On the Spectral Bias of Neural Networks

Nasim Rahaman* Aristide Baratin* Devansh Arpit

Felix Draxler Min Lin

Fred A. Hamprecht Yoshua Bengio Aaron Courville

The good old question:

Why do massive neural networks generalize when they can learn random labels?

UNDERSTANDING DEEP LEARNING REQUIRES RE-THINKING GENERALIZATION

Chiyuan Zhang*
Massachusetts Institute of Technology
chiyuan@mit.edu

Samy Bengio
Google Brain
bengio@google.com

Moritz Hardt
Google Brain
mrtz@google.com

Benjamin Recht†
University of California, Berkeley
brecht@berkeley.edu

Oriol Vinyals
Google DeepMind
vinyals@google.com

A Closer Look at Memorization in Deep Networks

Devansh Arpit [*1 2] **Stanisław Jastrzębski** [*3] **Nicolas Ballas** [*1 2] **David Krueger** [*1 2] **Emmanuel Bengio** [4] **Maxinder S. Kanwal** [5] **Tegan Maharaj** [1 6] **Asja Fischer** [7] **Aaron Courville** [1 2 8] **Yoshua Bengio** [1 2 9] **Simon Lacoste-Julien** [1 2]

The good old question:

Why do massive neural networks generalize when they can learn random labels?

Implicit Regularization in Deep Learning

by

Behnam Neyshabur

Theory of Deep Learning III: explaining the non-overfitting puzzle

by

T. Poggio[†], K. Kawaguchi [†⊤], Q. Liao[†], B. Miranda[†], L. Rosasco[†]

with

X. Boix[†], J. Hidary[††], H. Mhaskar[◇],

[†]Center for Brains, Minds and Machines, MIT

[⊤]CSAIL, MIT

[††]Alphabet (Google) X

[◇]Claremont Graduate University

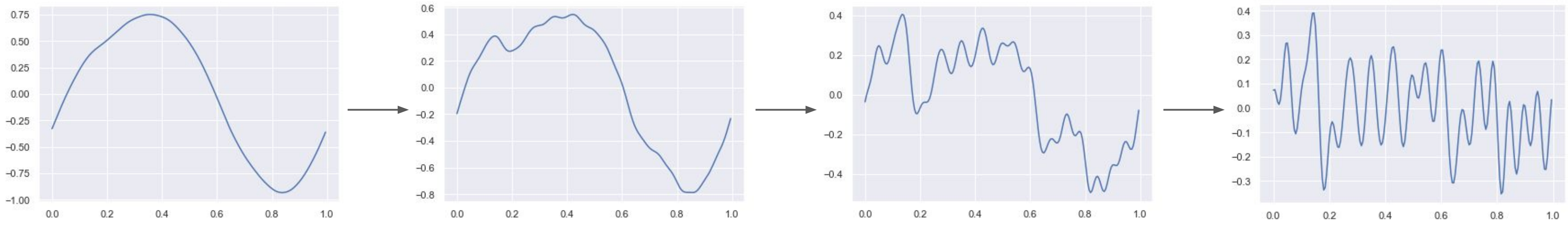
Our proposal:

Neural networks learn simpler functions first.

But how do we quantify *simplicity*?

Our approach:

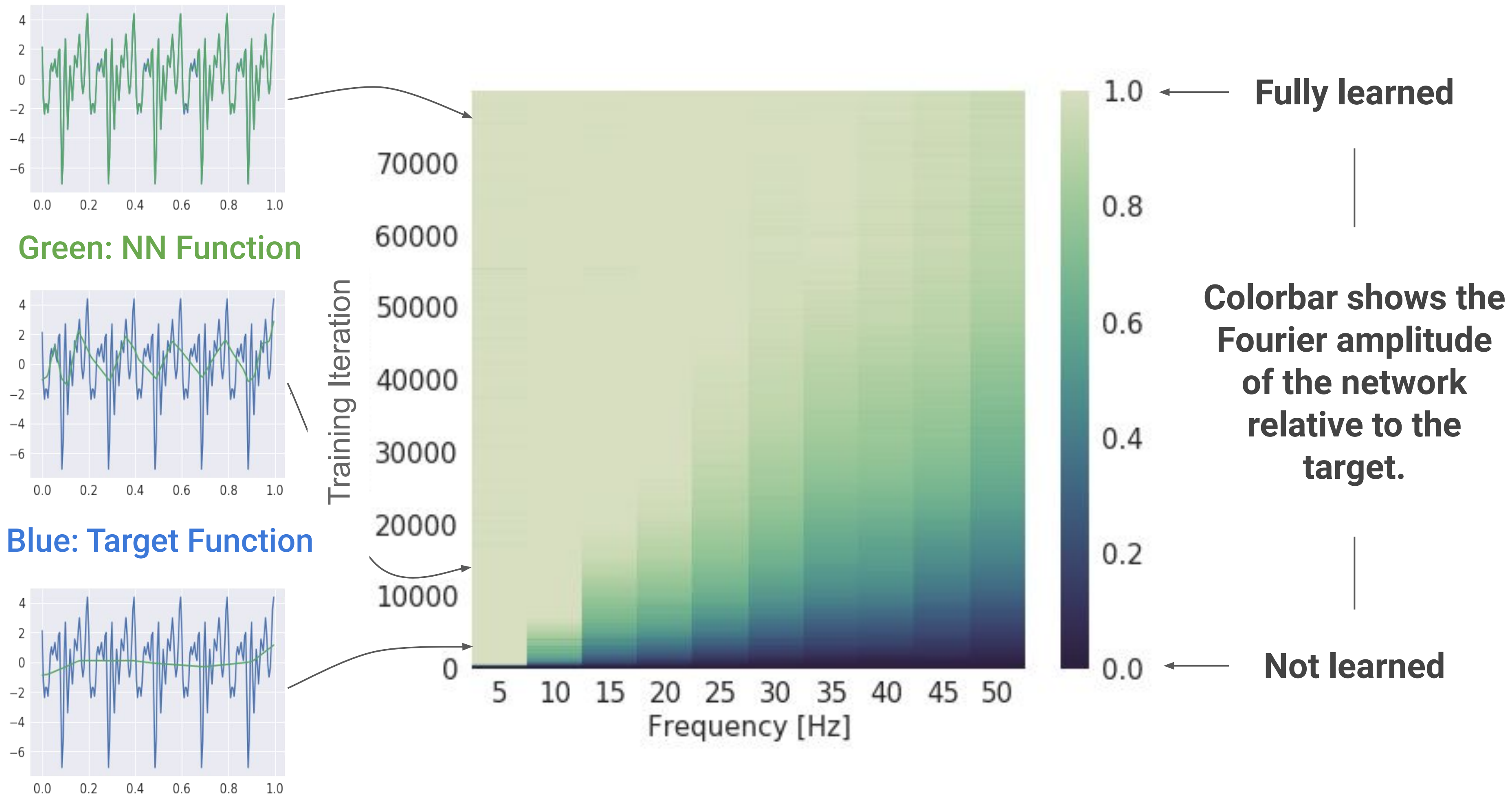
We use the (Fourier) Spectrum.



Lower Frequency Functions **Higher Frequency Functions**

Our proposal becomes:

Neural networks learn lower frequencies first.


Green: NN Function
Blue: Target Function
Training Iteration
70000
60000
50000
40000
30000
20000
10000
0
5 10 15 20 25 30 35 40 45 50
Frequency [Hz]
1.0
0.8
0.6
0.4
0.2
0.0
Fully learned
Colorbar shows the Fourier amplitude of the network relative to the target.
Not learned

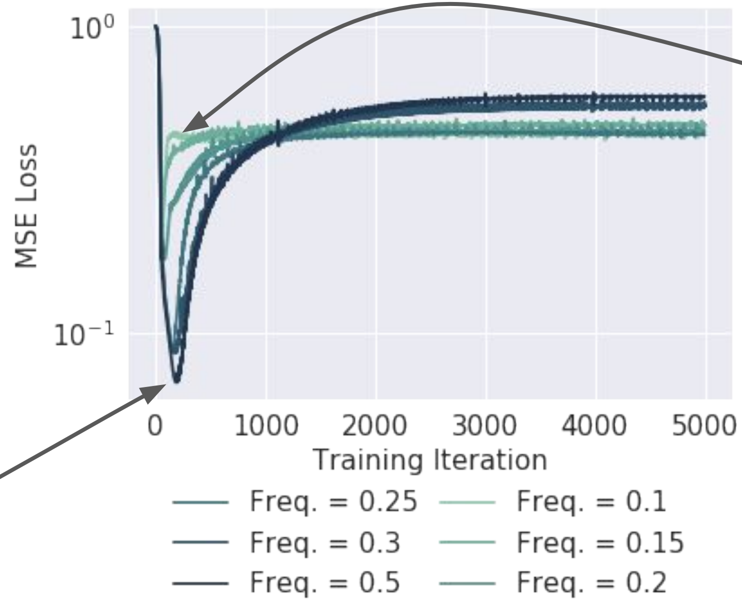
Why should I care?

One of the many reasons:

NN training is vulnerable against low frequency label noise.

Training with label noise

High frequency label noise leads to a *dip* in the validation loss.



Validation loss on MNIST (w.r.t pure targets)

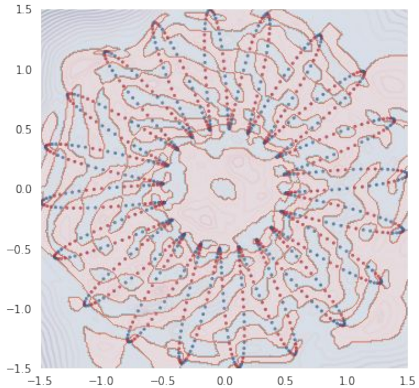
Low frequency label noise does not… :(

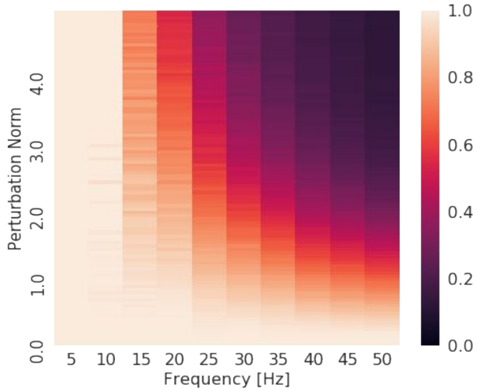
To learn how the manifold complexity attenuates the spectral bias,

drop by at our poster!

#72

Spoilers





Learning gets ***easier*** with ***increasing*** manifold complexity.

To express **complex functions**, the parameters must "***work together in harmony***".

Thank you for your attention!