Mason Experimental Geometry Lab

Topological Analysis of Neural Networks

Finn Brennan, Diane Hamilton, Joseph Jung, Charles Alexander Lee,
Shrunal Pothagoni, Benjamin Schweinhart

GEORGE MASON UNIVERSITY

Introduction

Although the method of training a neural network (NN) is well understood, how the data is transformed within the NN is not. We create a new metric that measures the changing topological complexity (TC) of the data as it passes through the layers of a NN.

The Neural Network

A neural network is a composition of functions which take vectors as input data.

$$g \circ f_k \circ f_{k-1} \circ \cdots \circ f_2 \circ f_1(x)$$

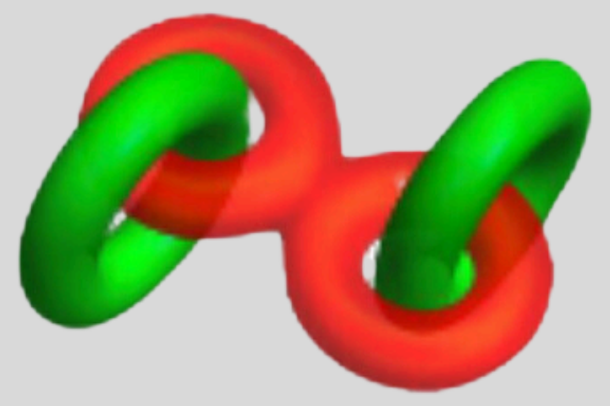
Each function looks like,

$$f_i = \sigma(W_i x + b_i)$$

where W_i and b_i are the adjustable weights and biases and σ is a non-linear activation function. The weights and biases are updated via back propagation to improve model performance.

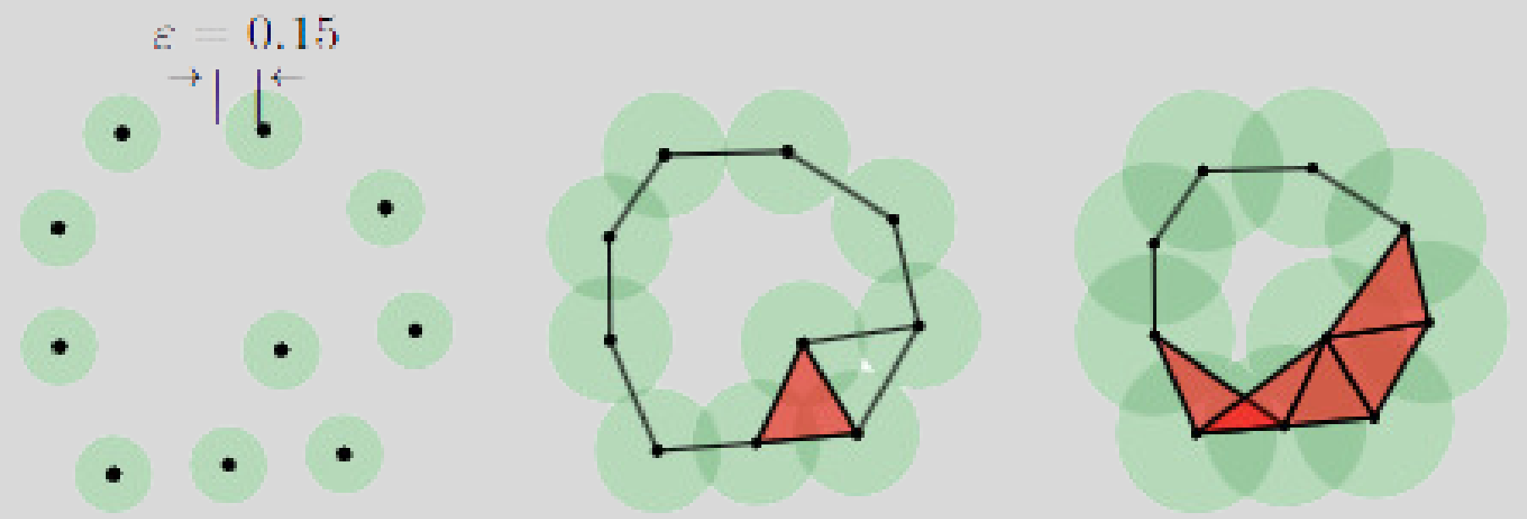
Data Has Shape

Data has shape and that shape has meaning **[1]**. We assume that our data lives on a manifold. With this we can measure the topological complexity of our data at each layer of the neural network by forming simplicial complexes.



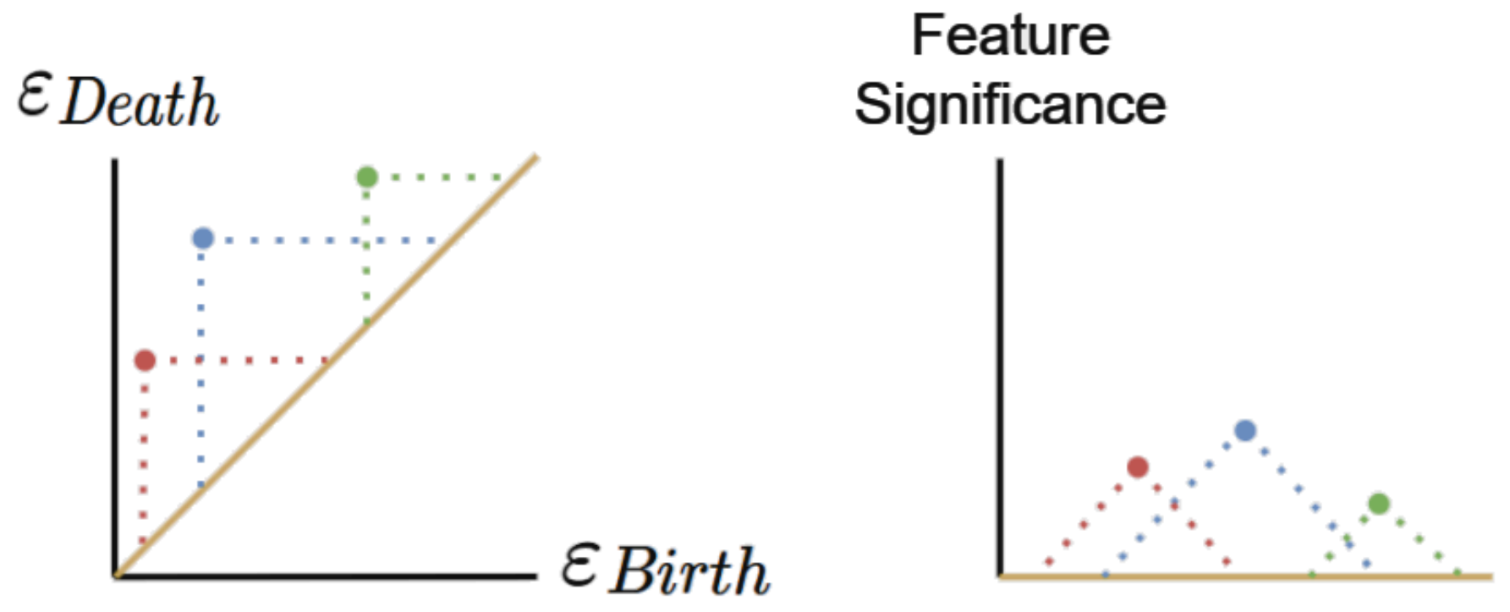
Simplicial Complex

To analyze the shape of our data, we construct a simplicial complex. We can now view our data set as points existing on a manifold.



Persistence Diagram

A persistence diagram captures which topological features persist the longest as we create the simplicial complex.



Activation Landscapes

A persistence diagram can be turned into an activation landscape curve, which we denote as $\lambda(t)$ **[2]**

The L^2 norm of this curve gives the topological complexity of our data.

$$||\lambda(t)||^2_{L^2} = \int_0^\infty (\lambda(t))^2 dt$$

Our New Metric

Our new metric measures how the topological complexity of a dataset changes as it is passed through a NN. By building off the norm of the activation landscape as shown in [2] we can measure the TC of the dataset at the layer. i^{th}

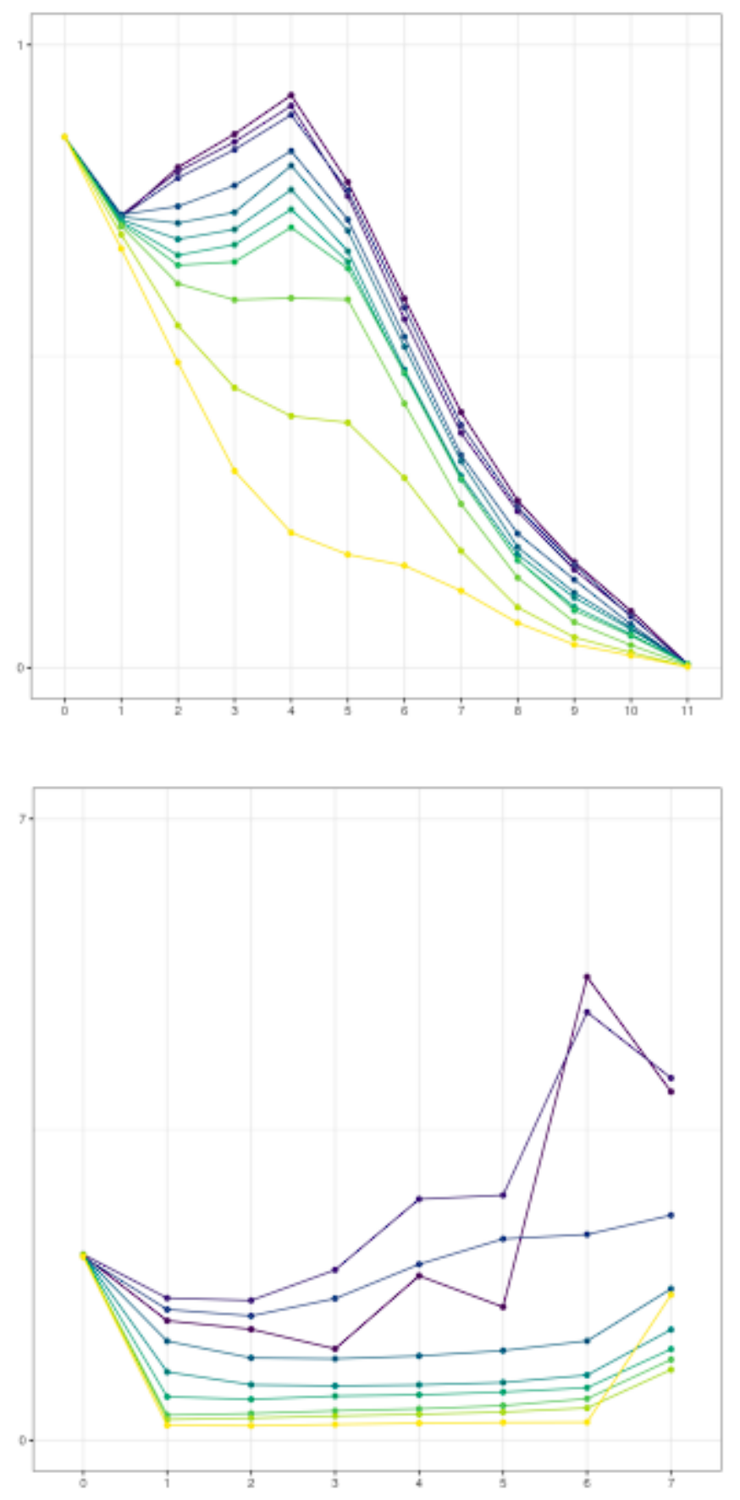
$$TC_i := ||\lambda_i(t)||_{L^2}$$

We then make a new function which plots the TC and connects them by lines. Our final metric is the norm of this new function:

$$||\hat{T}(t)||_{L^2} = \sqrt{\int_0^\infty (\hat{T}(t))^2 dt}$$

This measures how the topological complexity of our data changes as it passes through the layers of our NN.

Topological Complexity



The topological complexity of a neural network can be found with the activation landscape at each layer of the NN. Doing this can give us valuable insight on the spatial characteristics of the data

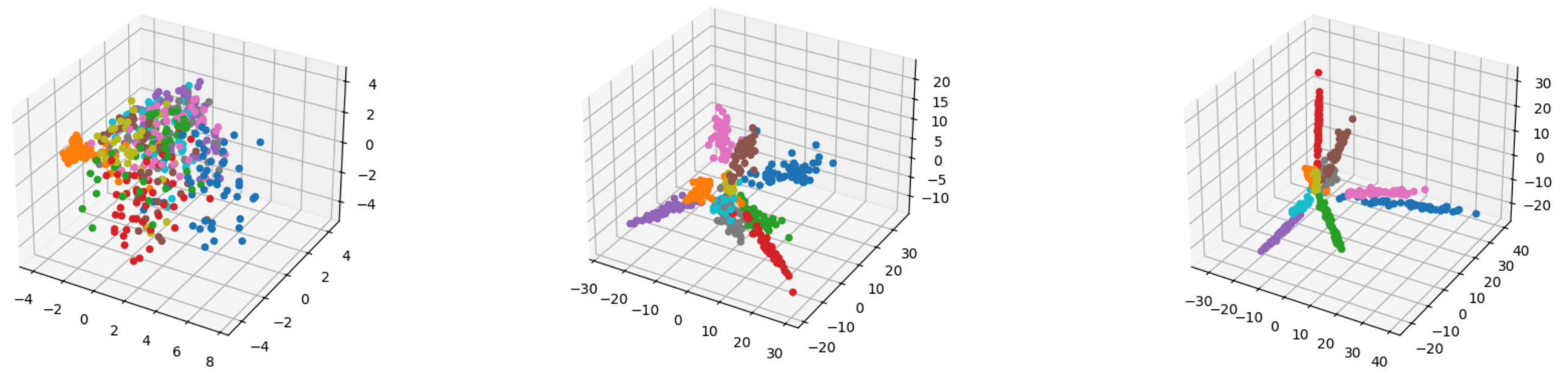
Average norms of activation landscapes using a larger number of training accuracy thresholds with consecutive layers [2]

Citations

[1] Gregory Naitzat, Andrey Zhitnikov, and Lek-Heng Lim. Topology of deep neural networks. The Journal of Machine Learning Research, 21(1):7503–7542, 2020

[2] Matthew Wheeler, Jose Bouza, and Peter Bubenik. Activation landscapes as a topological summary of neural network performance. In 2021 IEEE International Conference on Big Data (Big Data), pages 3865–3870. IEEE, 2021

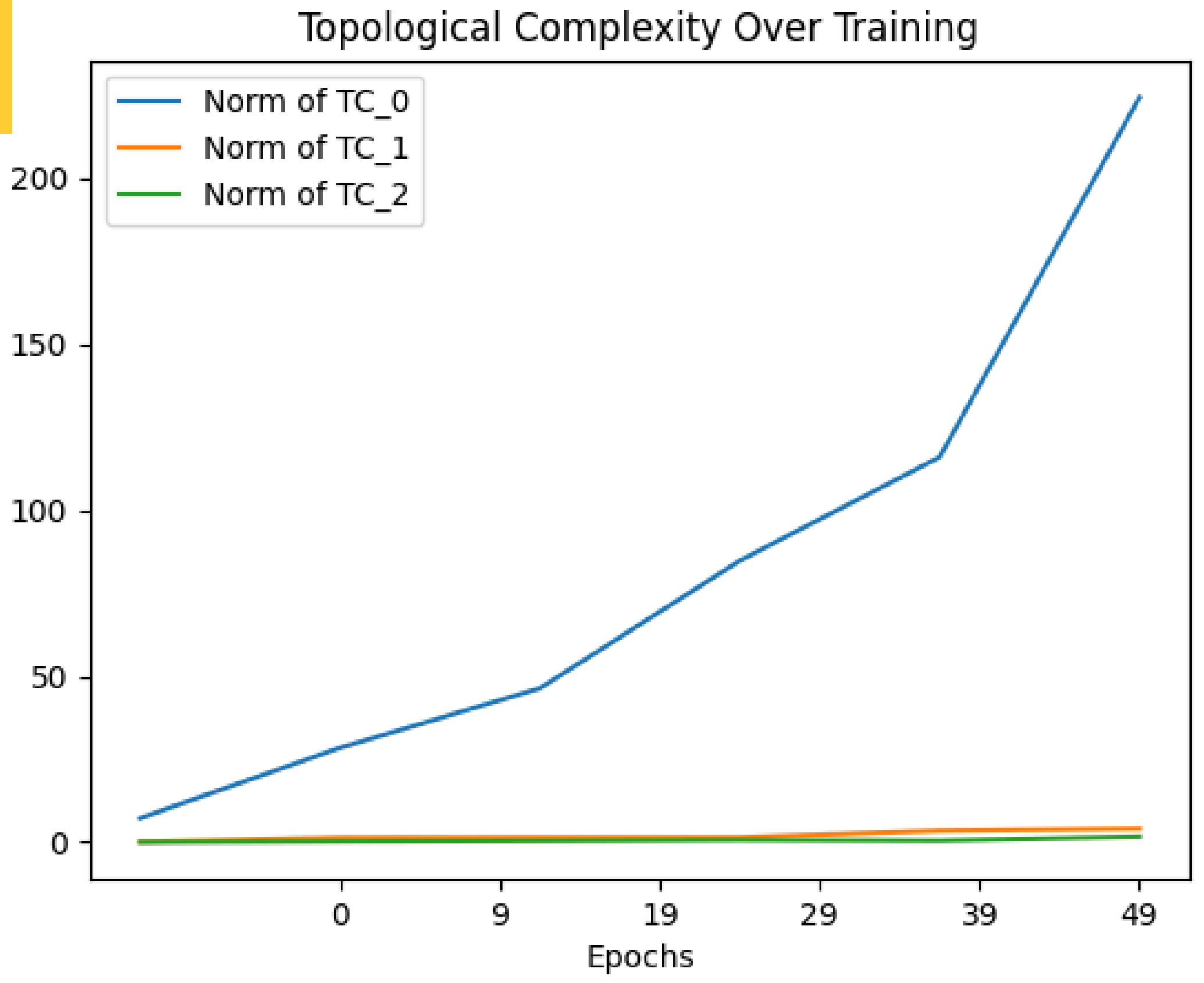
“Seeing” into a NN



Above is a visualization of the MNIST dataset (images of hand written digits) as they pass through the NN at several epochs of training. The image data begins with high topological complexity and as the model is trained the complexity of the data reduces, distinguishing the classes.

Current Experiment

We hypothesize that during training, the overall topological complexity of a neural network—averaged across all layers—temporarily increases at intermediate epochs before decreasing towards final epochs. Much like solving a puzzle where pieces are first spread out and then reassembled, a well-trained neural network operates similarly.



Let $\mathcal{D} = \{v_i\}_{i=1}^n$ with $v_i \in \mathbb{R}^d$, and let $N(x) = g \circ f_N \circ \cdots \circ f_1(x)$ represent an N-layer fully trained neural network.

Notation: The i-th activation layer of the network is denoted as $N^{(i)}(x) = f_i \circ \cdots \circ f_1(x)$ for $i < N$. Starting from an initialized model $N_0(x)$ at $t = 0$, we say the model is trained when $N_t(x) \to N(x)$ as $t \to T$, where T is the total number of training epochs.

Conjecture: Let $\mathcal{D}_t^{(k)}$ denote the transformed data up to the k-th layer after t training epochs, and let $\mathrm{TC}_t^{(k,p)}$ represent the p-th dimensional topological complexity of $\mathcal{D}_t^{(k)}$. Then for some $K < N$, if $N_t(x) \to N(x)$, there exists a training epoch t_i such that

$$\mathrm{TC}_{t_i}^{(k,0)} > \mathrm{TC}_{t_i}^{(k,p)} \quad \text{for all } p > 0$$

and for some $K \leq k \leq N$.