Analyzing deep neural networks with persistent homology

Thomas Gebhart

University of Minnesota

UNIVERSITY OF MINNESOTA

Neural Networks

Large number of parameters

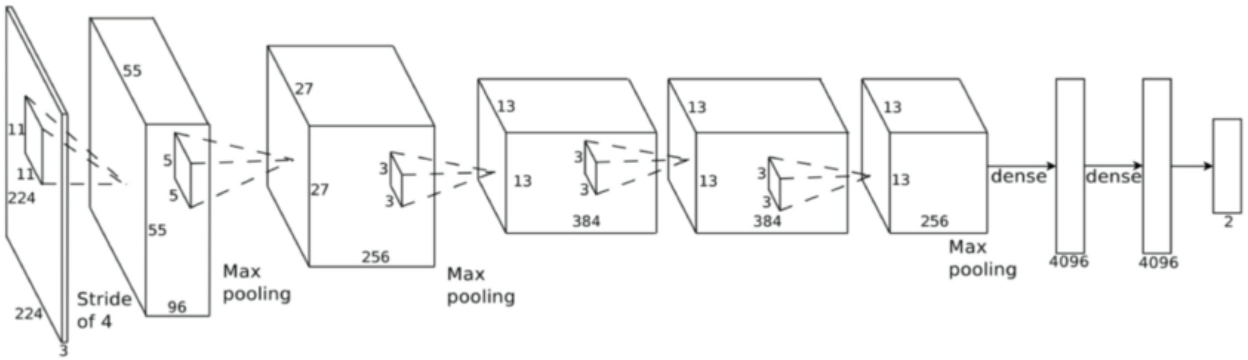
+

Nonlinearities via activation functions

+

Layer-wise functional components

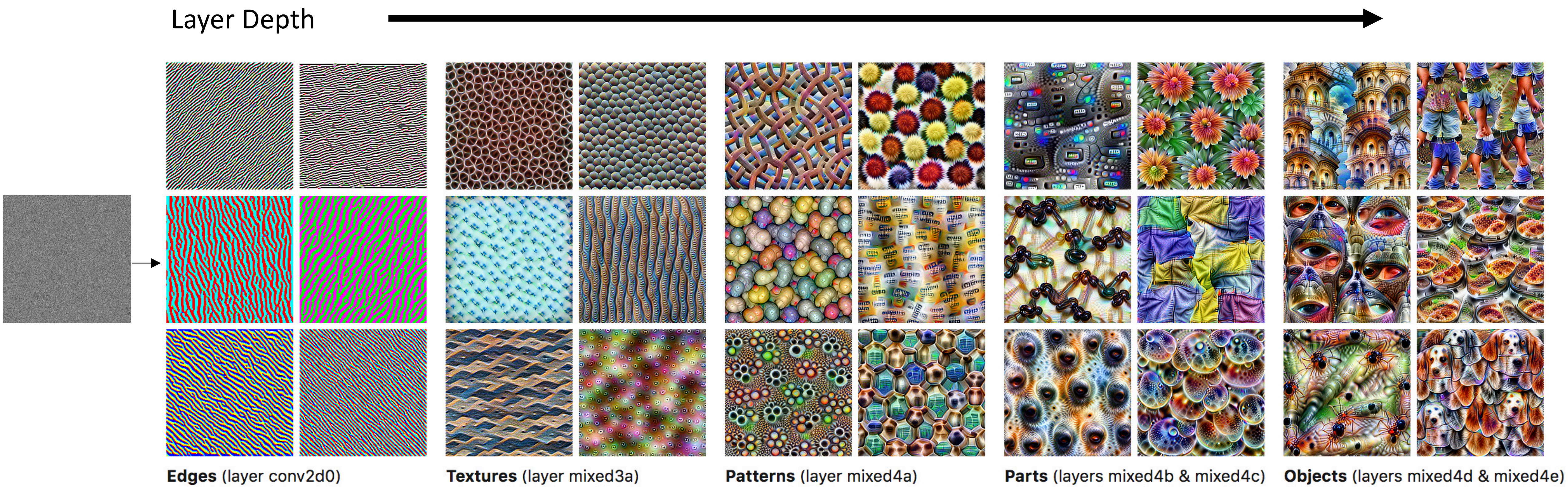
Difficult to interpret



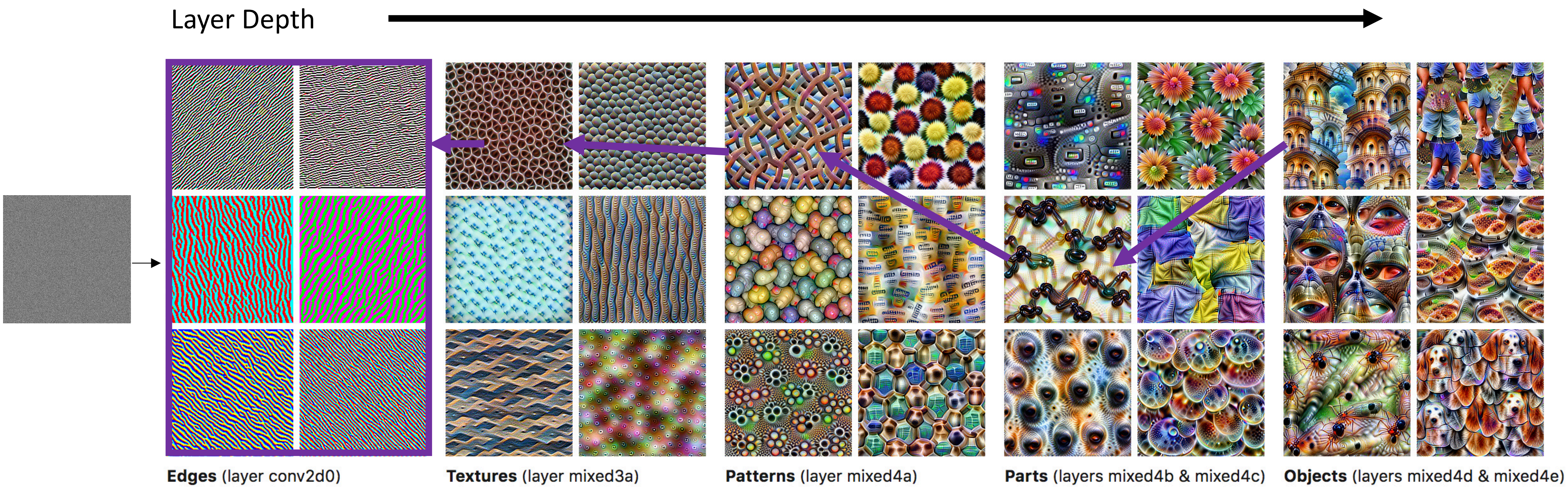
Layer Name	Tensor Size	Weights	Biases	Parameters
Input Image	227x227x3	0	0	0
Conv-1	55x55x96	34,848	96	34,944
MaxPool-1	27x27x96	0	0	0
Conv-2	27x27x256	614,400	256	614,656
MaxPool-2	13x13x256	0	0	0
Conv-3	13x13x384	884,736	384	885,120
Conv-4	13x13x384	1,327,104	384	1,327,488
Conv-5	13x13x256	884,736	256	884,992
MaxPool-3	6x6x256	0	0	0
FC-1	4096×1	37,748,736	4,096	37,752,832
FC-2	4096×1	16,777,216	4,096	16,781,312
FC-3	1000×1	4,096,000	1,000	4,097,000
Output	1000×1	0	0	0
Total				**62,378,344**

https://www.learnopencv.com/number-of-parameters-and-tensor-sizes-in-convolutional-neural-network/

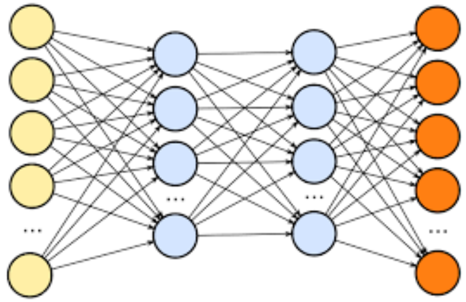
Non-local Representations



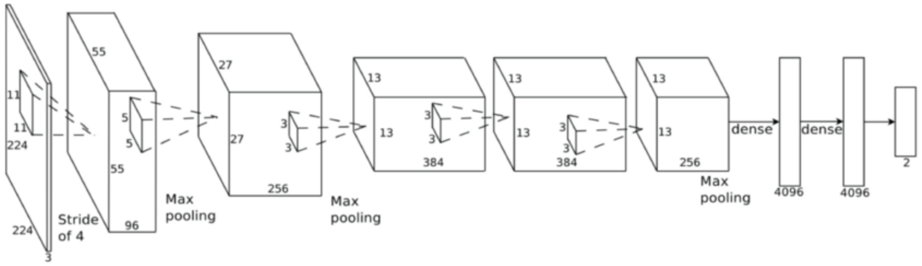
Non-local Representations



Representations Distributed Layer-wise



Network Structure

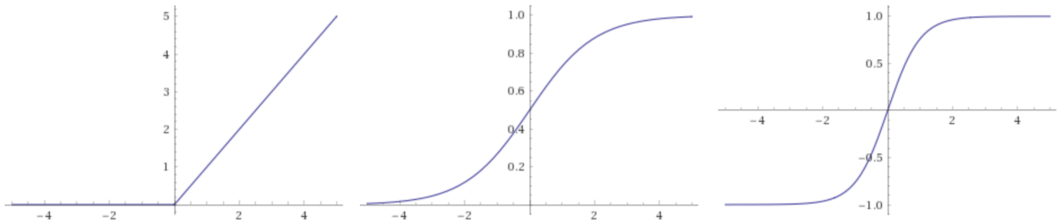


High dimensionality and multiple scales

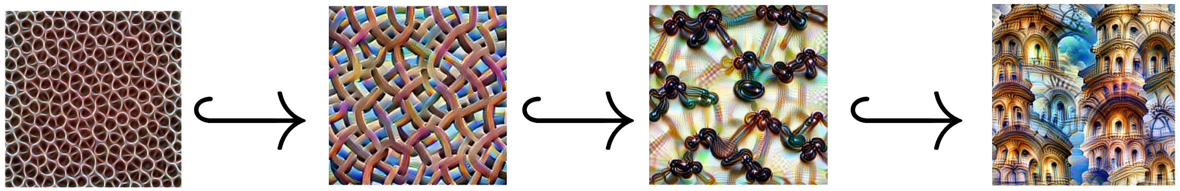
Persistent Homology

Nonlinearities

Global representations with inclusion



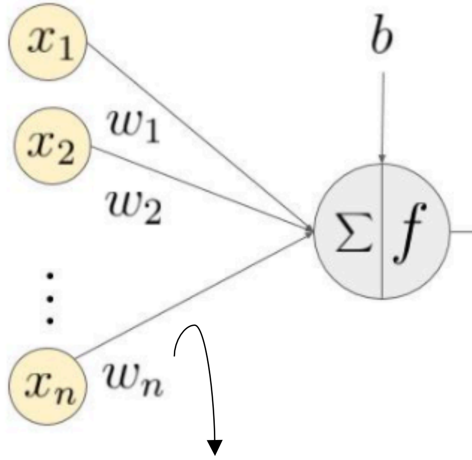
ReLU Sigmoid Tanh



https://towardsdatascience.com/breaking-neural-networks-with-adversarial-attacks-f4290a9a45aa

Neural Networks as Graphs

- Two Views:
  - *Static Network*
    - Gabella et al.
    - Rieck et al.
  - *Induced Network*
    - This talk
- Connectivity Types:
  - Fully Connected
  - Convolutional
  - Pooling



$$f\left(b + \sum_{i=1}^{n} x_i w_i\right)$$

$$\phi(u, v) = |w_{u \to v} h_u|$$

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Activations layer l

$$\boldsymbol{h}_l$$

Neuron v_1 ... Neuron v_n

$$\begin{pmatrix} w_{1,1} & \cdots & w_{1,n} \\ w_{2,1} & \ddots & \\ \vdots & & \\ w_{n,1} & \cdots & w_{n,n} \end{pmatrix}$$

Parameters layer $l+1$

$$\boldsymbol{W}_l$$

Big Picture

Let $G^{\mathcal{I}} = (V, E, \phi)$ be a network's graphical representation induced by input $\mathcal{I}$

$V = V_0 \sqcup V_1 \sqcup \cdots \sqcup V_{L-1}$ where $u \in V_k,\, v \in V_l$ and $(u, v) \in E$ only if $k = l - 1$

The edge weighting for edge $(u, v) \in E$ given by $\phi(u, v) = |w_{u \to v} h_u|$ defines the filtration:

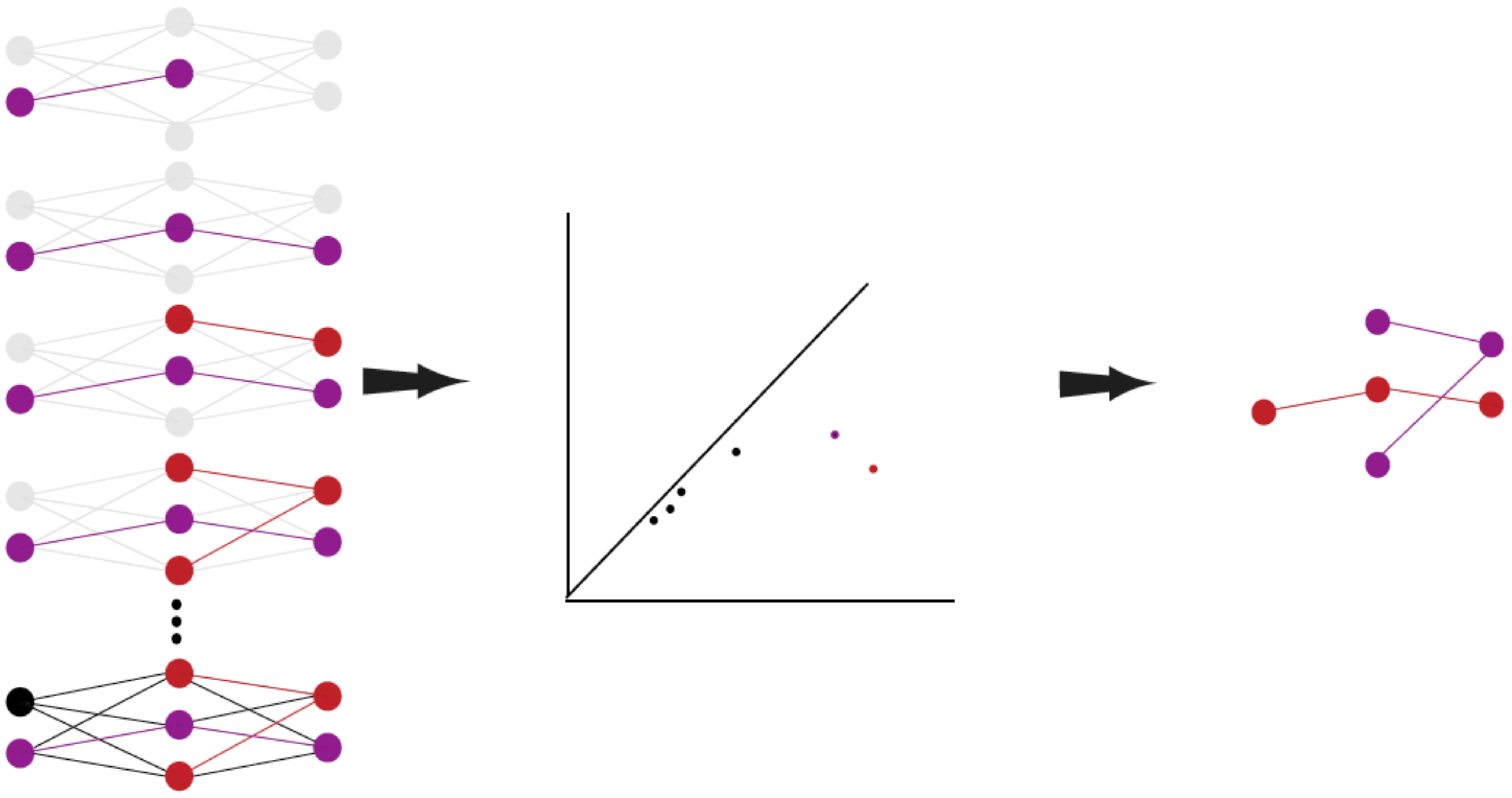
$$\emptyset \subset G_0^{\mathcal{I}} \subset G_1^{\mathcal{I}} \subset \cdots \subset G_N^{\mathcal{I}} = G^{\mathcal{I}}$$

Big Picture

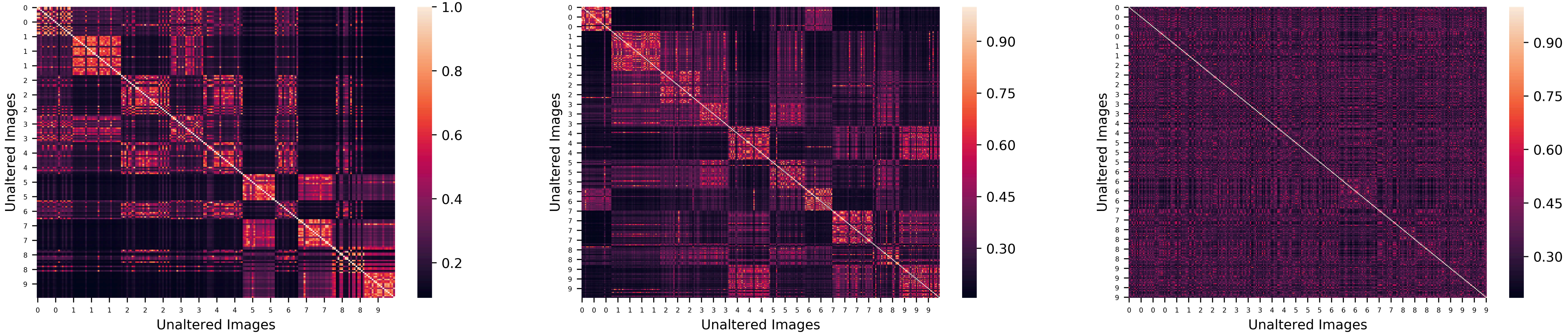
The persistent structure of this network filtration relates to semantic information about the input. We hypothesize:

$$
\begin{array}{ccccccccccc}
0 & \longrightarrow & H_0(G_0^{\mathcal{I}}) & \longrightarrow & H_0(G_1^{\mathcal{I}}) & \longrightarrow & \cdots & \longrightarrow & H_0(G_{N-1}^{\mathcal{I}}) & \longrightarrow & H_0(G_N^{\mathcal{I}}) \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \\
0 & \longrightarrow & \mathcal{I}_0 & \longrightarrow & \mathcal{I}_1 & \longrightarrow & \cdots & \longrightarrow & \mathcal{I}_{N-1} & \longrightarrow & \mathcal{I}_N
\end{array}
$$

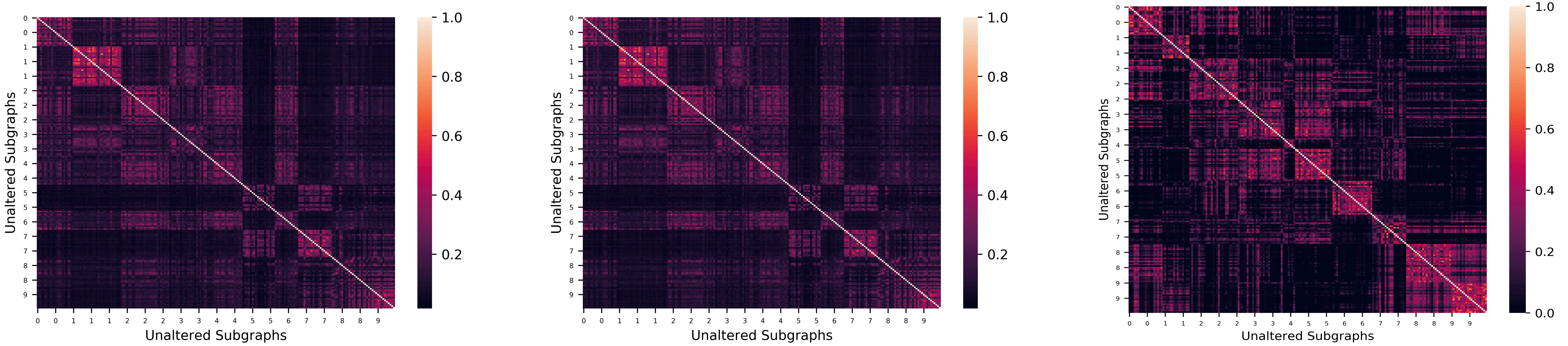
For some decomposition $\emptyset \subset \mathcal{I}_0 \subset \mathcal{I}_1 \subset \cdots \subset \mathcal{I}_N = \mathcal{I}$ of input $\mathcal{I} \in \mathbb{R}^n$



Homology generator similarity mirrors image space similarity



MNIST FashionMNIST CIFAR10



Persistent subgraph structure is highly predictive for classification of input

Network	Subgraph SVM Accuracy	Network Accuracy	Recovery Accuracy
CCFF-Relu	89.3%	97.6%	70.3%
CCFF-Sigmoid	89.1%	88.8%	83.4%
CCFF-Relu	89.3%	90.0%	80.3%
CCFF-Sigmoid	85.3%	84.7%	79.3%

Task-relevant information is retained by the generators

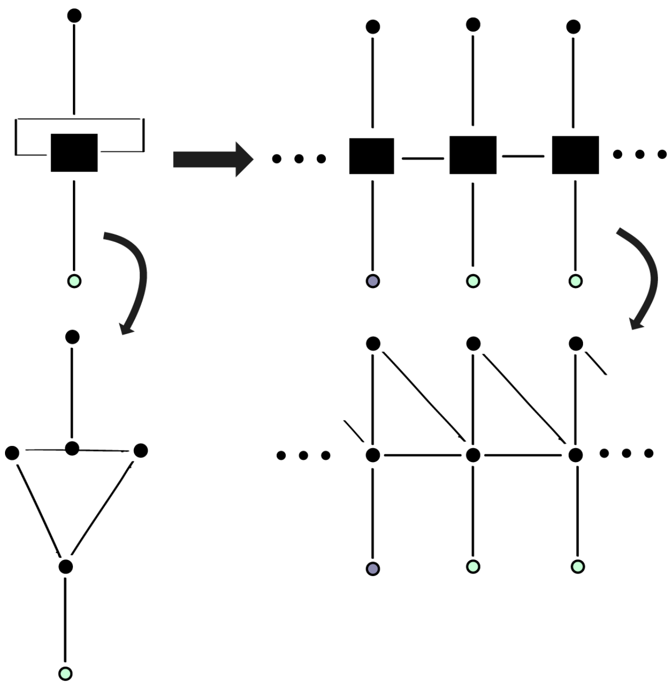
Future Work

Future Work

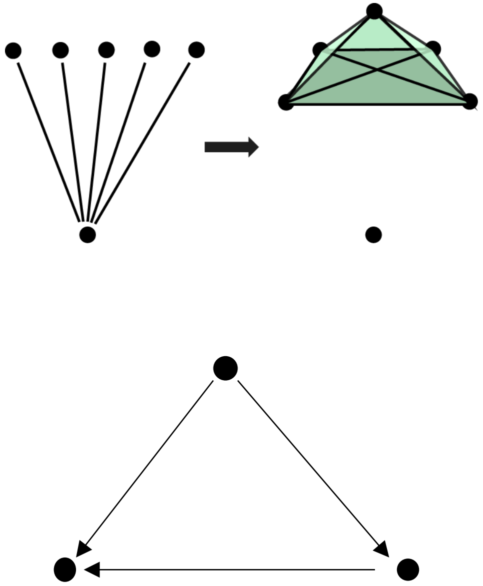
More Math

$$
\begin{array}{ccccccccccc}
0 & \longrightarrow & H_0(G_0^{\mathcal{I}}) & \longrightarrow & H_0(G_1^{\mathcal{I}}) & \longrightarrow & \cdots & \longrightarrow & H_0(G_{N-1}^{\mathcal{I}}) & \longrightarrow & H_0(G_N^{\mathcal{I}}) \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \\
0 & \longrightarrow & \mathcal{I}_0 & \longrightarrow & \mathcal{I}_1 & \longrightarrow & \cdots & \longrightarrow & \mathcal{I}_{N-1} & \longrightarrow & \mathcal{I}_N
\end{array}
$$

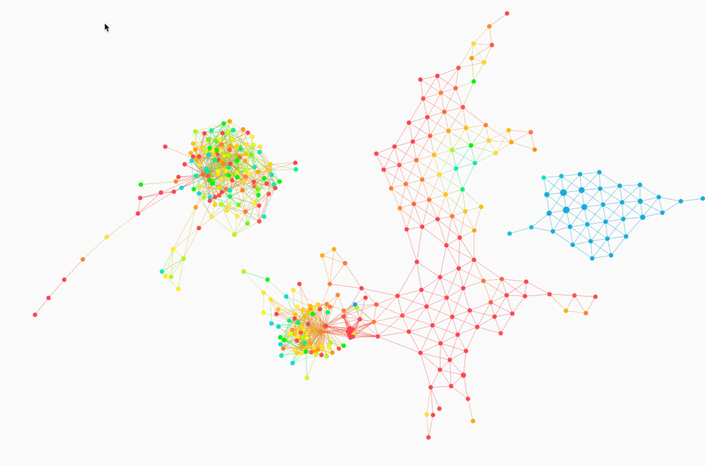
Different architectures

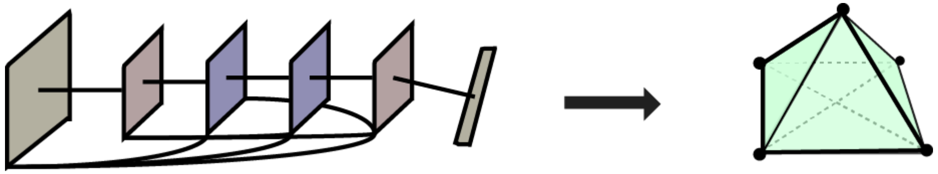


New complex constructions



Structure-preserving scaling



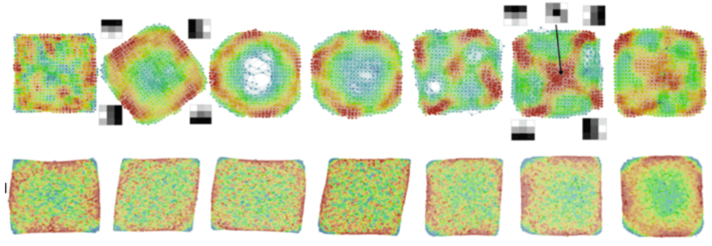


https://www.ayasdi.com/blog/bigdata/understanding-distinction-clustering-tda/

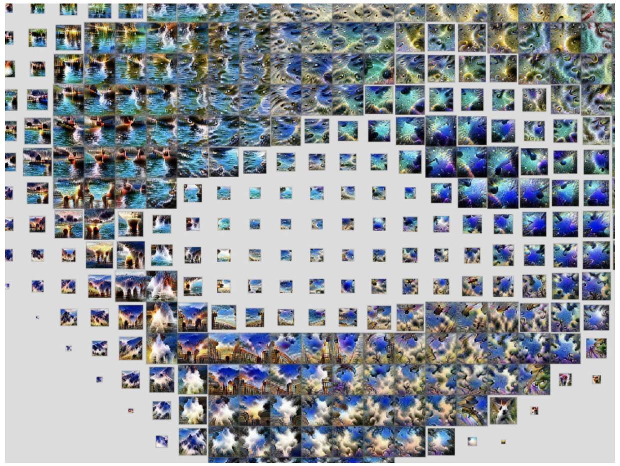
$$
\begin{array}{ccccccccc}
0 & \longrightarrow & H_0(G_0^{\mathcal{I}}) & \longrightarrow & H_0(G_1^{\mathcal{I}}) & \longrightarrow \cdots \longrightarrow & H_0(G_{N-1}^{\mathcal{I}}) & \longrightarrow & H_0(G_N^{\mathcal{I}}) \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{I}_0 & \longrightarrow & \mathcal{I}_1 & \longrightarrow \cdots \longrightarrow & \mathcal{I}_{N-1} & \longrightarrow & \mathcal{I}_N
\end{array}
$$

Input space homology

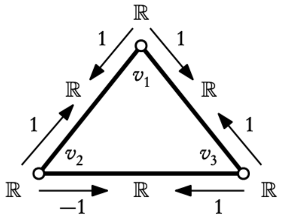
$$H_p(\mathcal{I})$$

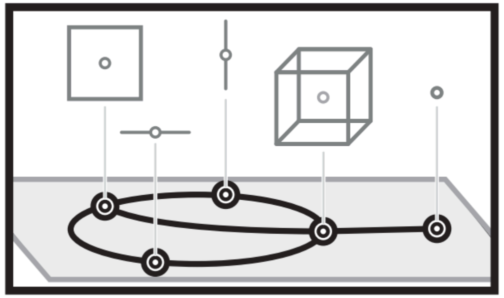


Visualization & Regularization



Sheaf-like constructions





https://www.ayasdi.com/blog/artificial-intelligence/using-topological-data-analysis-understand-behavior-convolutional-neural-networks/

https://www.math.upenn.edu/~jhansen/publications/gentleintroduction.pdf

More Info

Adversarial Examples Target Topological Holes in Deep Networks T Gebhart, P Schrater - *arXiv preprint arXiv:1901.09496, 2019*

Email: gebhart@umn.edu

References

- Maxime Gabella, Nitya Afambo, Stefania Ebli, and Gard Spreemann. Topology of learning in artificial neural networks*. arXiv preprint arXiv:1902.08160, 2019.*
- Bastian Rieck, Matteo Togninalli, Christian Bock, Michael Moor, Max Horn, Thomas Gumbsch, and Karsten Borgwardt. Neural persistence: A complexity measure for deep neural networks using algebraic topology. *arXiv preprint arXiv:1812.09764, 2018.*
- Ghrist, Robert W. *Elementary applied topology*. Vol. 1. Seattle: Createspace, 2014.
- Carter, et al., "Activation Atlas", *Distill*, 2019.