

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Manifold Structure of Artificial and Biological Neural Networks

Zhang Liu

Supervisor: Francesca Spagnuolo

Co-supervisors: Steven W. Zucker, Luciano Dyballa

April 13, 2022

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Motivation

Motivation: biological vision vs computer vision

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

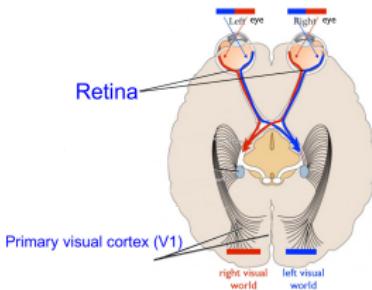


Figure 1: Early visual pathway [21].

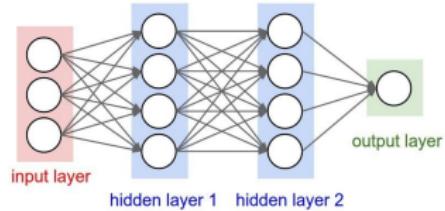


Figure 2: Illustration for a simple artificial neural network with two hidden layers.

- ① How do computer vision (CV) models compare to biological vision in terms of their respective **neural circuits**?
Neural circuit: the circuit in which the neurons connect to carry out a specific function.
- ② What specific mechanisms are important in causing such differences and/or similarities?

Significance

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- **Neuroscience:** modeling the visual system.
 - Despite decades of research, we have not fully understood the neural circuits responsible for visual perception [12].
- **CV:** reverse-engineer desired features of biological vision.
 - Neocognitron [9], CNN [17] were inspired by earliest models of primate vision [14].
 - CV performs poorly on visual relational reasoning:

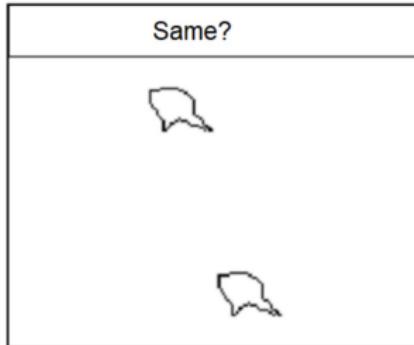


Figure 3: Same-different relations [8], [16].

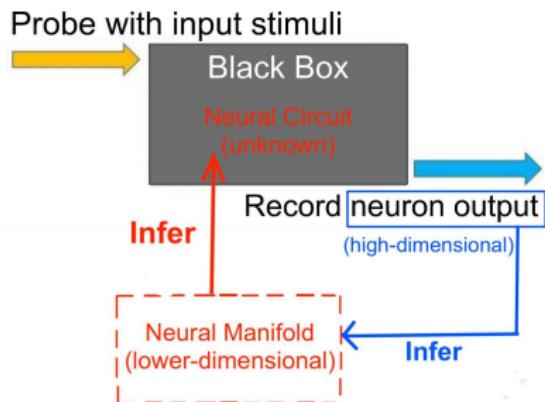
Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Framework of Comparison

Framework of comparison

Biological and artificial neural networks are black box problem: we do not know which are activated and how they connect to form neural circuits.

- ① Probe with appropriate input stimuli.
 - ② Obtain neuron output.
 - ③ Proxy for comparison:
neural manifold
(justified by the two-step inference).



What is a manifold?

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Definition (Homeomorphism)

Let $f : X \rightarrow Y$ be a function between topological spaces X and Y . If f is bijective, f and f^{-1} are continuous, then f is a homeomorphism; X and Y are said to be homeomorphic.

Definition (Manifold)

In brief, a real n-dimensional manifold is a topological space \mathcal{M} for which every point $x \in \mathcal{M}$ has a neighborhood homeomorphic to Euclidean space \mathbb{R}^n . Informally, a manifold is a topological space that locally resembles Euclidean space.

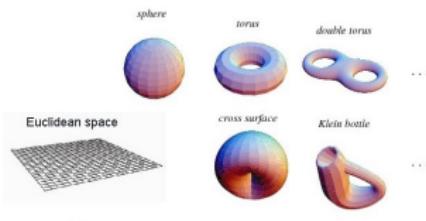


Figure 4: Examples of 2-dimensional manifolds

What is a neural manifold?

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References

Neural manifold is a lower-dimensional representation of the high-dimensional neural output.

- **“Manifold Hypothesis”**: natural data form lower-dimensional manifolds in their embedding space [19].
- Neuroscientific justification: all possible patterns of neural output are confined to a lower-dimensional manifold [10].

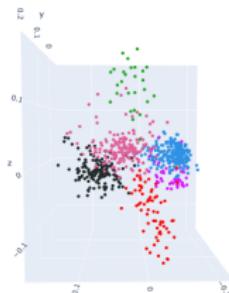


Figure 5: A 3D visualization of a neural manifold. Each point represents a neuron. Distance indicates how similar a pair of neurons respond.

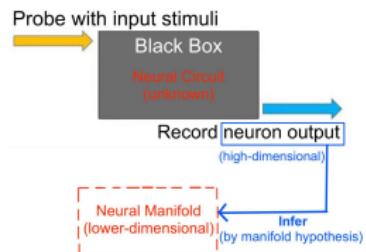


Figure 6

What can we analyze from neural manifold?

Neural manifold is an intermediate step from the neural output towards inferring the neural circuits. This step is justified by the **duality** between network and manifold.

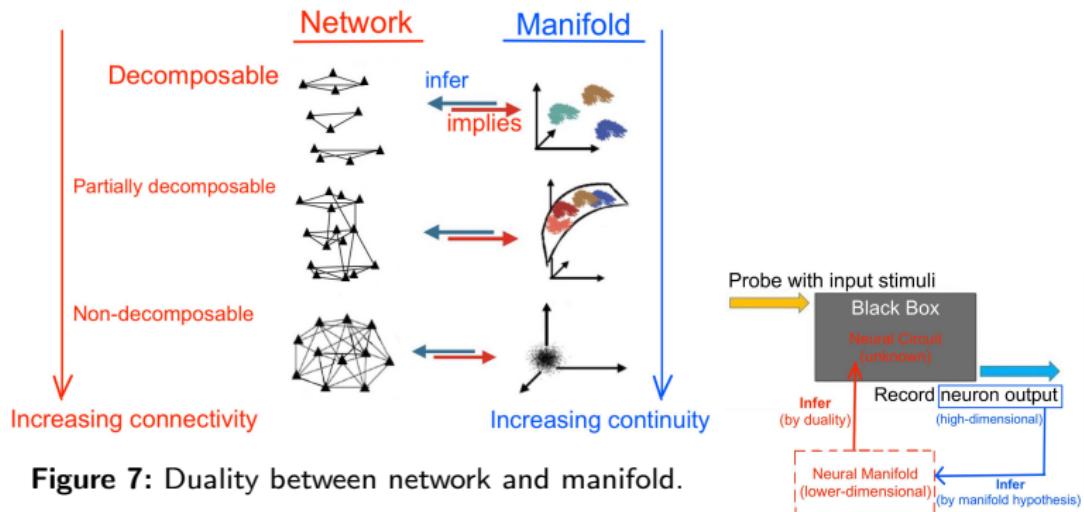


Figure 7: Duality between network and manifold.

Framework of comparison

	Biological	Artificial
Models	retina, V1	CNN, ViT, RNN
Input	flow stimuli [7]	ImageNet images [5]
Output	single-neuron recording	artificial neuron output

Table 1: Summary of comparison.

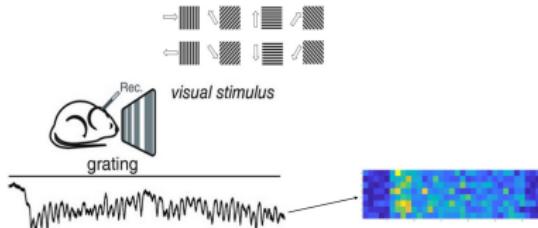


Figure 8: Input and output in biological neural networks.

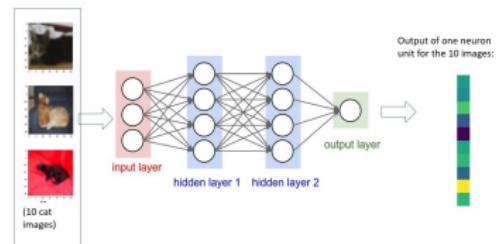


Figure 9: Input and output in artificial neural networks.

Starting point: retina vs V1

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Neural manifold for V1 is much more continuous than that for the retina, suggesting higher connectivity in the neural circuit of V1.

Questions: Do artificial neural networks in computer vision form continuous manifold like V1 or discontinuous manifold like the retina?

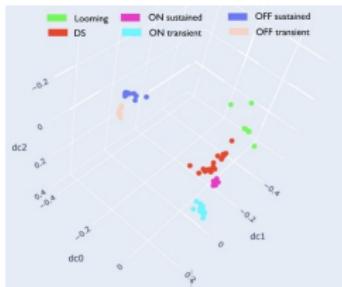


Figure 10: Retina neural manifold [6].

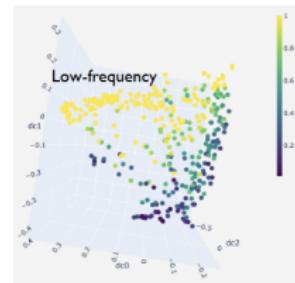


Figure 11: V1 neural manifold [6].

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Methods

Dimensionality reduction

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

The neuron output is encoded in neural tensor:

Definition (Neural tensors)

Each neural tensor encodes the output of N neurons to K flow stimuli over T time steps and is a 3-way N -by- K -by- T tensor. An N -way tensor is an element of the tensor product of N vector spaces. A 2-way tensor is a matrix.

The neural tensor has a high-dimensional representation (N , K , T are very large) and its structure cannot be visualized directly.

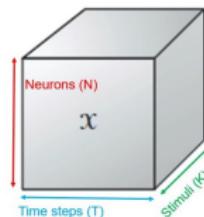
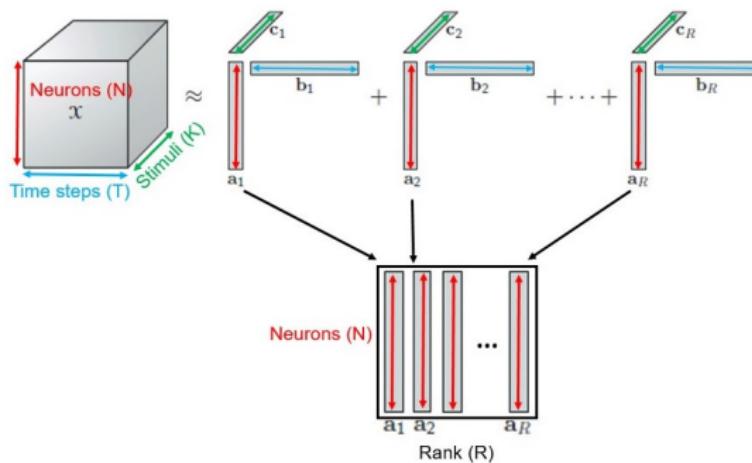


Figure 12: Visualization of the neural tensor.

Dimensionality reduction

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References

- ① We use Tensor CANDECOMP/PARAFAC (CP) decomposition with nonnegativity constraint, called "Nonnegative Tensor Factorization (NTF)."
- ② Apply NTF on the neural tensor to get the **neural factors**.
- ③ Concatenate the **neural factors** to get an N -by- R neural matrix.



Dimensionality reduction

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- ④ Apply PCA on the neural matrix (consisting of **neural factors**) to get the orthogonal principal components. This step has the same effect as other manifold learning methods such as diffusion map.
- ⑤ Use the first three principal components to plot the 3D projection. This provides a 3D visualization of the neural manifold.

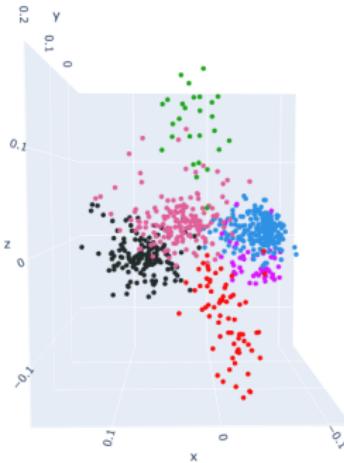


Figure 13: A 3D visualization of a neural manifold.



Quantifying continuity: mean flow ratio (ϕ_G)

- We adapt the mean flow ratio (MFR) developed in [6] to quantify the continuity of a given neural manifold.
 - $\phi_G \in [0, 1]$ and the greater the ϕ_G is, the more continuous the neural manifold is, and vice versa.

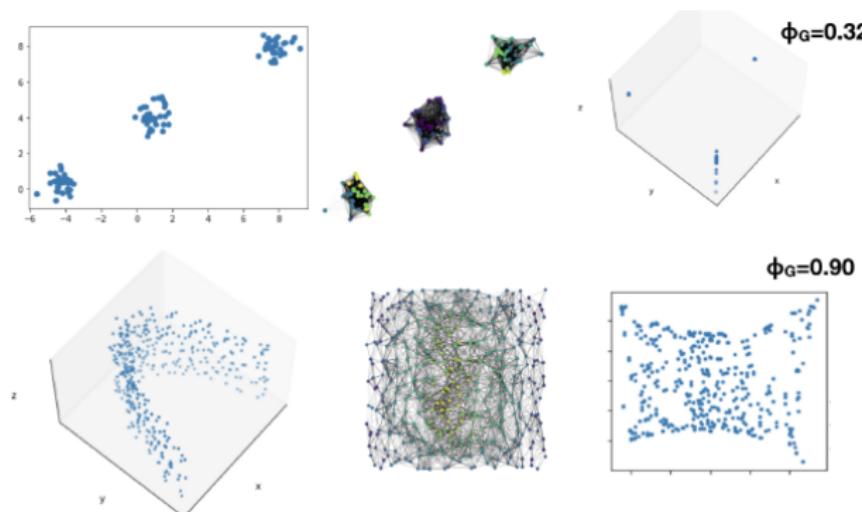


Figure 14: Intuition for mean flow ratio. (Adapted from [6].)

Motivation
Framework
Methods
**Results:
biological**
Results:
artificial
Conclusions
References
Q & A

Results: Biological Neural Networks

Retina: neural factors

Neural tensor with 698 neuron recordings:

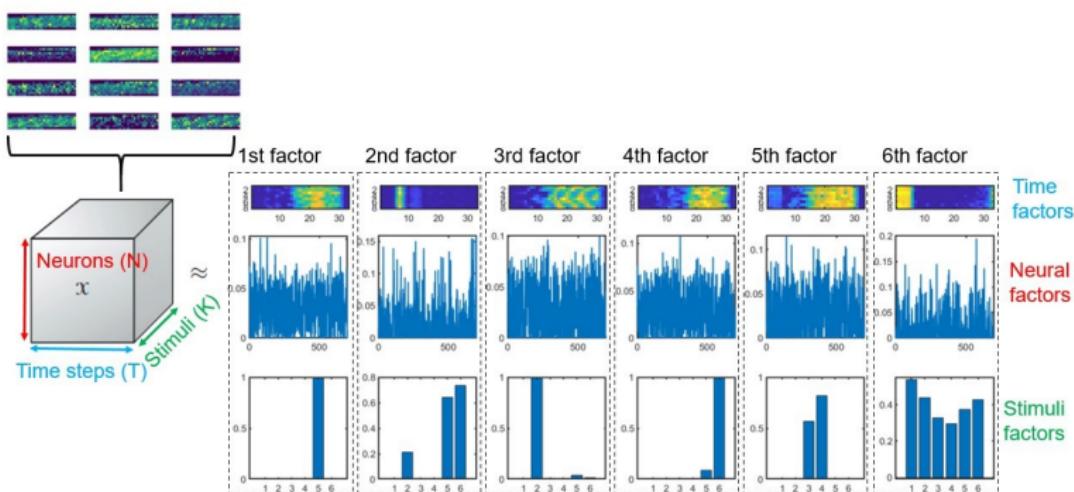
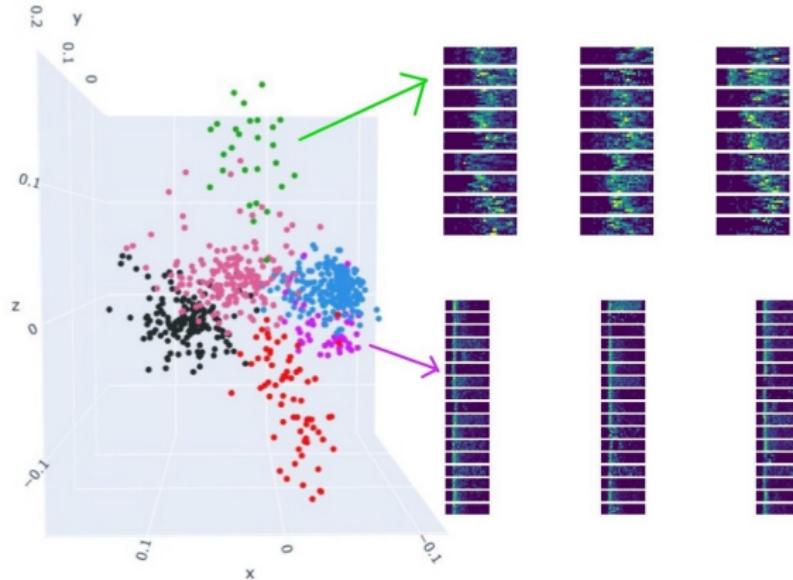


Figure 15: Visualizing the resulting tensor factors.

Retina: neural manifold

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- Neurons with similar firing patterns organize into distinct clusters.
- Retina neural manifold is discontinuous.
- Retina neural circuit form isolated subcircuits with low inter-connectivity.



Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Results: Artificial Neural Networks

CNN: algorithm for sampling neuron output

- Convolutional layer has several feature maps, each corresponding to a filter of certain size and weights.
- Neurons in the same feature map share the same weights and each attends to a subregion of the image (within its receptive field).

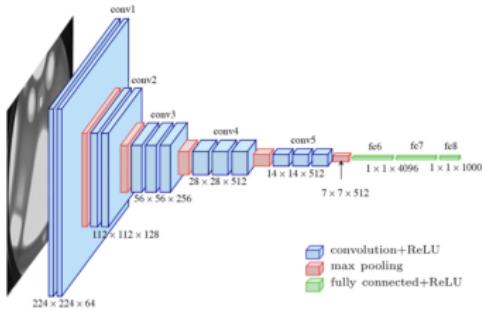


Figure 16: Structure of VGG16.

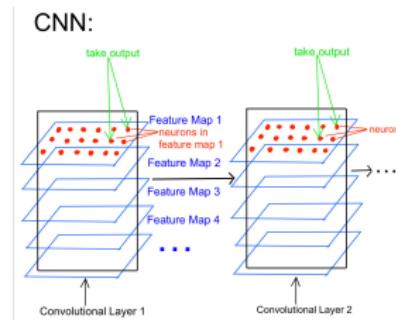


Figure 17: Taking neuron output from VGG16.

CNN: neural manifold

- Neurons from the same feature map form distinct clusters.
 - CNN neural manifold is discontinuous.
 - CNN neural circuit form isolated subcircuits with low inter-connectivity.
 - From shallow to deep layers, the clusters become more diffused but still distinct.

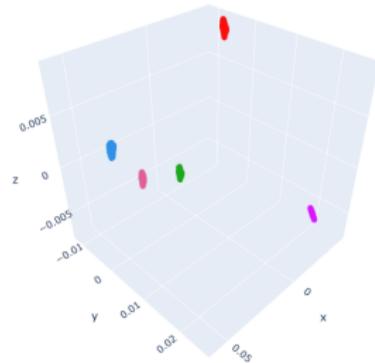


Figure 18: CNN neural manifold by sampling neuron output from convolutional layers in block 1 (shallow).

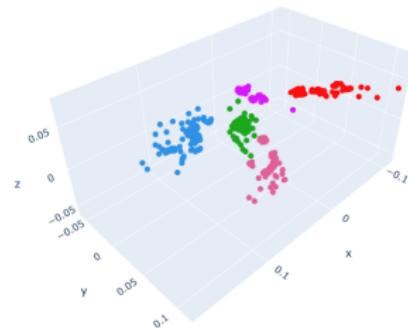
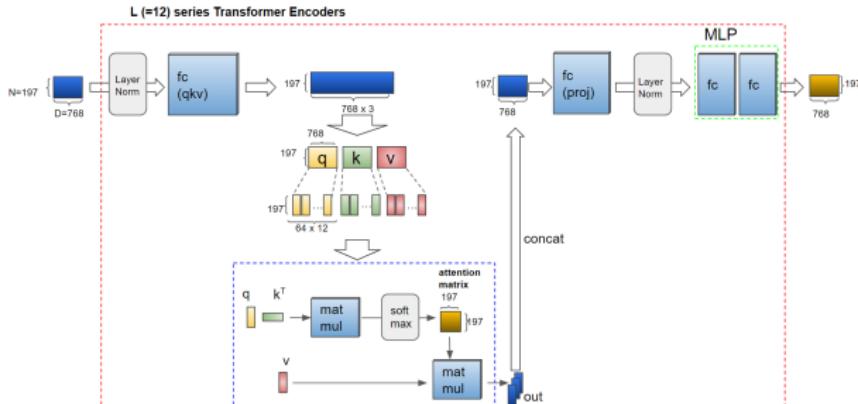
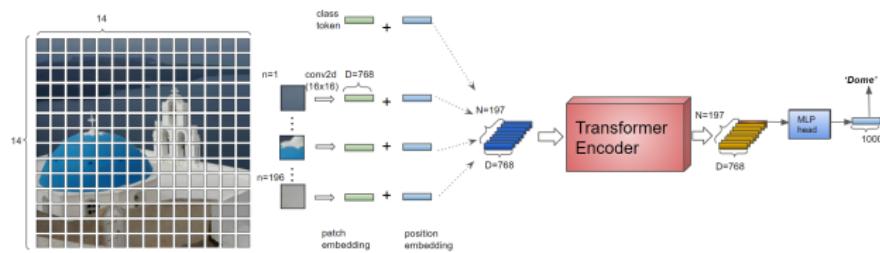


Figure 19: CNN neural manifold by sampling neuron output from convolutional layers in block 3 (deep).

Vision Transformer (ViT)

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Inspired by Transformer for NLP applications: an image is a sequence of 16-by-16 “words.” The main mechanism in ViT is the self-attention in the encoder layer.



ViT: algorithm for sampling neuron output

The diagram illustrates two stacked convolutional layers. Each layer consists of multiple feature maps. A green arrow labeled "take output" points to a specific neuron in Feature Map 1 of the first layer. Another green arrow labeled "take output" points to a specific neuron in Feature Map 1 of the second layer. The layers are labeled "Convolutional Layer 1" and "Convolutional Layer 2".

(a) Taking neuron output from CNN to build the corresponding neural tensor.

The diagram illustrates the ViT architecture. It starts with **196 image patches** (each having **768 features**) which are processed by a **Convolutional Layer** to produce **196 patch embeddings** (**196 neurons**). These embeddings, along with a **CLS token**, are fed into the first **Encoder Layer 1**. This layer contains a **Multi-Head Attention** block followed by an **MLP** block. The output of **Encoder Layer 1** is then passed to **Encoder Layer 2**, which also consists of a **Multi-Head Attention** block and an **MLP** block. Finally, the output of **Encoder Layer 2** is taken, and the **CLS token** is extracted as the final output.

(b) Taking neuron output from ViT to build the corresponding neural tensor.

ViT: neural manifold

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- Similar to CNN, the neural manifold is discontinuous, suggesting a neural circuit with limited inter-connectivity.
- Unlike CNN, the clusters remain the same shape through the layers
- This aligns with the findings in [4], [22]: convolution receptive fields are highly local and grow gradually; self-attention has global receptive field and more uniform representations between shallow and deep layers.

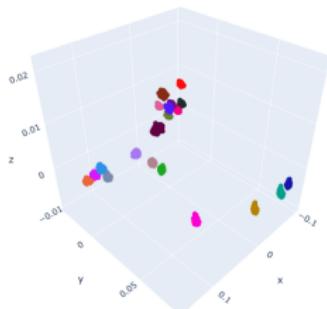


Figure 21: ViT encoder layer 1 neural manifold.

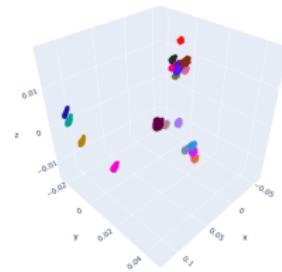


Figure 22: ViT encoder layer 12 neural manifold.

Qualitative comparison: neural manifold

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References

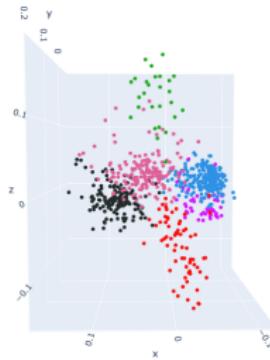


Figure 23: Retina.

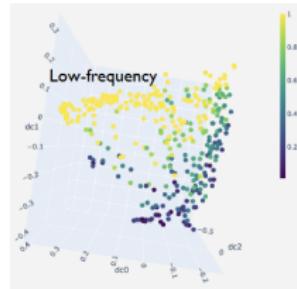


Figure 24: V1 [6].

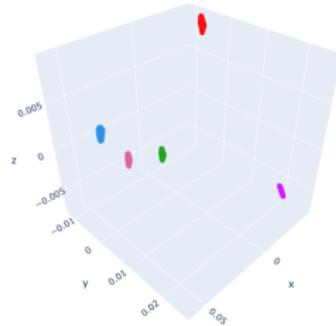


Figure 25: CNN.

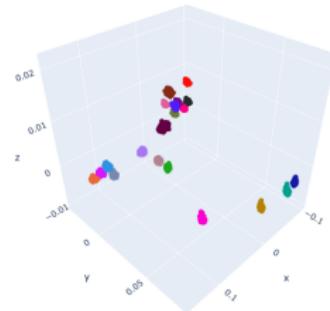
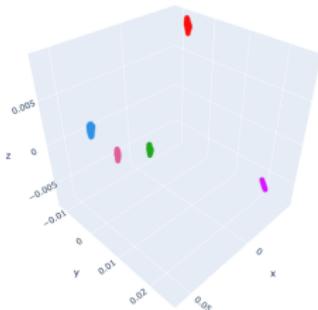


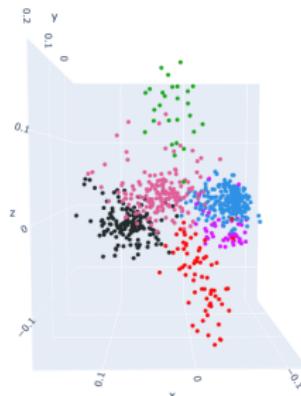
Figure 26: ViT.

Quantitative comparison: mean flow ratio (ϕ_G)

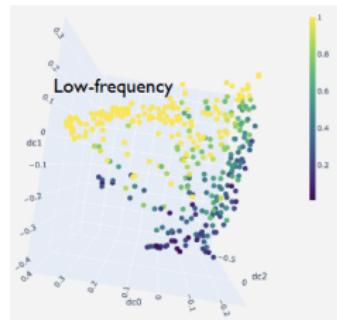
Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A



(a) $\phi_G = 0.27$ for CNN.



(b) $\phi_G = 0.56$ for retina.

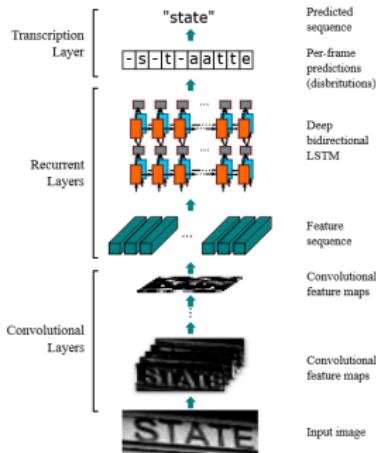


(c) $\phi_G = 0.91$ for V1 [6].

The quantitative comparison aligns with our qualitative comparison!

Hypothesis: adding recurrence

- Both CNN and ViT yield discontinuous neural manifold like the retina (limited connectivity in their underlying neural circuits).
 - Hypothesis: their feed-forward property leads to this limited connectivity. Incorporating recurrent connections might increase the continuity of the neural manifold (and hence connectivity of the neural circuit).
 - Next model to investigate: CRNN.



CRNN neural manifold

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- After adding recurrent layers to the base CNN model, clusters of neurons share significant overlaps.
- CRNN neural manifold is more continuous than CNN, suggesting greater inter-connectivity in the neural circuit.

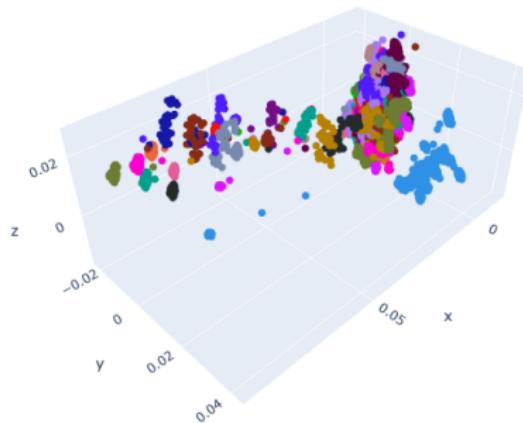


Figure 29: Neural manifold for CRNN (shallow layer).

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Conclusions and Future Directions

Conclusions

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- *How do computer vision models (VGG16, ViT, and CRNN) compare to biological vision, at retina and V1, in terms of their respective neural circuits?*

Conjecture 1. The underlying neural circuits in feed-forward networks, including CNN and ViT, are closer to that in retina than that in V1. Thus, contrary to popular belief, current computer vision models are poor approximations of the neural circuit of visual cortex.

- *What specific mechanisms are important in causing such differences and/or similarities?*

Conjecture 2. Incorporating recurrent mechanisms in computer vision models seems to be crucial in increasing the connectivity of the underlying neural circuits.

Highlights

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

During the project, we have made the following original contributions:

- ① conducted new computational experiments on neural data in the retina,
- ② developed original algorithms to investigate the manifold structure of CNN, ViT, CRNN,
- ③ compared and analyzed (both qualitatively and quantitatively) the manifold structure of CNN, ViT, and CRNN, against that of the neurobiological networks in retina and V1,
- ④ new results that have significant implications for the fields of neuroscience and computer vision.

Limitations and Future Directions

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

- Building on the empirical results, we can further develop a theoretical analysis on how recurrent connections increase the connectivity of neural circuits.
- Simply adding recurrent layers has limited effect. We are currently testing other CV models where recurrent connections play a more prominent role, in particular γ -net [18] that adds recurrence directly onto the convolutional layers.
- It would be of interest to extend the mean flow ratio analysis to different recurrent models for a more precise comparison.

Acknowledgements

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

I would like to thank my supervisor Prof. Francesca Spagnuolo and co-supervisor and mentor Prof. Steven W. Zucker and Dr. Luciano Dyballa for their generous guidance and advice. They have made possible the many serendipitous moments in this project.

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

References

References |

-  **How to visualize filters and feature maps in convolutional neural networks.**
<https://machinelearningmastery.com/how-to-visualize-filters-and-feature-maps-in-convolutional-neural-networks/>, 2021.
Accessed: 2021-11-08.
-  **Notes for cs231n: Convolutional neural networks for visual recognition.**
<https://cs231n.github.io/>, 2021.
Accessed: 2021-11-08.
-  **Abien Fred Agarap.**
Deep learning using rectified linear units (relu).
CoRR, abs/1803.08375, 2018.
-  **Zihang Dai, Hanxiao Liu, Quoc V. Le, and Mingxing Tan.**
Coatnet: Marrying convolution and attention for all data sizes, 2021.
-  **Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei.**
Imagenet: A large-scale hierarchical image database.
In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
-  **Luciano Dyballa.**
The Manifold of Neural Responses Informs Physiological Circuits in the Visual System.
Ph.D., Yale University, New Haven, 2021.
ISBN: 9798534651942 Publication Title: ProQuest Dissertations and Theses 28321876.
-  **Luciano Dyballa, Mahmood S. Hoseini, Maria C. Dadarlat, Steven W. Zucker, and Michael P. Stryker.**
Flow stimuli reveal ecologically appropriate responses in mouse visual cortex.
Proceedings of the National Academy of Sciences, 115(44):11304–11309, 2018.

References II

-  François Fleuret, Ting Li, Charles Dubout, Emma K. Wampler, Steven Yantis, and Donald Geman.
Comparing machines and humans on a visual categorization test.
Proceedings of the National Academy of Sciences, 108(43):17621–17625, 2011.
-  Kunihiko Fukushima.
Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.
Biol. Cybernetics, 36(4):193–202, 1980.
-  Juan A. Gallego, Matthew G. Perich, Lee E. Miller, and Sara A. Solla.
Neural manifolds for the control of movement.
Neuron, 94(5):978–984, 2017.
-  Xavier Glorot and Yoshua Bengio.
Understanding the difficulty of training deep feedforward neural networks.
In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.
-  Laura Gwilliams and Jean-Rémi King.
Performance-optimized hierarchical models only partially predict neural responses during perceptual decision making.
bioRxiv, 2017.
-  Dan Hendrycks and Kevin Gimpel.
Bridging nonlinearities and stochastic regularizers with gaussian error linear units.
CoRR, abs/1606.08415, 2016.

References III

-  D H Hubel and T N Wiesel.
Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.
The Journal of physiology, 160(1):106–154, January 1962.
-  Tim C. Kietzmann, Courtney J. Spoerer, Lynn K. A. Sørensen, Radoslaw M. Cichy, Olaf Hauk, and Niklaus Kriegeskorte.
Recurrence is required to capture the representational dynamics of the human visual system.
Proceedings of the National Academy of Sciences, 116(43):21854–21863, October 2019.
-  Junkyung Kim, Matthew Ricci, and Thomas Serre.
Not-so-clevr: learning same-different relations strains feedforward neural networks.
Interface Focus, 8(4):20180011, 2018.
-  Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner.
Gradient-based learning applied to document recognition.
Proceedings of the IEEE, 86(11):2278–2324, 1998.
-  Drew Linsley, Junkyung Kim, Alekh Ashok, and Thomas Serre.
Recurrent neural circuits for contour detection.
2020.
-  Christopher Olah.
Neural networks, manifolds, and topology, 2014.
-  Randall O'Reilly, Dean Wyatte, Seth Herd, Brian Mingus, and David Jilk.
Recurrent Processing during Object Recognition.
Frontiers in Psychology, 4, 2013.

References IV

-  **Jonathan Pillow.**
Vision: From Eye to Brain (Chap 3, Part II).
page 34, 2022.
-  **Maithra Raghu, Thomas Unterthiner, Simon Kornblith, Chiyuan Zhang, and Alexey Dosovitskiy.**
Do Vision Transformers See Like Convolutional Neural Networks?
page 13, 2021.
-  **Matthew Ricci, Junkyung Kim, and Thomas Serre.**
Same-different problems strain convolutional neural networks.
page 6, 2018.
-  **Daniel A. Spielman and Nikhil Srivastava.**
Graph sparsification by effective resistances.
2008.
-  **Courtney J. Spoerer, Patrick McClure, and Nikolaus Kriegeskorte.**
Recurrent Convolutional Neural Networks: A Better Model of Biological Object Recognition.
Frontiers in Psychology, 8, 2017.
-  **Sebastian Stabinger, Antonio Rodríguez-Sánchez, and Justus Piater.**
25 years of cnns: Can we compare to human abstraction capabilities?, 2016.
-  **Ruben S van Bergen and Nikolaus Kriegeskorte.**
Going in circles is the way forward: the role of recurrence in visual inference.
Current Opinion in Neurobiology, 65:176–193, December 2020.

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Thank you for listening! Any questions?

Which one is the neural manifold?

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Finding the true CP rank of a n -tensor is NP-hard for $n \in \mathbb{Q}$. Thus, we do not know the true CP rank of the neural manifold. The only thing we can do is to approximate the rank using repeated dimensionality reduction steps. The neural manifold is a lower-dimensional manifold spanned by the independent neural patterns, which is thus the best intermediate representation of the original neuron output in the sense that it requires the least number of dimensions while still preserving the structure of the original neuron output. Both the neural tensors and the 3D projection plots are only approximations of the true neural manifold (since we DO NOT know the true intrinsic dimension!)

More on neural manifold

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Neural manifold is an intermediate step from the neural output towards inferring the neural circuits. On the neural manifold, each point represents a neuron. The distance between a pair of points indicates how similar the firing patterns of the corresponding neurons are, and equivalently, how likely they participate in the same neural circuit. To infer the (unknown) neural circuit from a neural manifold (known), we note that there are three possible cases: if neurons were sampled

- ① from a collection of isolated neural circuits (decomposable networks), then the neurons would have distinct stimuli preferences, yielding a discontinuous neural manifold.
- ② from overlapping neural circuits (partially decomposable networks), most neurons would respond to multiple stimuli, leading to a smooth transition in stimuli preferences and a continuous neural manifold.
- ③ from fully connected neural circuits (non-decomposable networks), all neurons would respond to all stimuli, and there would be no stimuli preferences across groups of

Recurrence

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

However, recent research starts to suggest that these feed-forward neural networks fail to model various key features in early primate vision [20], [25], [23], [15], [27]. In addition, more recently, recurrent models of the visual system has inspired a new topic of research that focus on applying the idea behind recurrence model to show and/or address limitations of computer vision models. Among other similar works ([20], [25], [15], [27]), two notable recent works are [23], which proposed that feedback mechanisms including attention and perceptual grouping are key to visual relational reasoning tasks and [15], which showed that recurrence is necessary to model the representational dynamics in the visual system.

Visual reasoning

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

A growing number of research starts to reveal tasks that biological vision excels at while the state-of-the-art computer vision models fail. One notable example is that CNN is highly inferior to human vision in solving visual problems that involve relational reasoning [11], such as recognizing same-different relations in images [23], categorizing images based on the constituent parts [8], and comparing features in images [26]. The hope is that with a rigorous comparison, we can discover the specific mechanisms responsible for certain desired features of biological vision that the current computer vision models are lacking and adapt those mechanisms into computer vision models with relevant computational constraints in mind.

Algorithm for NTF for retina tensor

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Algorithm .1: Algorithm for NTF on retina tensor.

Data: Neural tensor for retina.

Result: R tensor factors for retina.

- 1 import data as X, $X = \text{tensor}(X)$
 - 2 chosen rank, $R = 35$
 - 3 $n_repetitions = 10$, initialize errors with zeros
 - 4 **for** r from 1 to R **do**
 - 5 **for** rep from 1 to $n_repetitions$ **do**
 - 6 NTF: $M = \text{cp_opt}(X, r, \text{'init'}, \text{'rand'}, \text{'lower'}, 0)$
 - 7 error = $\text{norm}(X - \text{tensor}(M)) / \text{norm}(X)$;
 - 8 add error to the array of errors
 - 9 $F_{\text{neuron}} = M.u\{1\}$, $F_{\text{stimuli}} = M.u\{2\}$, $F_{\text{time}} = M.u\{3\}$
 - 10 save the factors
-

Algorithm for computing the mean flow ratio

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Algorithm .2: Algorithm for computing the mean flow ratio.

Data: points on the neural manifold, $x_1, \dots, x_N \in \mathcal{M}$.

Result: mean flow ratio for the neural manifold, ϕ_G .

- 1 kernel bandwidth, $\varepsilon := \frac{1}{N} \sum_{i=1}^N \min_{j: x_j \neq x_i} \|x_i - x_j\|^2$
 - 2 Gaussian kernel^a, $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \varepsilon)$
 - 3 weight matrix W , $W_{i,j} := k_\varepsilon(x_i, x_j)$
 - 4 degree of node x_i , $d(x_i) := \sum_{j=1}^N W_{i,j}$
 - 5 anisotropic Gaussian kernel, $\tilde{k}_\varepsilon(x_i, x_j) := \frac{k_\varepsilon(x_i, x_j)}{d(x_i)d(x_j)}$
 - 6 anisotropic weight matrix, \tilde{W} , $\tilde{W}_{i,j} := \tilde{k}_\varepsilon(x_i, x_j)$
 - 7 build graph $G = (V, E)$, weight of the edge (x_i, x_j) is $\tilde{W}_{i,j}$
-

^aIn fact, we can choose any symmetric, positive semi-definite similarity kernel. But Gaussian kernel gives a physically intuitive construction as it effectively considers a ball of radius ε around each data point.

Algorithm .3: Algorithm for computing the mean flow ratio (cont.).

- 1 edge resistance, $R_{i,j} = 1/\tilde{W}_{i,j}$
 - 2 compute effective resistance $R_{i,j}^{\text{eff}}$, using [24]
 - 3 effective conductance, $C_{i,j}^{\text{eff}} = 1/R_{i,j}^{\text{eff}}$
 - 4 build graph G_C , weight of the edge (x_i, x_j) is $C_{i,j}^{\text{eff}}$
 - 5 **for** i from 1 to N **do**
 - 6 **for** j from 1 to N , $j \neq i$ **do**
 - 7 SumAll += maxflow(i, j)
 - 8 **for** k : k is a neighbor of i **do**
 - 9 SumNeighbors += maxflow(i, k)
 - 10 flow ratio of node x_i , $\rho_i = \text{SumAll}/\text{SumNeighbors}$
 - 11 mean flow ratio of graph G , $\phi_G = \sum_{i=1}^N \rho_i.$
-

Algorithm for sampling neurons from CNN

Algorithm .4: Algorithm for CNN artificial neural tensor.

- 1 **Function** compute_neuron_output(*model*,
image_all_shifts):
- 2 **for** *layer* in all convolutional layers **do**
- 3 pass the image through the model
- 4 take the neuron output from the specified layer
 after ReLU
- 5 (optional) remove the neurons at the edges
- 6 compute the average neuron output in each
 feature map
- 7 select the feature maps with highest average
- 8 normalize the neuron outputs in each feature map

Algorithm for sampling neurons from ViT

Algorithm .5: Algorithm for ViT artificial neural tensor.

```
1 Function compute_neuron_output(model,  
    image_all_shifts):  
2     for layer in all encoder layers do  
3         pass the image through the model  
4         take the neuron output from the specified layer  
5         apply ReLU activation on the neuron output  
6         compute the average of neuron outputs in each  
           feature  
7         select the feature maps with highest average  
8         normalize the neuron outputs in each feature map
```

CNN: spatial ordering of neurons is preserved

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

The spatial ordering of neurons' receptive fields are continuously mapped onto the manifold.

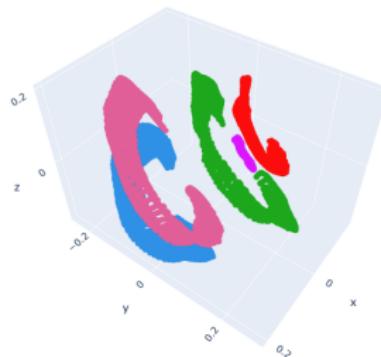
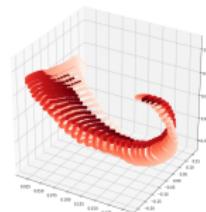
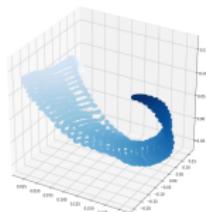


Figure 30: VGG16 block 1 neural manifold with spatial order preserved.



ViT: spatial ordering of neurons is preserved

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References

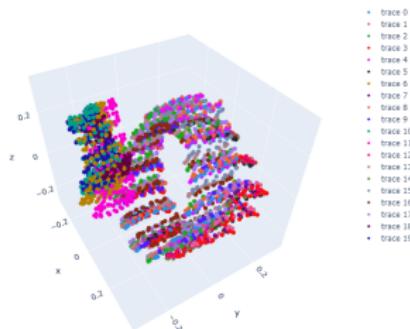


Figure 33: ViT encoder layer 1 neural manifold with spatial order preserved.

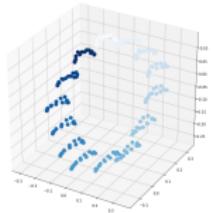


Figure 34: First cluster of neurons organized by vertical order.

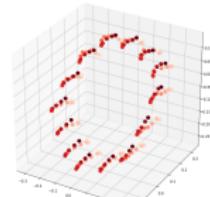


Figure 35: First cluster of neurons organized by horizontal order.

Remarks I

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Remark

For ease of reference, we also use the term “neural manifold” for the point-cloud of data points with an underlying manifold structure. In addition, neural manifolds from real neural data are often no longer “manifolds” in a mathematical sense, due to the sparse input sampling and the presence of neural noise.

Remarks II

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Remark

At first it might seem that using different stimuli for biological and artificial neural networks will lead to an unfair comparison. However, this approach is appropriate for the following reason: The mouse visual system was trained on the flow stimuli is shown to resemble the naturalistic visual information that the mouse's vision has adapted to^a. The visual stimuli for VGG16 model came from the same data set that it is trained on (ImageNet). Thus, in both cases, the visual stimuli are similar to what the respective network was trained to recognize.

^aBased on the paper that proposed flow stimuli [7], flow stimuli are more like what the mouse would encounter in the natural world than are sine-wave gratings but is more tractable for analysis than are natural images.

Remarks III

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

Remark

On line 5, we need to impose ReLU^a activation to make the neuron output non-negative. The non-negative constraint is desired because the tensor factors from NTF are more interpretable. The pre-trained ViT uses GELU^b as the activation function. This results in negative values and would prevent us from using non-negative tensor decomposition. Fortunately, the negative portion of GELU is negligible given the purpose of our investigation, which justifies this step.

^aReLU: Rectified Linear Unit [3].

^bGELU: Gaussian Error Linear Unit [13].

“Visual stimuli” in CNN

Analogous to the flow stimuli in lab experiments, the “visual stimuli” for CNN are natural images of different objects selected from the benchmark dataset ImageNet [5]. To keep the size of the artificial neural tensor manageable, we randomly selected 20 images from four classes as the visual stimuli. In order to simulate the movement of flow stimuli over time, we create multiple shifts of the original image in vertical and horizontal direction, with the shift step equal to the size of the filter.

Analogy II

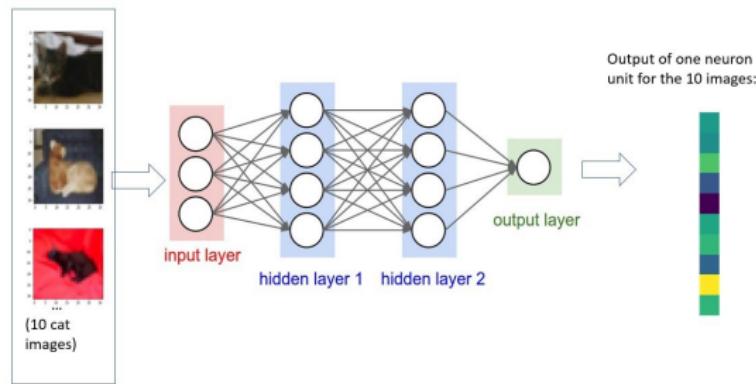


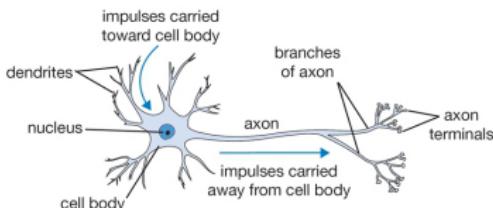
Figure 36: Computing output of one neuron unit in the CNNs given 10 input cat images.

Analogy III

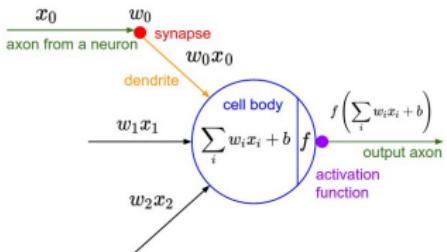
Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

“Individual neuron” in CNN In the lab experiments for biological neural networks, moving visual stimuli trigger spikes in the activation potential in the neurons in the retina. The individual neuron response is the firing rate of the neuron which is dependent on the electrostatics processes that take place in the synapse. The biological neuron is coarsely modeled by the neuron unit in the CNNs. At the “dendrite,” the input is taken from the “axons” of the connected neurons.¹ The weight of each neuron models the synaptic process in biological neuron. At the “cell body,” all the inputs are multiplied with the weights and summed. We then add the bias term to the sum and apply the nonlinear activation to obtain the final output from the individual neuron unit at the “axon.” The parallel between biological and artificial neuron is shown in the figures below.

Analogy IV



(a) An individual neuron in biological neural networks. Adapted from [2].



(b) An "individual neuron" in CNNs. Adapted from [2].

"Receptive field" of an artificial neuron Each layer in the CNN has several different feature maps. Each feature map corresponds to a filter of a specific size and specific weights. The filter is essentially a matrix that, when taking inner product with the input image, give an output that highlights some specific features in the input image. To illustrate this, we can visualize the 64 feature map in the first convolutional layer of the VGG16 model, given the input image.

Analogy V

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A



(a) Input image. Adapted from [1].



(b) Visualizing the 64 feature maps in the first convolutional layer of the VGG16 model. Adapted from [1].

The neurons within each feature map share the same weights and each attends to a specific subregion of the image which is referred to as their receptive field. This term originated from biological vision defined as a restricted region of visual space where a stimulus could illicit responses in a retinal ganglion cell. The size of the receptive field of a neuron is the same as the filter size corresponding to the feature map in which the neuron is located.

Analogy VI

Motivation
Framework
Methods
Results:
biological
Results:
artificial
Conclusions
References
Q & A

“Neuron population response” in CNN Analogous to biological neural networks, we computed the neural output given the input image. Unlike biological neurons, most of the artificial neurons will have output of small values because most pre-trained weights decay to zero. In other words, the “firing” of neurons in CNN will be sparse. We select the feature maps that contain the neurons with highest average firing rate.

¹Note, however, that in our experiments, instead of taking the input from previously connected neuron unit, each neuron takes the image as the input.