Inter-layer Information Similarity Assessment of Deep Neural Networks Via Topological Similarity and Persistence Analysis of Data Neighbour Dynamics

Andrew Hryniowski[1,2,3] and Alexander Wong[1,2,3]

[1] Vision and Image Processing (VIP) Research Group, University of Waterloo [2] Waterloo Artificial Intelligence Institute [3] DarwinAI Corp.

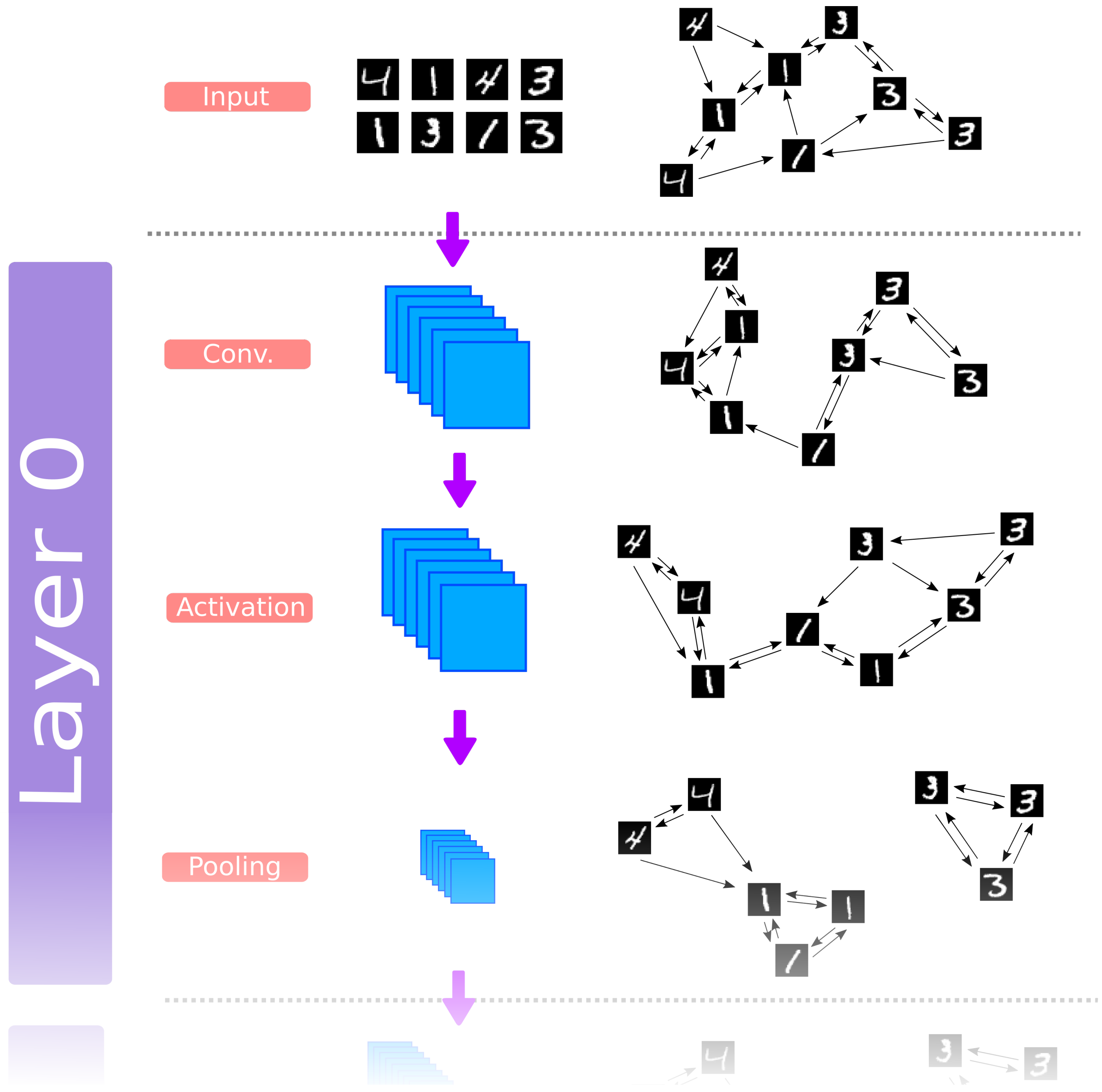
Abstract

- Quantitative analysis of information structure through a deep neural network (DNN) can unveil new insights
- Two very promising avenues related to this objective are 1) layer similarity (LS) and 2) intrinsic dimensionality (ID)
- We introduce two novel complimentary methods of analysis inspired by LS and ID methods
- **Nearest Neighbour Topological Similarity** (NNTS) for quantifying the information topology similarity between layers
- **Nearest Neighbour Topological Persistence** (NNTP) for quantifying the inter-layer persistence of data neighbourhood relationships throughout a DNN.
- The proposed strategies facilitate the efficient inter-layer information similarity assessment by leveraging only local topological information

Nearest Neighbour Graphs of DNN Feature Maps

- For each given operation in a DNN generate the nearest neighbour graph (NNG) using the feature maps of every sample
- The NNG uses directional neighbours
- Each sample has k neighbours

Example of NNGs on LeNet-5 MNIST (k=2)



Nearest Neighbour Topological Similarity

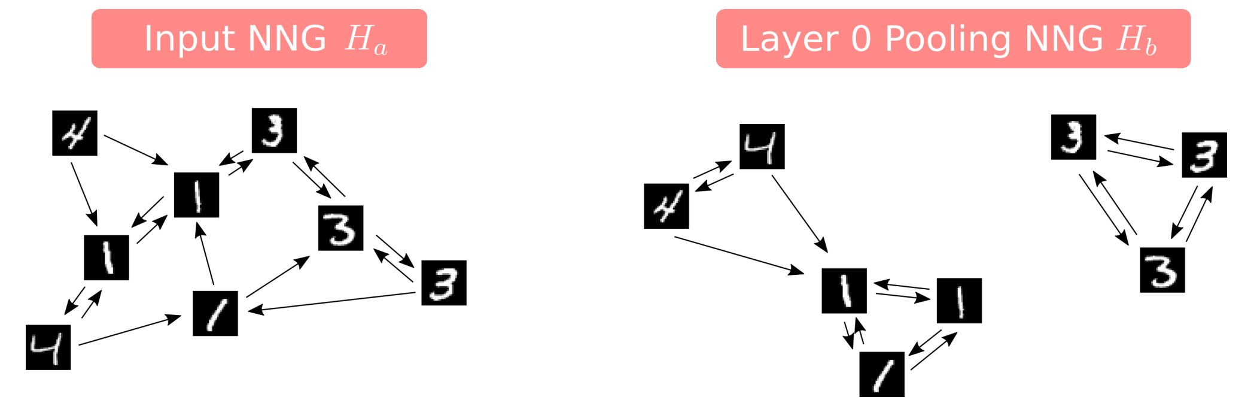
Idea

Compare the similarity of data representation between two operations (v_i, v_j) using nearest neighbour topological similarity (NNTS):

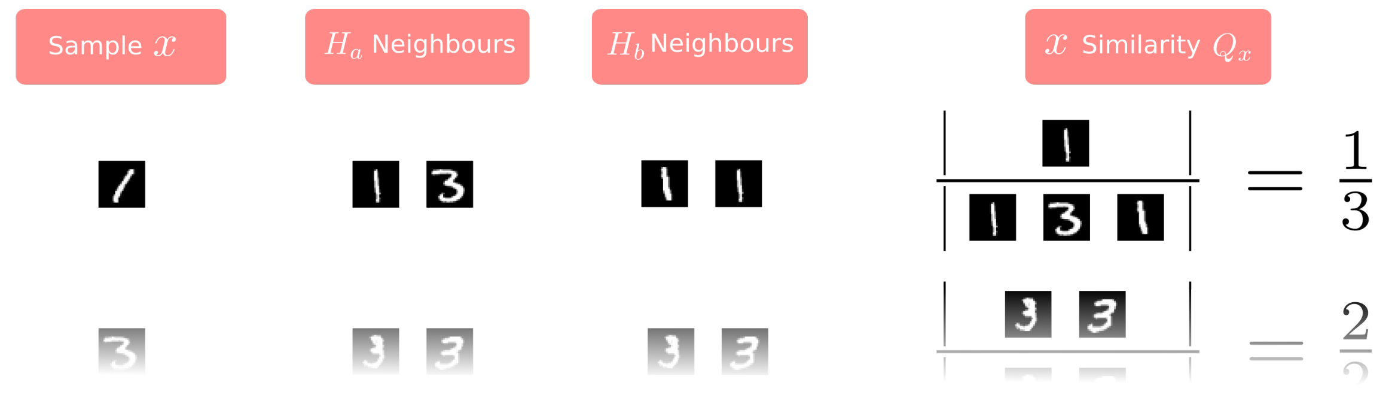
1. Calculate the NNG for each of the two operations
2. Score each sample based on the IOU of neighbours between operations
3. Take the average score of all samples between two layers

Example of NNTS between two operations in LeNet-5 MNIST

1. NNG Calculation



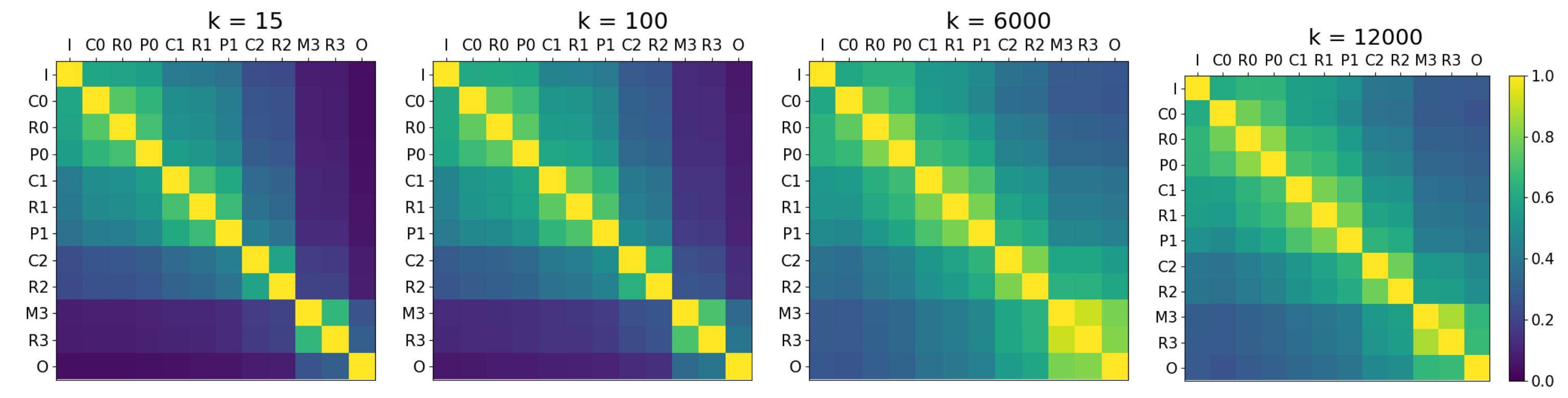
2. Compare samples between the graphs



3. Aggregate the results

$$Q(H_a, H_b) = \frac{1}{|X|} \sum_{x \in X} Q_x \approx 0.42$$

NNTS on the entire LeNet-5 MNIST model for different k



- Each operation is shown by a character followed by the layer number
- Notice the block-like pattern associated with each layer
- The small difference between $(P1, M3)$ and $(R2, M3)$ indicates that layer 2 may not be fully utilized
- As the number of k nearest neighbours increases from $k = 15$ to $k = 6000$ there is an increase in the similarity between all layer pairs.
- $k = 6000$ to $k = 12000$ sees a decrease in similarity since samples are now just as likely to be neighbours of other number

Nearest Neighbour Topological Persistence

Idea

- NNTS reduces similarity to a single digit, finer detail may be useful
- Look at how long samples are neighbour
  - What operation does the neighbour relation start at/end at?
- Use nearest neighbour topological persistence (NNTP) to measure such relations between operations
  - For a given operation tuple (v_{first}, v_{last}), NNTP counts the number of neighbour pairs that first appear at operation v_{first} and last appear at operation v_{last}

NNTP on the entire LeNet-5 MNIST model (k=15)

Operation of last appearance (columns); **Operation of first appearance** (rows)

	I	C0	R0	P0	C1	R1	P1	C2	R2	M3	R3	O
I	455	122	6.99	5.95	2.51	3.02	36.0	0.99	0.59	3.94	0.32	28.7
C0		427	2.94	5.30	0.65	0.52	2.27	1.23	0.06	0.18	0.01	0.62
R0			550	2.72	0.57	0.61	10.2	0.13	0.02	0.09	0.01	0.46
P0				494	19.0	22.1	6.08	0.54	0.35	2.50	0.42	0.53
C1					110	114	286	1.71	0.17	0.82	0.10	2.83
R1						67.2	57.7	1.33	0.23	0.29	0.04	0.39
P1							205	4.53	2.54	0.57	0.12	0.55
C2								528	36.3	8.79	0.57	0.46
R2									291	104	10.4	168
M3										170	19.1	141
R3											149	128
O												180

*in the thousands

- Notice the large number of transient connection along the diagonal
- The number of connections present in the top right persist throughout all operations in the LeNet-5 model, indicating that they are likely to be true neighbours on the data's intrinsic manifold
- C2 and R2 seem to have little effect on the data manifold as they largely add persistent connections while allowing most other connections to pass though
- C2 seems to be undoing connections created by C1 as indicated by (C1, P1)'s large value of 286000

Future Directions

Future directions involve using NNTS and NNTP to study aspects of a DNN:

1. **DNN Architecture:** We plan on studying the effects of common atomic operations and blocks of operations on a DNN's NNGs
2. **Data Augmentation:** The NNG based approaches provide a useful structure to study how augmentations of a given sample affect its position relative to other samples throughout a DNN
3. **Learning:** NNG allows one to study the changing characteristics of a DNN's NNGs during learning