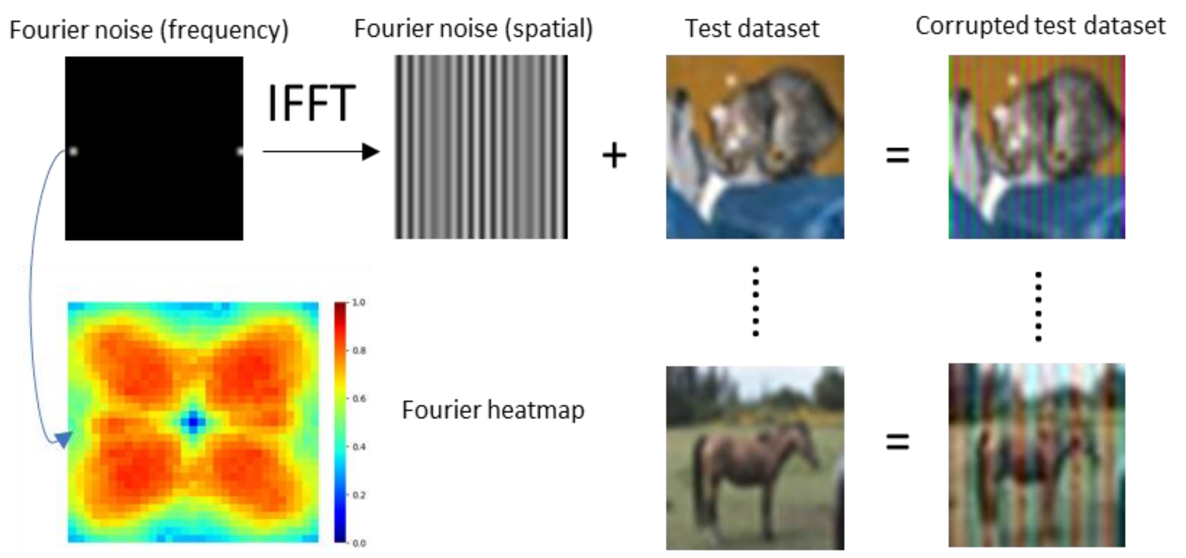
Frequency Analysis of Neural Network

1. Introduction

Though neural networks (NN) have been widely used for a decade, its intrinsic mechanism remains a black-box and is still a popular topic to explore. Unlike mathematical models which are interpretable, the mechanism of neural networks are hard to understand. Namely, the way that neural network uses to extract information from data remains unknown and how the network affected by corruptions is unknown as well. Many researchers have worked on analysing neural networks from frequency perspective, such as Fourier heatmap [1] – a tool to measure how different frequency noises influence a classifier. Differently, Rahaman et al. [2] investigate what frequency component is learned first, and claim that there is a spectral bias during learning (low-frequency is learned first and high-frequency is learned later). However, it is unknown whether such a claim holds for image classification. Besides, the usage of Fourier heatmap in intermediate layers of neural networks (Fig. 2) does not provide good enough information to interpret how neural network reacts to different frequencies.

Figure 1. Process of generating Fourier heatmap (Pixel values of Fourier heatmap indicates either classification error or the changes of representations (measured by cosine similarity or other distance measurements)



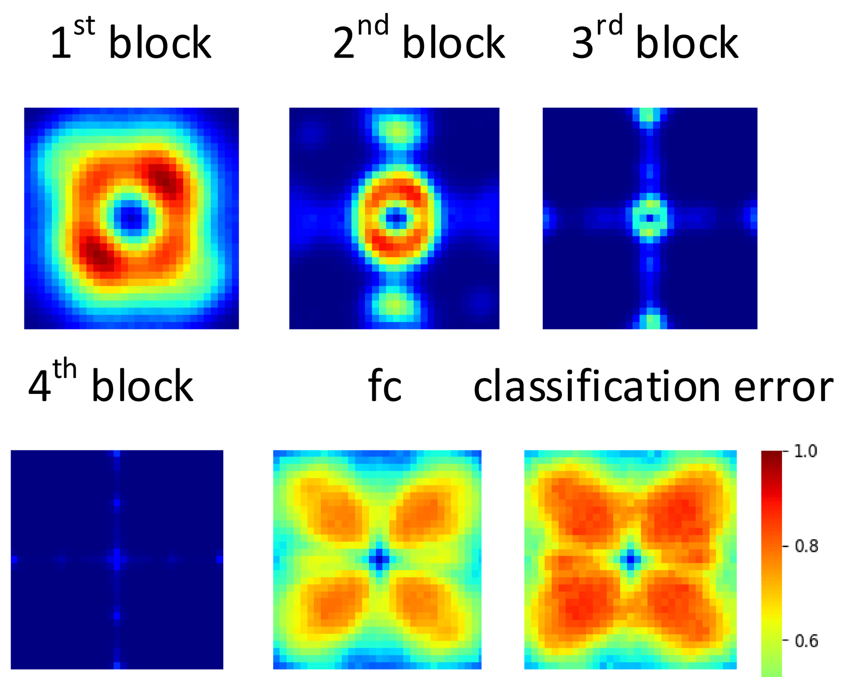


Figure 2. Fourier heatmaps of ResNet18 trained on CIFAR10: Maps indexed by '1st block' to '4th block' indicate the heatmap of the last convolutional layer of each block, which measure the changes of response maps of intermediate layers. Index 'fc' refer to the classifier which is actually a fully connected layer

In this work, we analyse neural networks from frequency perspective and tend to relate such analysis to robustness of neural networks towards image corruption. We investigate the learning behaviour and frequency preference of neural network (specific to image classification) towards different frequency components.

References

[1] Yin, D., Lopes, R. G., Shlens, J., Cubuk, E. D., & Gilmer, J. (2019). A Fourier Perspective on Model Robustness in Computer Vision. doi:10.48550/ARXIV.1906.08988

[2] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., & Courville, A. (2019). On the Spectral Bias of Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (pp. 5301–5310). PMLR.

Shunxin Wang, s.wang-2@utwente.nl

Supervisors: Raymond Veldhuis (Promoter),
Nicola Strisciuglio,
Christoph Brune (Co-supervisor)

UNIVERSITY OF TWENTE.

2. Methods – Frequency Controlled Dataset

The details of each class in the datasets (synthetic images with size 32x32) are as follows:

1) Class 0: a,b,c + special pattern
2) Class 1: a,b,c - special pattern
3) Class 2: a,b,c,d - special pattern
4) Class 3: d - special pattern

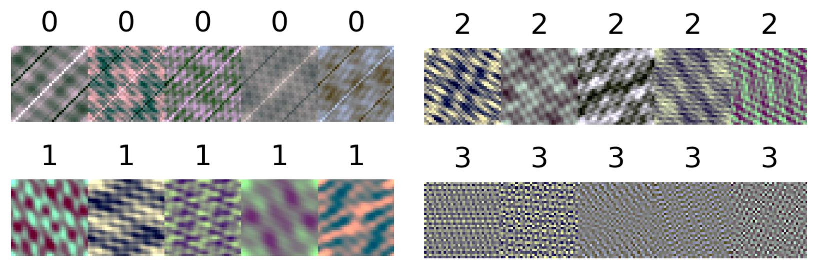


Figure 3. Samples of dataset *Syn_4*

where a, b, c, d are frequency bands. The special pattern contains frequencies $[u, v]$ where $u = v \in \{1,3,5,7,9,11,13,15\}$. Datasets are noted as *Syn_d*, where *d* is the specific frequency band in class 3.

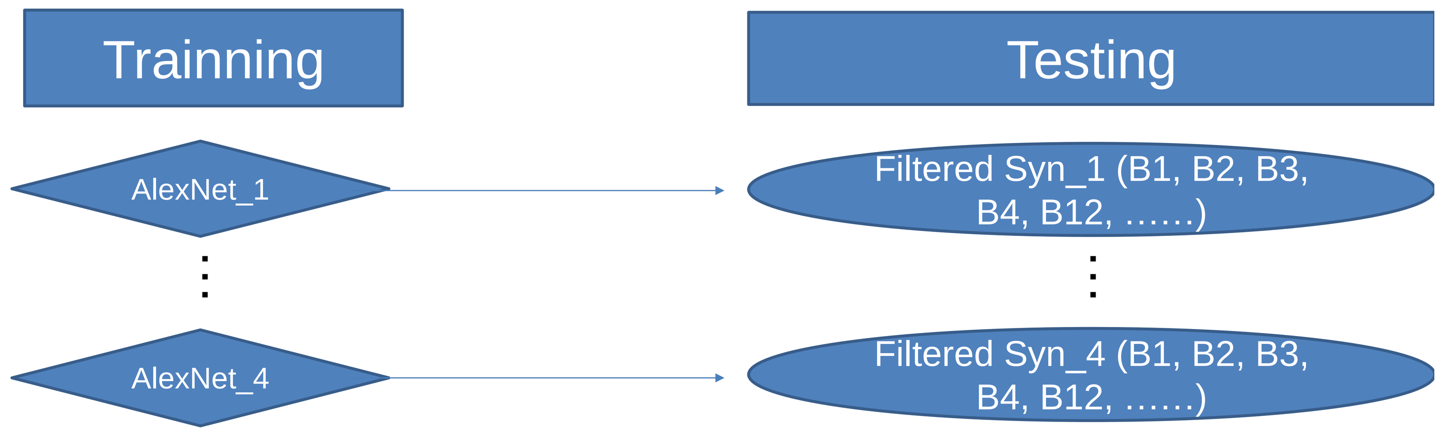


Figure 4. Training and testing for models on the 4 synthetic datasets (trained for 50 epochs with the same learning rate and batch size); during testing, the synthetic datasets are filtered to have different frequency components.

3. Results

- For classification, either low-frequency or high-frequency can be learned first, as long as they provide discriminative information.
  - In Fig. 5, class 3 is predicted correctly first, while it can only contains frequencies in the lowest frequency band or the highest frequency band.
  - Same phenomenon holds for other models (which are trained on *Syn_2/3*)

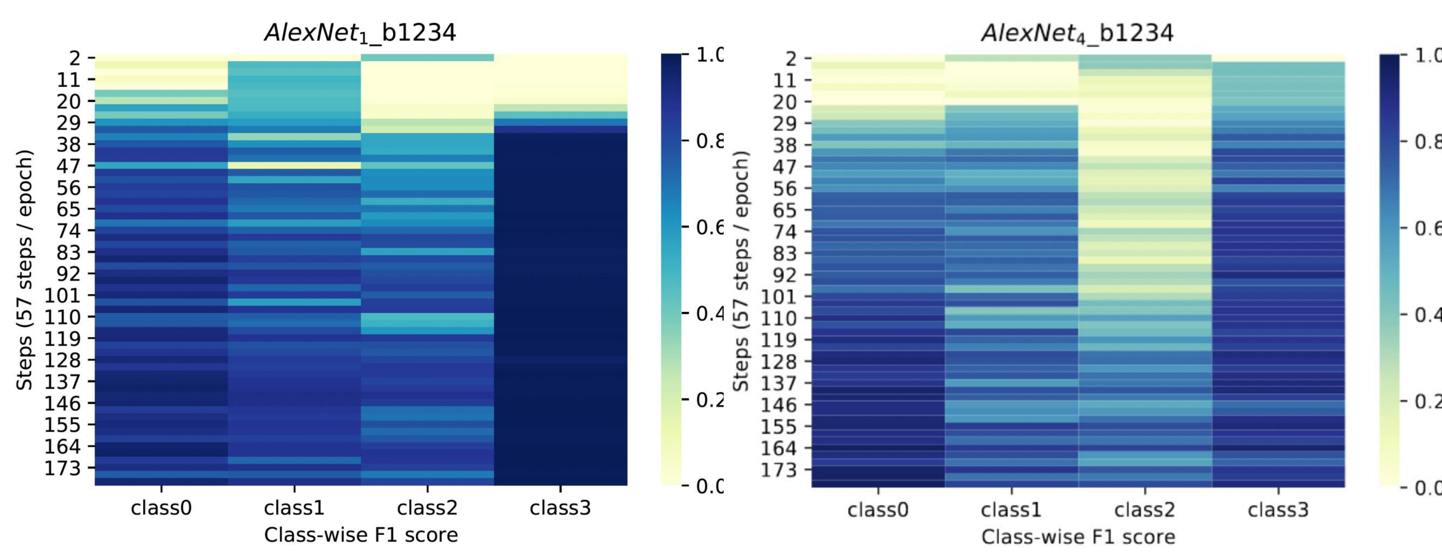


Figure 5. *AlexNet_1* and *AlexNet_4* tested on B1234 of *Syn_1* and *Syn_4 respectively* (the test performance of each class in the first 3 training epochs is measured by F1-score); Class 3 is the easiest to learn since images of it are generated from frequencies in one band only.

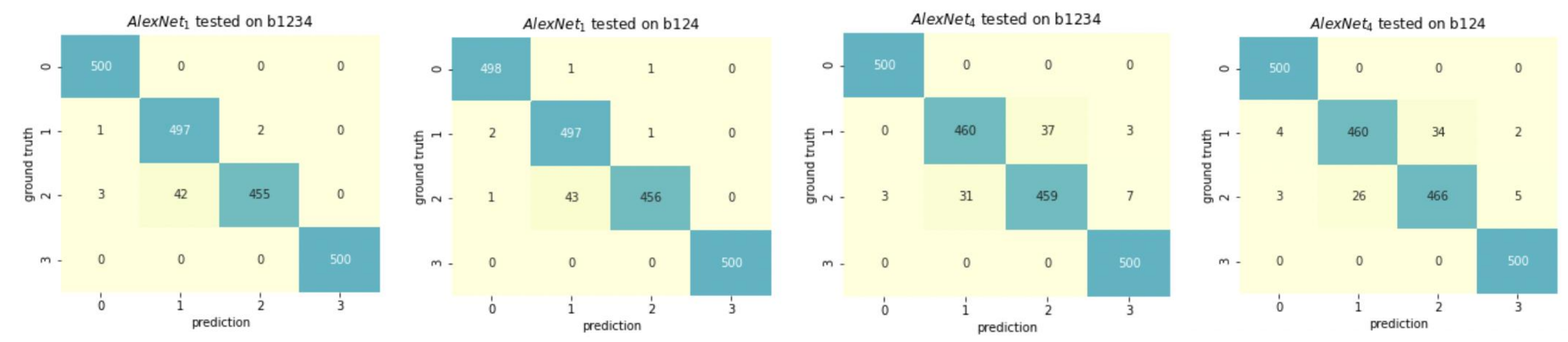


Figure 6. Confusion Matrices of *AlexNet_1* and *AlexNet_4* tested on B1234 and B124 of *Syn_1 and Syn_4 respectively* (removing one frequency band does not influence much on the performance)

- For each class, there is a shortcut to achieve classification
  - Class 0 & 2: No need to search for frequencies in all bands, but 2 or 3 bands are enough.
  - Class 1: The classification is not only based on the existence of the three bands, but when there is no special characteristics appears, models give predictions as class 1.

4. Future work

- Design an approach to find such shortcuts, which can be used either in data pre-processing or during training
- Investigate the relationship between complexity of shortcuts and model robustness