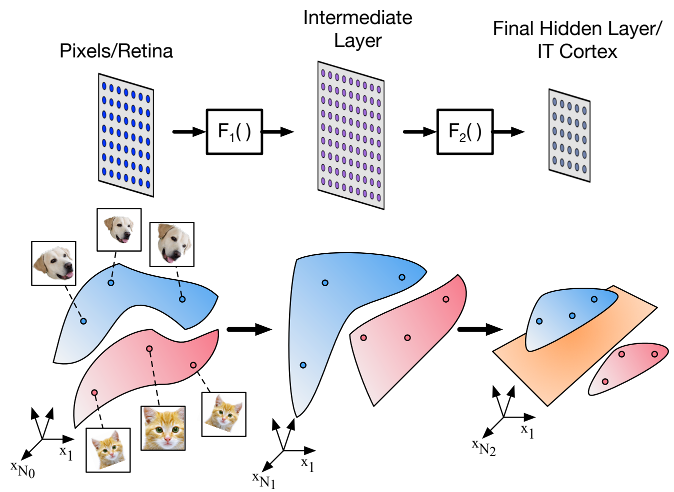
Processing of Object Manifolds in Deep Networks and the Brain

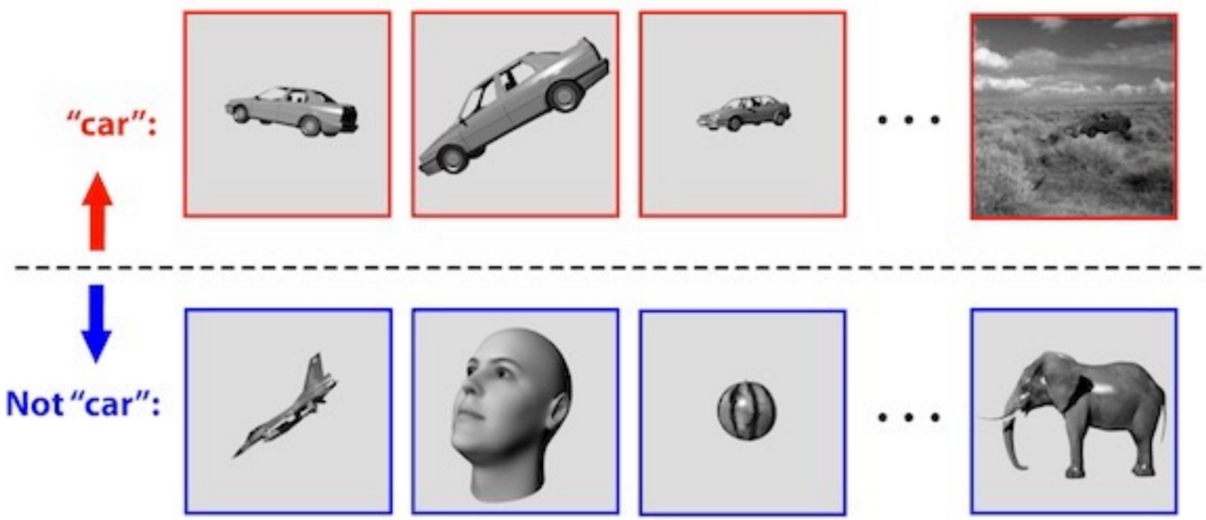


SueYeon Chung

Columbia University

Guest Lecture, Caltech
June 1, 2021

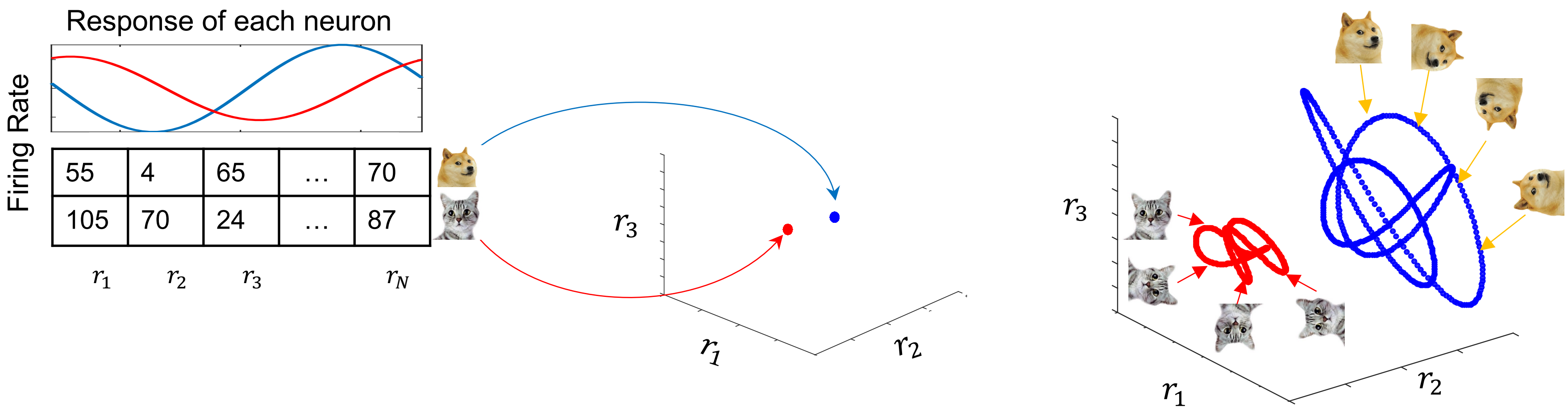
The brain needs to identify objects despite stimulus variability



Examples of variability: rotation, scaling, pose, background…

DiCarlo, James J., Davide Zoccolan, and Nicole C. Rust. "How does the brain solve visual object recognition?." *Neuron* 73.3 (2012): 415-434.

Object Manifolds

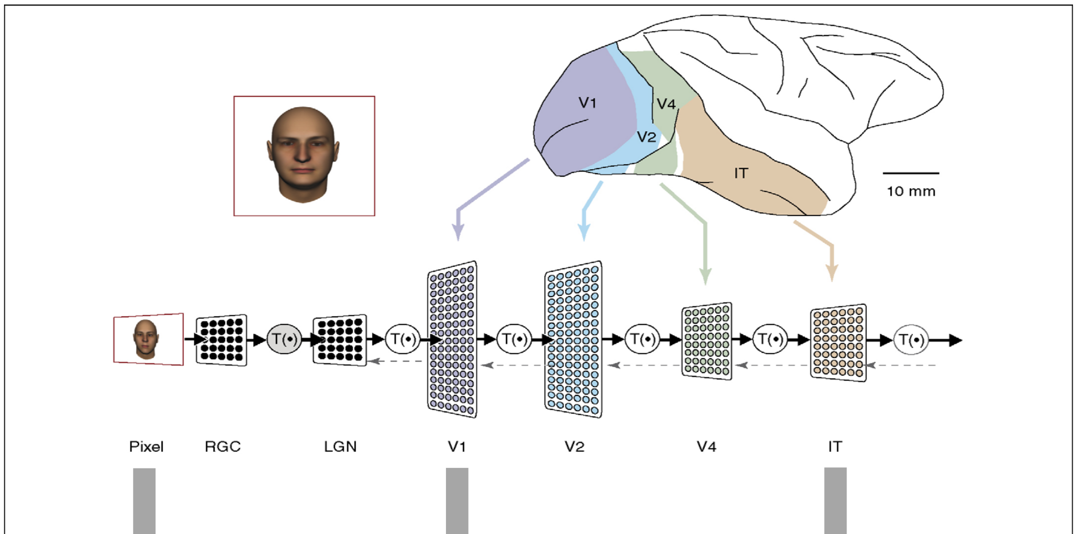


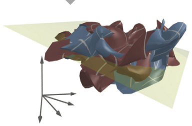
With stimulus variabilities

DiCarlo, James J., Davide Zoccolan, and Nicole C. Rust. "How does the brain solve visual object recognition?." *Neuron* 73.3 (2012): 415-434.

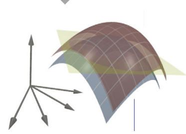
“Untangling” object manifolds by layers of sensory processing

In Visual Cortex…

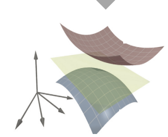




Pixel: Poor
(Nonlinearly Separable)

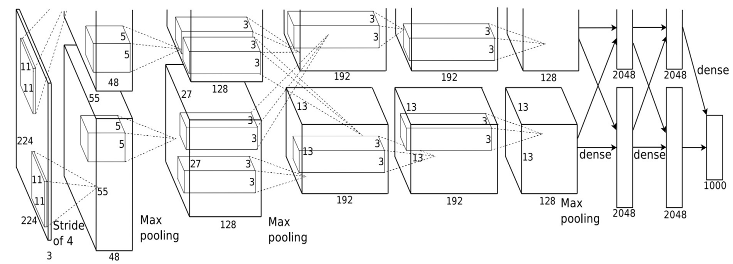


V1: Intermediate
(Nonlinearly Separable)



IT: Good
(Linearly Separable)

In Deep Networks…

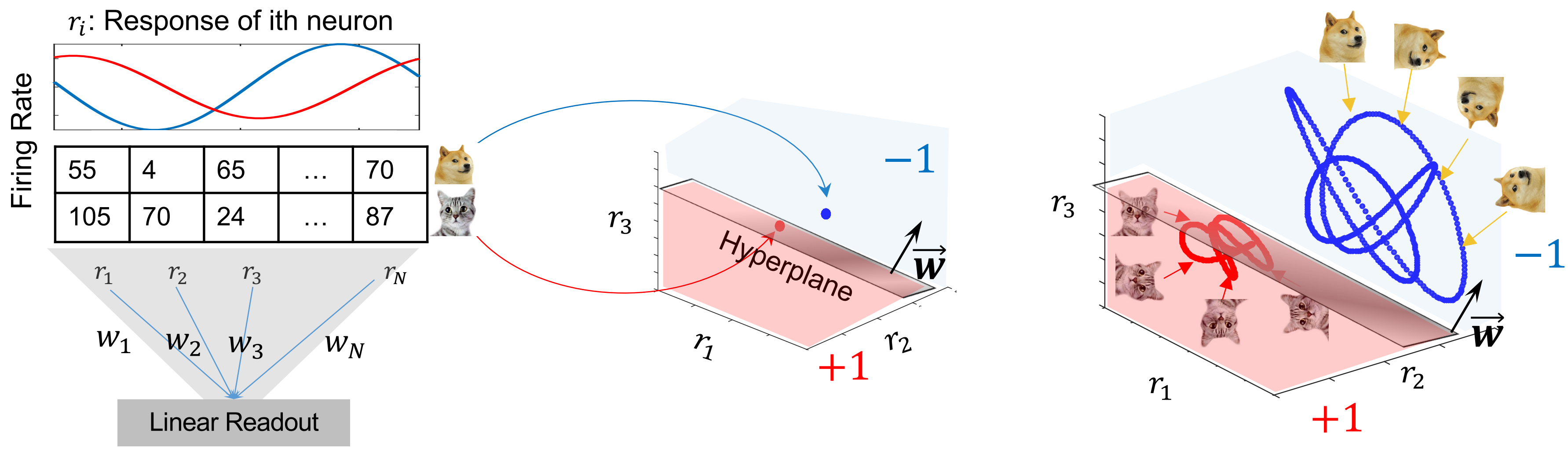


Krizhevsky, Sutskever, and Hinton. *NIPS.* 2012.

Untangling: reformatting manifolds across the layers to increase
linear separability

DiCarlo and Cox. *Trends in cognitive sciences*. 2007

Invariant object discrimination as **linear separation** of **manifolds**



$$\text{sign}\,(\boldsymbol{w}\cdot\boldsymbol{r}) \begin{cases} +1 & (if\ cat) \\ -1 & (otherwise) \end{cases}$$

Perceptron

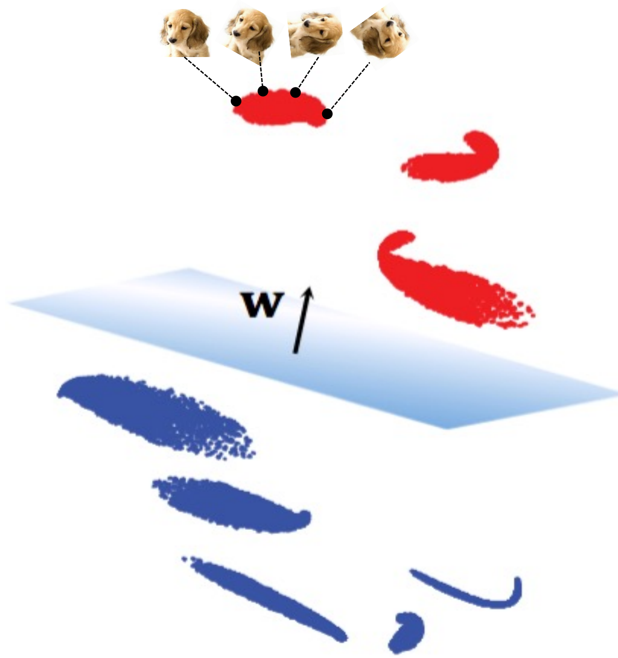
With stimulus variabilities

Outline

1. Introduction
2. **Theory of Linear Classification of Object Manifolds**
3. Object Manifolds in Visual Hierarchy
4. Object Manifolds in Auditory Hierarchy
5. Object Manifolds in Language Hierarchy
6. Understanding Generalization Dynamics using Object Manifolds

Which geometric properties determine the linear separability of manifolds?

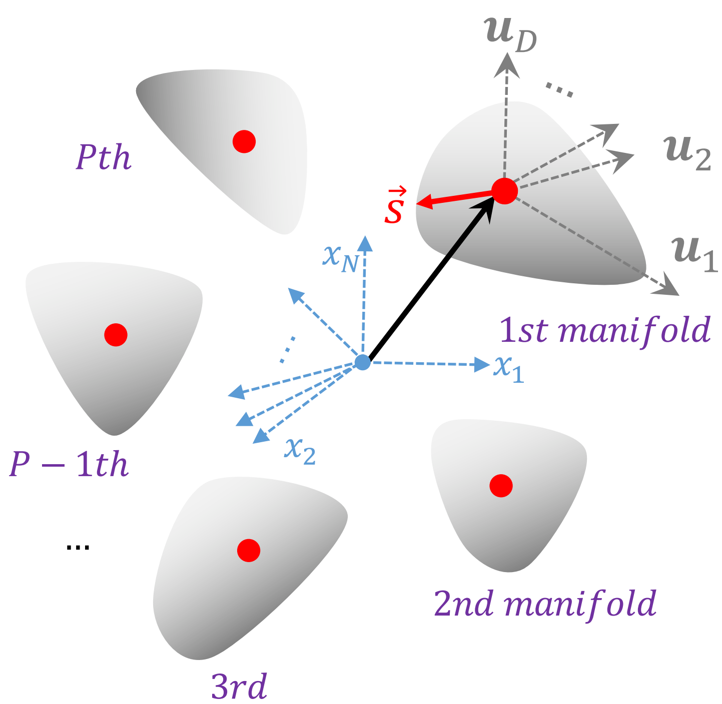
Statistical Mechanical Theory of Linear Classification of **Manifolds**



SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. ”Classification and Geometry of General Perceptual Manifolds." *Physical Review X* (2018)

Model of Object Manifolds

(or manifold-like object representations)



N: ambient dimension for data.

D: subspace spanned by the manifold

Each Point: $\boldsymbol{x}^{\mu} = \boldsymbol{x}_0^{\mu} + \sum_{i=1}^{D} s_i \boldsymbol{u}_i^{\mu}$,

Center: $\boldsymbol{x}_0^{\mu} \in \mathcal{R}^N$

Directors: $\boldsymbol{u}_i^{\mu} \in \mathcal{R}^N$

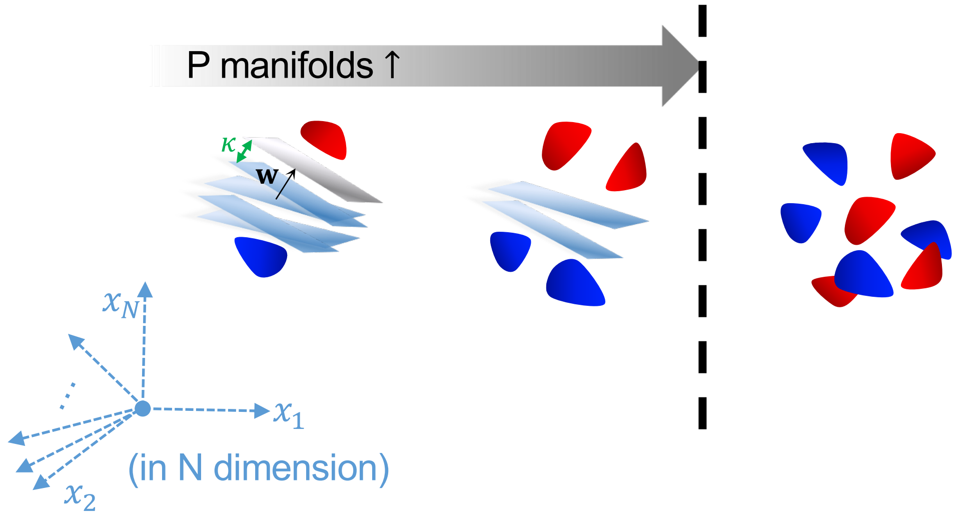
Shape: $f(\vec{s}) \leq 0$, $\vec{s} \in \mathcal{R}^D$

P: number of manifolds

No need to be smooth (i.e. data clouds ✅)

- Statistical Assumptions about **P manifolds**
  - Centers $\vec{x}_0$ are randomly oriented
  - Manifold subspaces are randomly oriented

Capacity of object manifolds

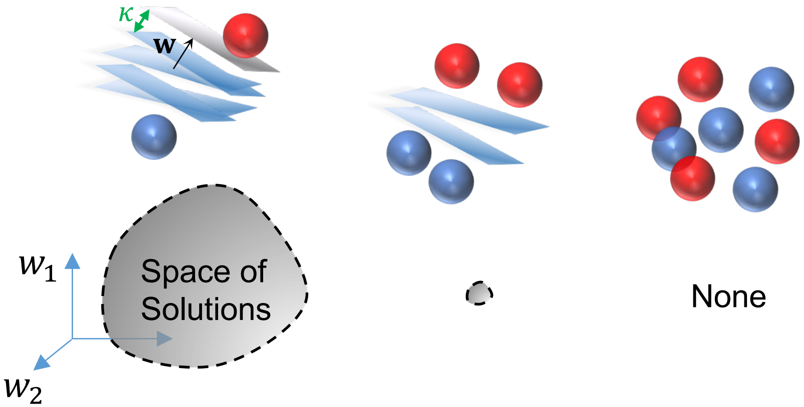


- Critical Manifold Capacity: maximum load (P/N) where most dichotomies of manifolds are linearly separable

- How is the geometry related to manifold capacity?

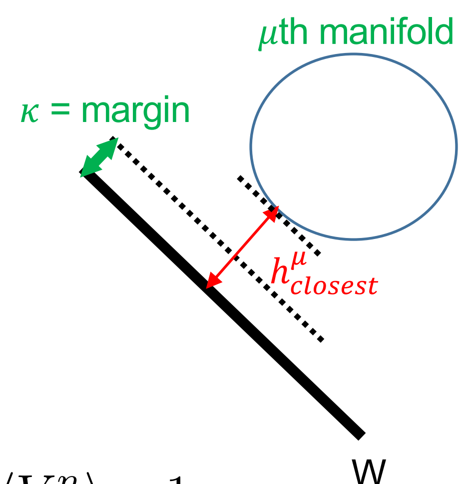
SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. "Linear readout of object manifolds." *Physical Review E* 93.6 (2016): 060301.

Statistical Mechanics: Volume of Solution



Critical capacity occurs when volume of valid weight vectors shrinks to zero

$$\boxed{V = \int d^N\vec{w}\,\delta(|\vec{w}|^2 - N)\prod_{\mu=1}^{P}\Theta\left(h^{\mu}_{closest} - \kappa\right)}$$



Θ : heavyside step function
1 when argument >0
0 when argument <=0

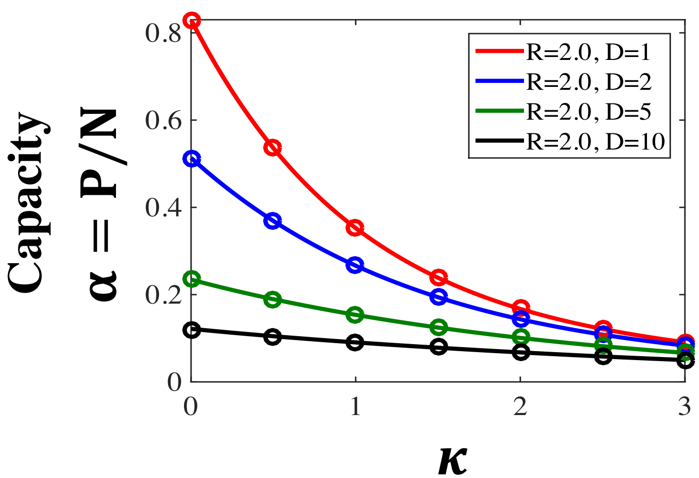
Replica trick: $\langle \log V \rangle = \lim_{n \to 0} \frac{\langle V^n \rangle - 1}{n}$

Capacity versus Geometry for L2 balls

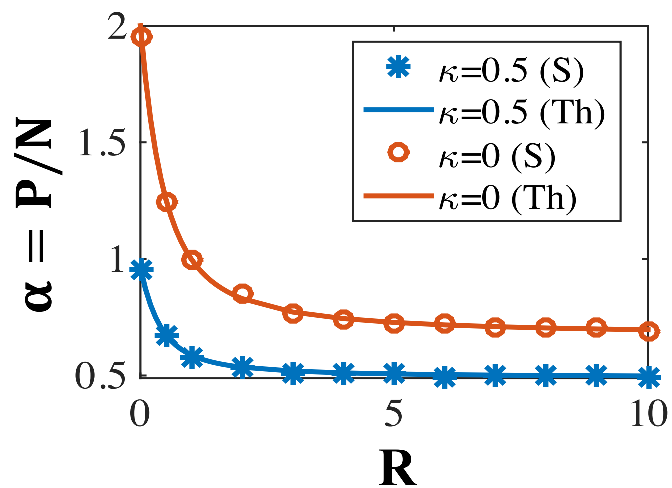
Exact Theoretical Result:

$$\alpha_{ball}{}^{-1}(\kappa, R, D) = \int_0^{\infty} dt\, \chi_D(t) \int_{\kappa-\frac{t}{R}}^{\kappa+Rt} Dt_0 \frac{(Rt+\kappa-t_0)^2}{R^2+1} + \int_0^{\infty} dt\, \chi_D(t) \int_{-\infty}^{\kappa-Rt} Dt_0 \left[(\kappa-t_0)^2 + t^2\right]$$

($\chi_D(t)$ is D-dimensional Chi distribution)



Manifold dimension ↓
Manifold capacity ↑.

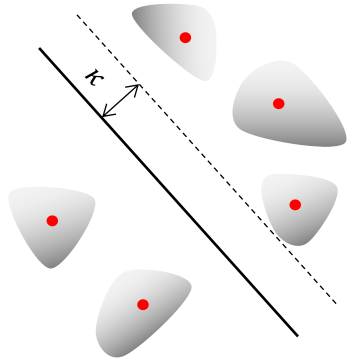


Manifold radius ↓
Manifold capacity ↑.

$\alpha = \frac{P\ (No. of\ Manifolds)}{N\ (Ambient\ Dimension)}$
κ = margin
R = radius of a ball
D = dimension of a ball
Line: Theory
Markers: Simulation

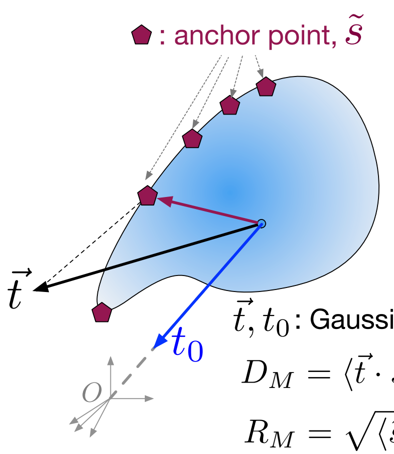
SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. "Linear readout of object manifolds." *Physical Review E* 93.6 (2016): 060301.

Capacity of **object manifolds**



General manifold capacity in high dimension:

$$\boxed{\boldsymbol{\alpha_{manifold}(\kappa) = \alpha_{ball}(\kappa, R_M, D_M)}}$$



$\vec{t}, t_0$: Gaussian vector

$D_M = \langle \vec{t} \cdot \hat{s} \rangle_{\vec{t}, t_0}$

$R_M = \sqrt{\langle \tilde{s}^2 \rangle_{\vec{t}, t_0}}$

$\boldsymbol{\alpha_{manifold}}$: general manifold capacity

$\boldsymbol{\alpha_{ball}}$: capacity for L2 balls of radius R,D

$\boldsymbol{\kappa}$: margin

$\boldsymbol{R_M}$: Effective Manifold Radius

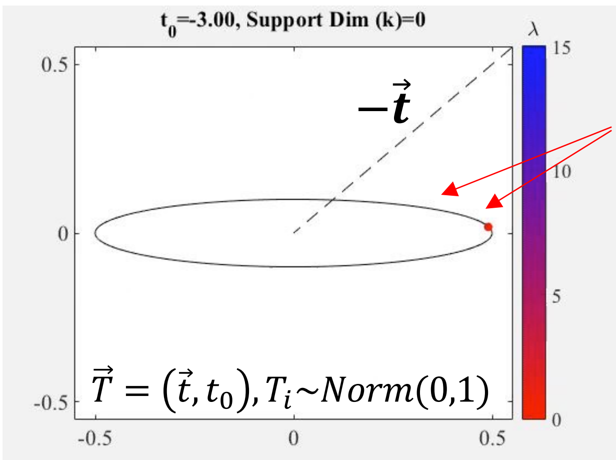
$\boldsymbol{D_M}$: Effective Manifold Dimension

SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. "Classification and Geometry of General Perceptual Manifolds." *Physical Review X* (2018)

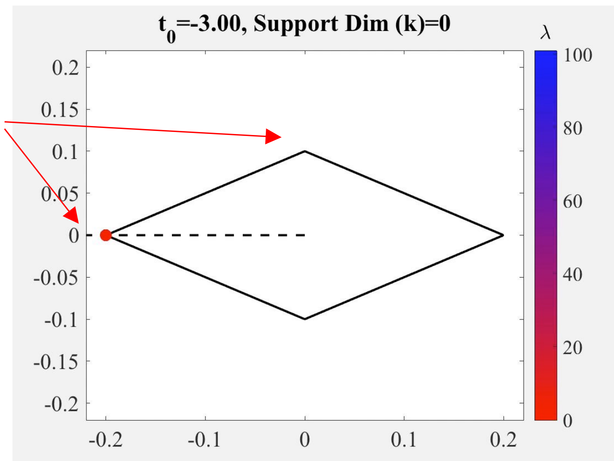
Size of General Manifolds :

Effective Radius (R_M), **Effective Dimension** (D_M)

- **Anchor Points:** $\tilde{s}(\vec{T})$, representative points for linear separation



Anchors $\tilde{s}(\vec{T})$



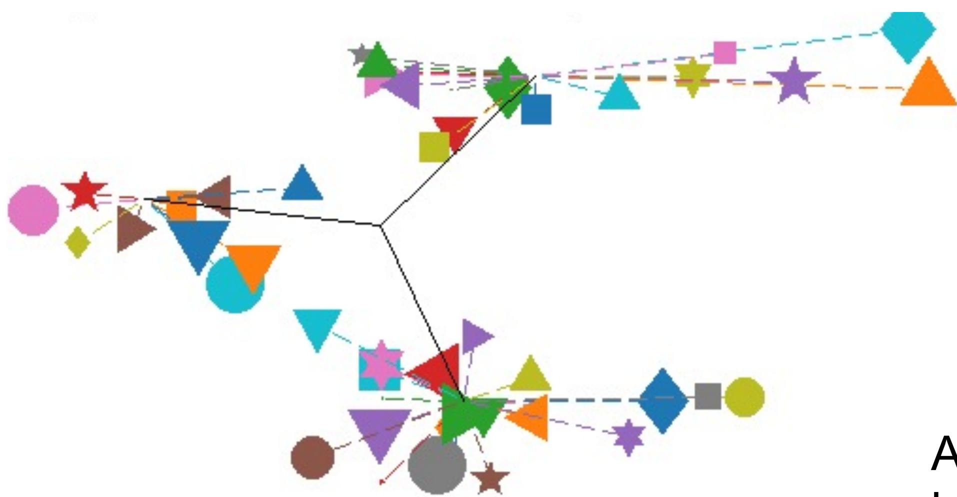
- **Effective Radius:** $R_M^2 = \langle \left| \tilde{s}(\vec{t}) \right|^2 \rangle_{\vec{t}}$
- **Effective Dimension:** $D_M = \dfrac{\langle \left| \vec{t} \cdot \tilde{s}(\vec{t}) \right| \rangle_{\vec{t}}^2}{R_M^2}$

For High Dimension ($D_M \gg 1$)

$$\alpha_{manifold} = \alpha_{ball}(\kappa, R_M, D_M) = \alpha_{point}\left(\kappa + R_M\sqrt{D_M}\right)$$

"Classification and Geometry of General Perceptual Manifolds", Chung, Lee, Sompolinsky (2018), Physical Review X

Correlations between Manifolds' Positions



Center Correlation:

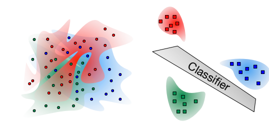
$$\langle \hat{x}_{0,i} \cdot \hat{x}_{0,j} \rangle_{i<j}$$

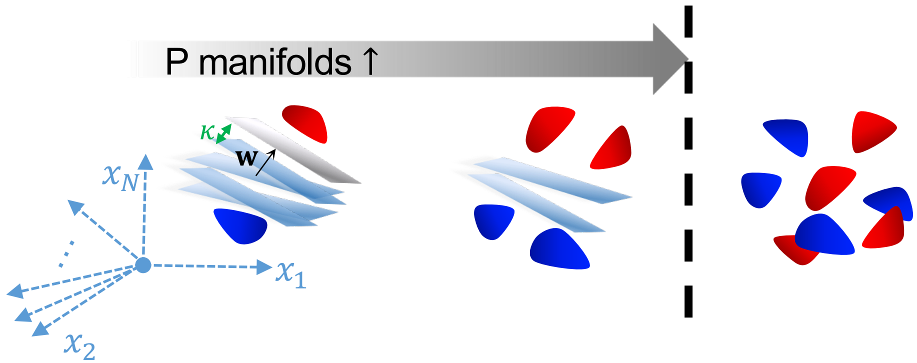
Average of signed pairwise overlap between manifold centers

- **Correlations** between manifold centers tend to **reduce** capacity
- Correlations are often low rank
- Readout weight vector projects data to the null space of centers

Cohen*, Chung*, Lee, and Sompolinsky Nature Comms (2020)

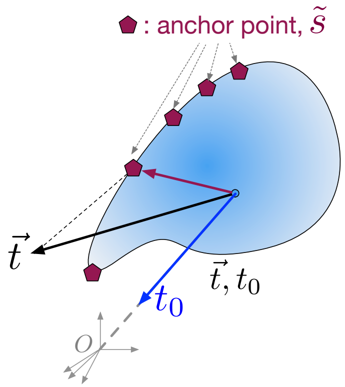
Theory connects the amount of object information (manifold capacity) with object manifolds' geometry





➢ Manifold Capacity ↓ ↑

$\alpha_{manifold}$: max #(Manifolds)/#(Features)
s.t. manifold dichotomies are separable



$$D_M = \langle \vec{t} \cdot \hat{s} \rangle_{\vec{t}, t_0}$$

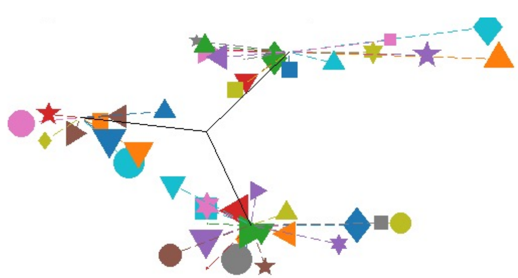
$$R_M = \sqrt{\langle \tilde{s}^2 \rangle_{\vec{t}, t_0}}$$

➢ Manifold Dimension, D_M ↑ ↓

➢ Manifold Radius, R_M ↑ ↓

captures each manifold's dimension, manifold's size related to capacity

$$\alpha_{manifold} = \alpha_{ball}(R_M, D_M)$$



➢ Center Correlation ↑ ↓

average of signed pairwise overlap between manifold centers

Chung, Lee, and Sompolinsky PRE (2016)
Chung, Lee, and Sompolinsky PRX (2018)
Cohen*, Chung*, Lee, and Sompolinsky Nature Comms (2020)

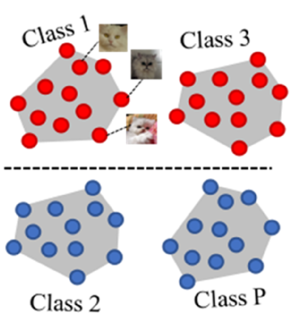
Characterizing neural population for invariant object recognition through geometry & capacity

- Task-dependent metric for object manifolds in neural population

Input:
Manifold-like
Point Cloud Data

Theory-backed
Geometric Analysis

Output:
High Level: Manifold **Capacity**
Low Level: Manifold **Dimension**,
Manifold **Radius, Correlation**



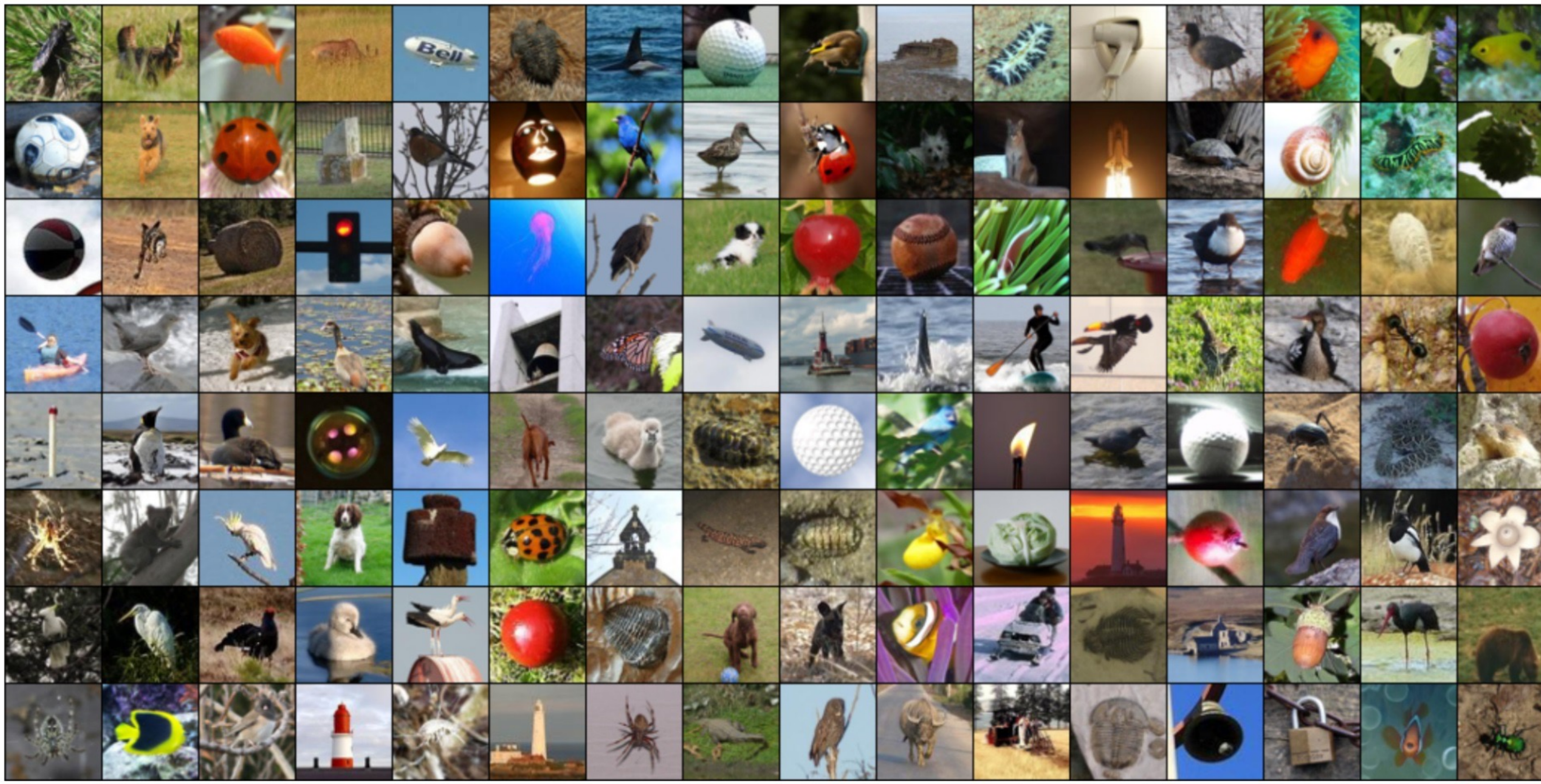
Classification and Geometry of General Perceptual Manifolds

SueYeon Chung,[1,2,6] Daniel D. Lee,[2,3] and Haim Sompolinsky[2,4,5]
[1]*Program in Applied Physics, School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA*
[3]*School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[4]*Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel*
[5]*Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem 91904, Israel*
[6]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

Outline

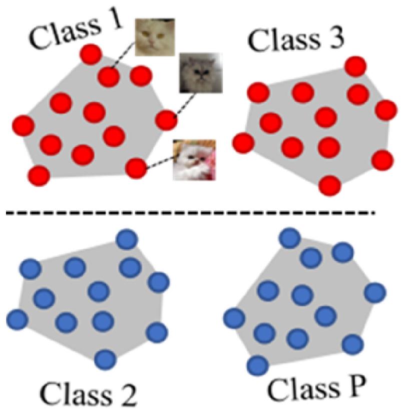
1. Introduction
2. Theory of Linear Classification of Object Manifolds
3. **Object Manifolds in Visual Hierarchy**
4. Object Manifolds in Auditory Hierarchy
5. Object Manifolds in Language Hierarchy
6. Understanding Generalization Dynamics using Object Manifolds

ImageNet

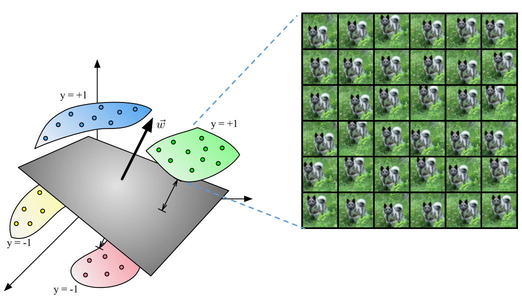


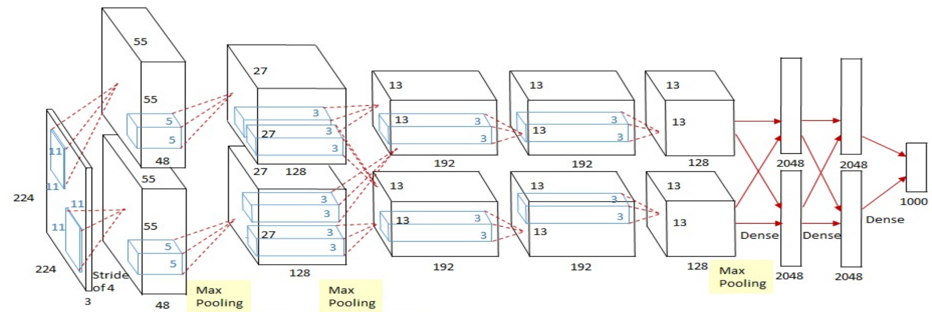
Visual Deep Convolutional Networks

Point-Cloud Manifolds



Affine-transformation manifolds

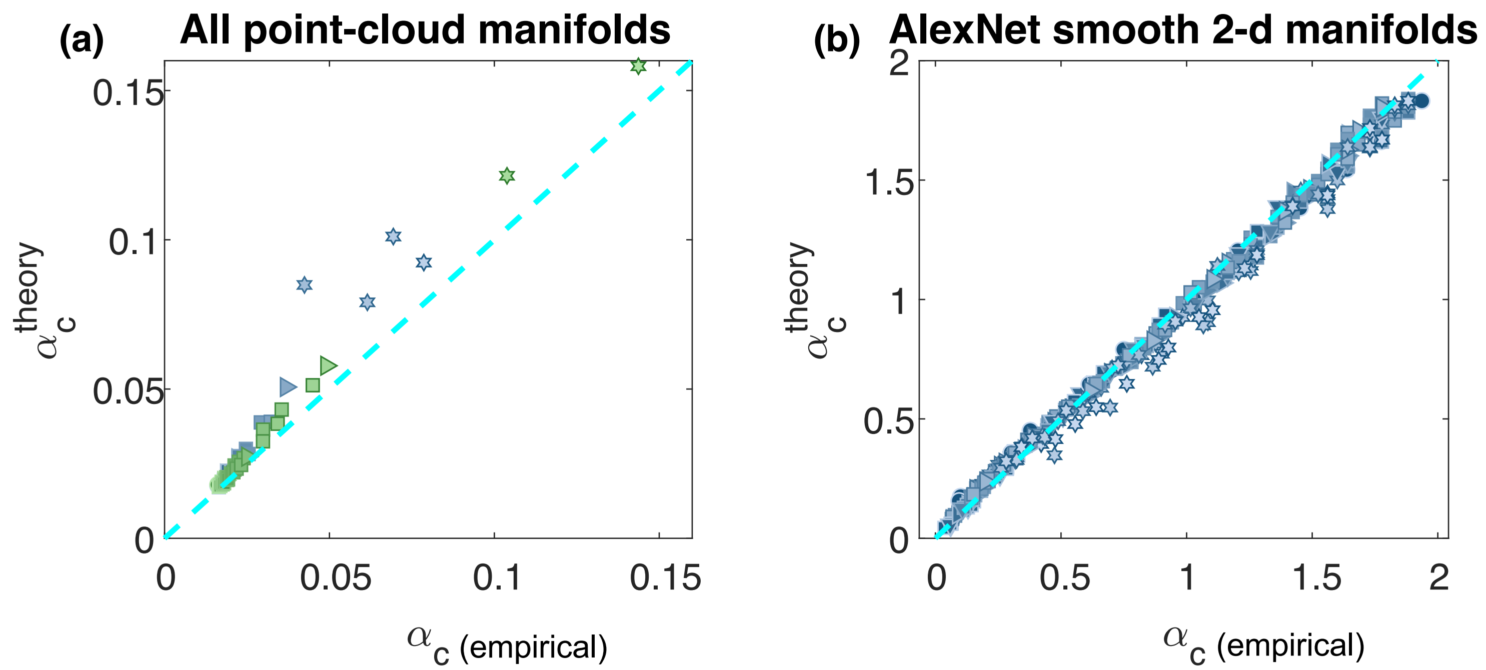




AlexNet

Cohen*, Chung*, Lee, and Sompolinsky. Nature Communications (2020)

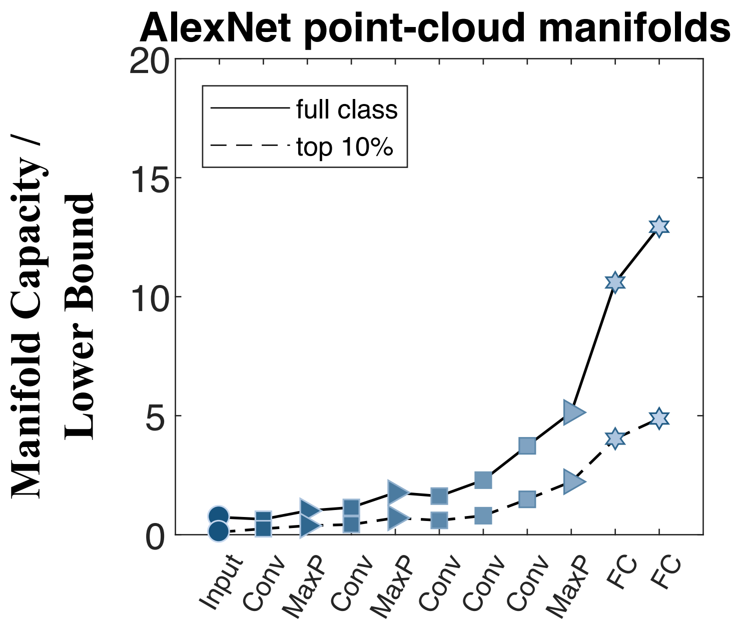
Measured manifold capacity is predicted by the theoretical manifold capacity (using geometry).

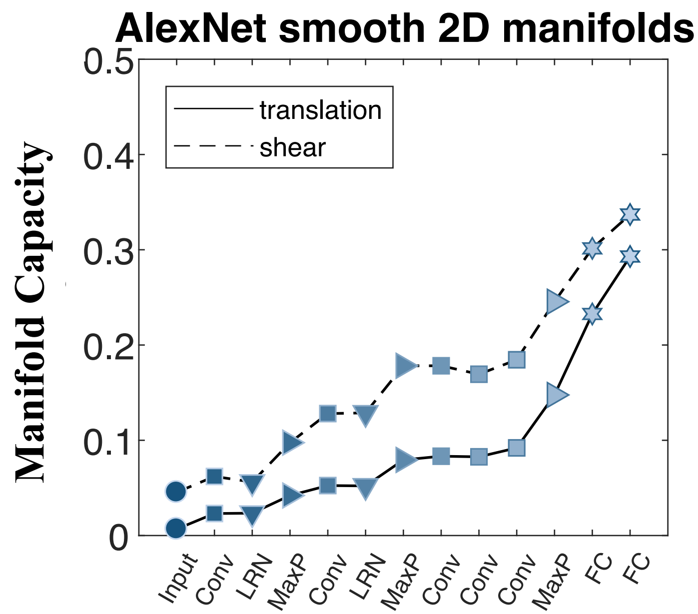


(theoretical) α_{theory} = theoretical prediction using D_M, R_M, correlations

(empirical) α_c = **<u>critical fraction of no. of manifolds / feature dimension</u>**
<u>s.t. majority of manifold dichotomies are linearly separable</u>

Manifold Capacity improves across deep network layers

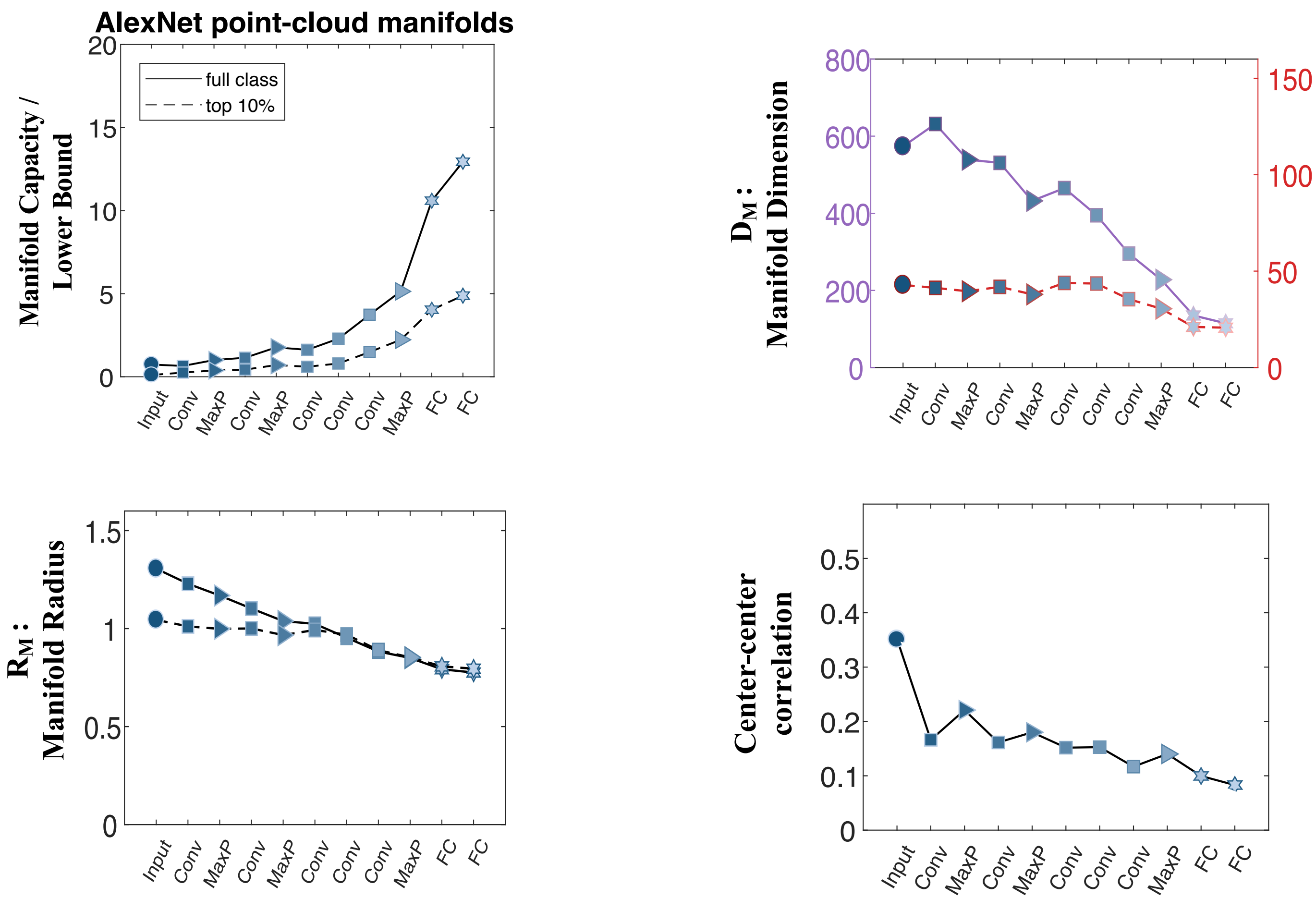




Cohen*, Chung*, Lee, and Sompolinsky. Nature Comms (2020)

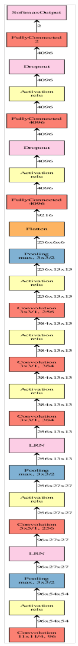
Manifold Capacity improves across layers

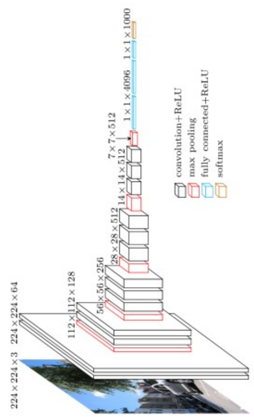
Due to reduced Dimension, Radius, Correlations



Cohen*, Chung*, Lee, and Sompolinsky. Nature Comms (2020)

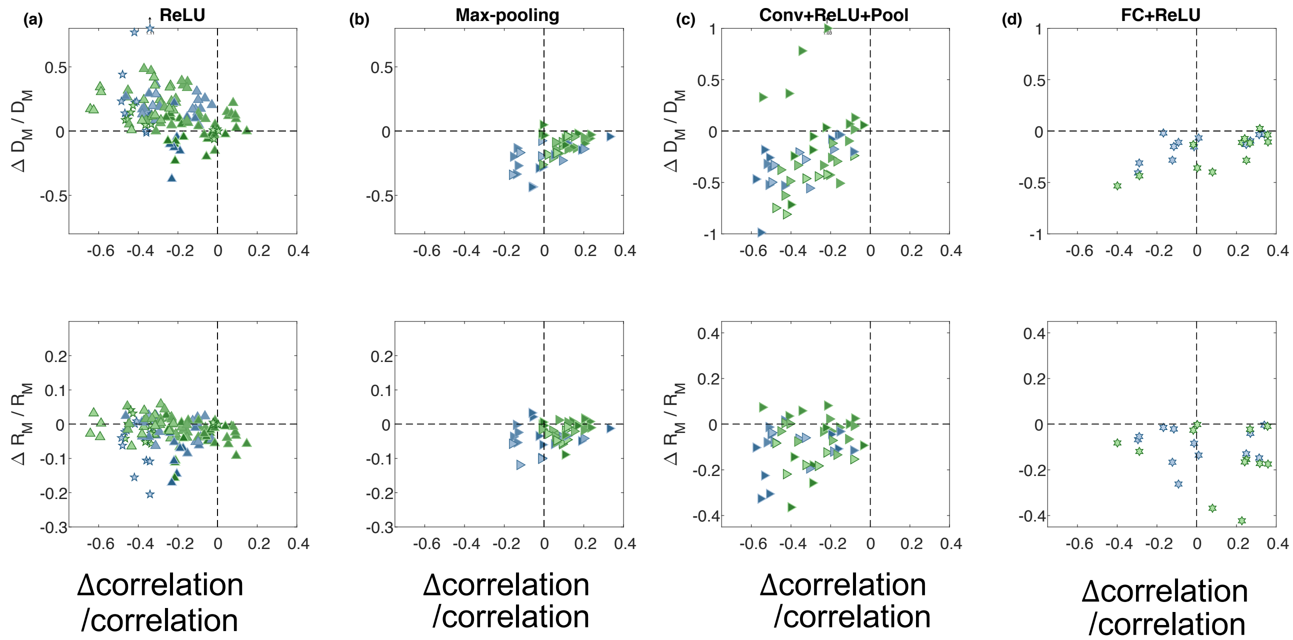
Each layer's role on untangling neural manifolds





AlexNet **VGG16**

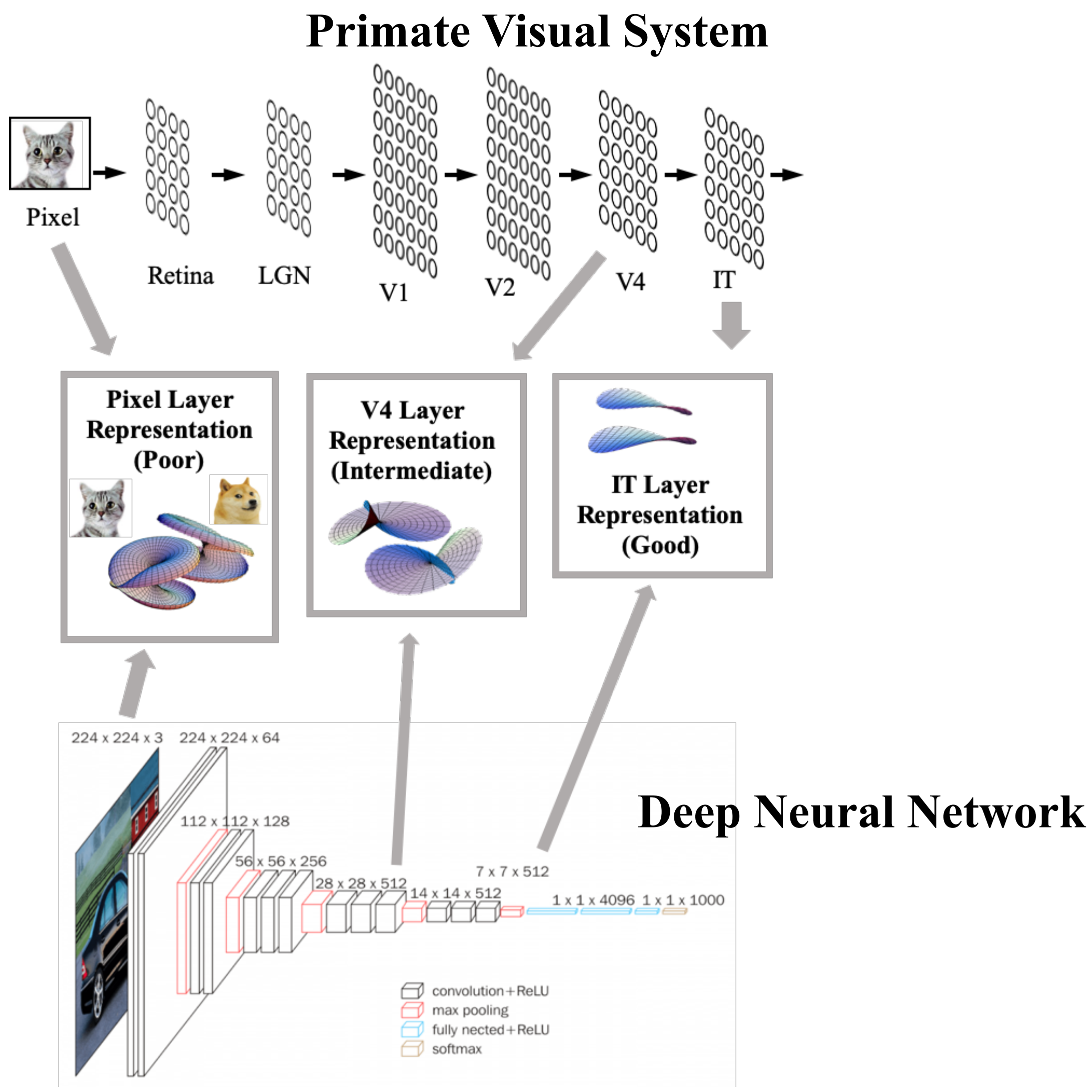
Measured before/after each computational unit (i.e. layer)



- **ReLu: decorrelates**
- **Max-pooling: reduces R, D**
- **Conv-ReLu-Pool: decorrelates & reduces R, D**
- **FC-ReLu: reduces R, D**

-> Improves manifold capacity

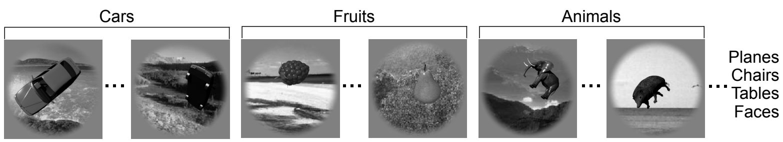
How do neural manifolds in macaque ventral stream compare with deep neural networks?



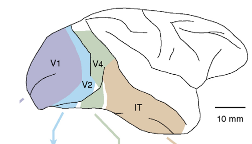
with Jim DiCarlo (MIT), Joel Dapello (MIT), Haim Sompolinsky (HUJI)

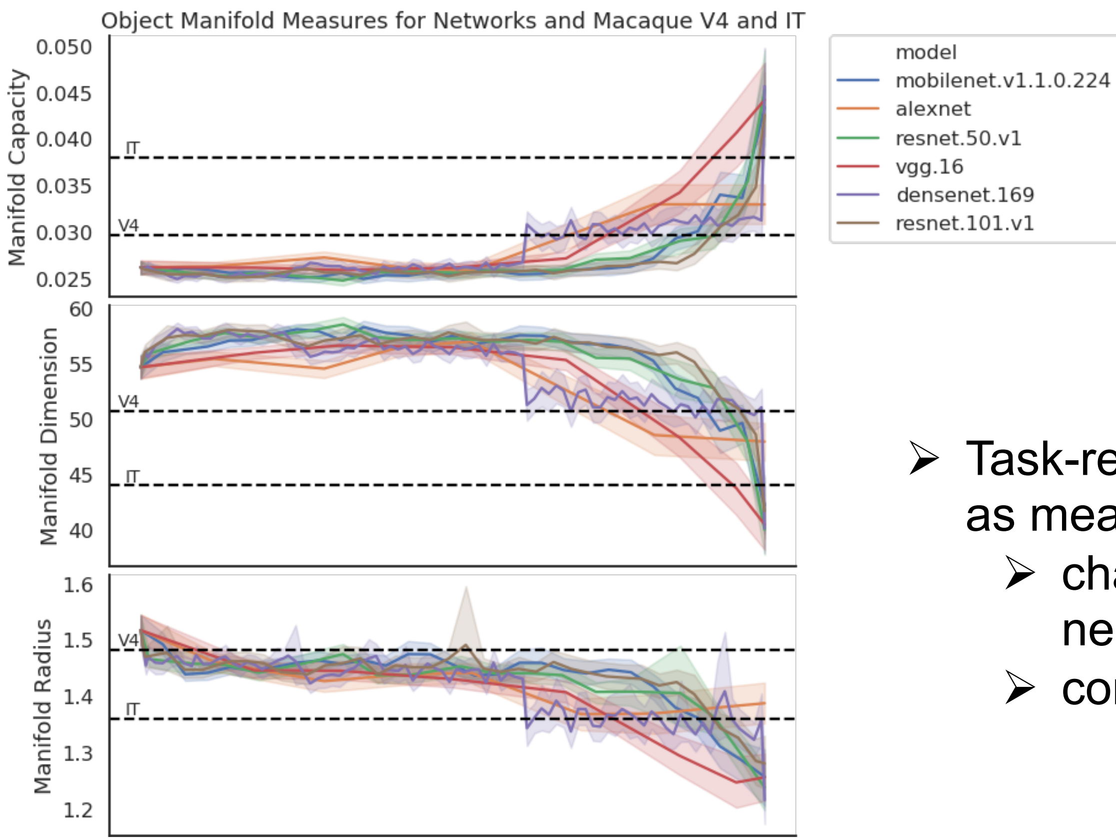
Neural Manifolds in Macaque Ventral Stream (vs. in DCNN)

- Dataset:
- 64 3D object models (varied in rendering, position, size) in random backgrounds



(stimuli from Majaj, DiCarlo et al, 2015)





- Task-relevant geometry can be used as measures for:
  - characterizing high-dimensional neural population
  - comparing representations

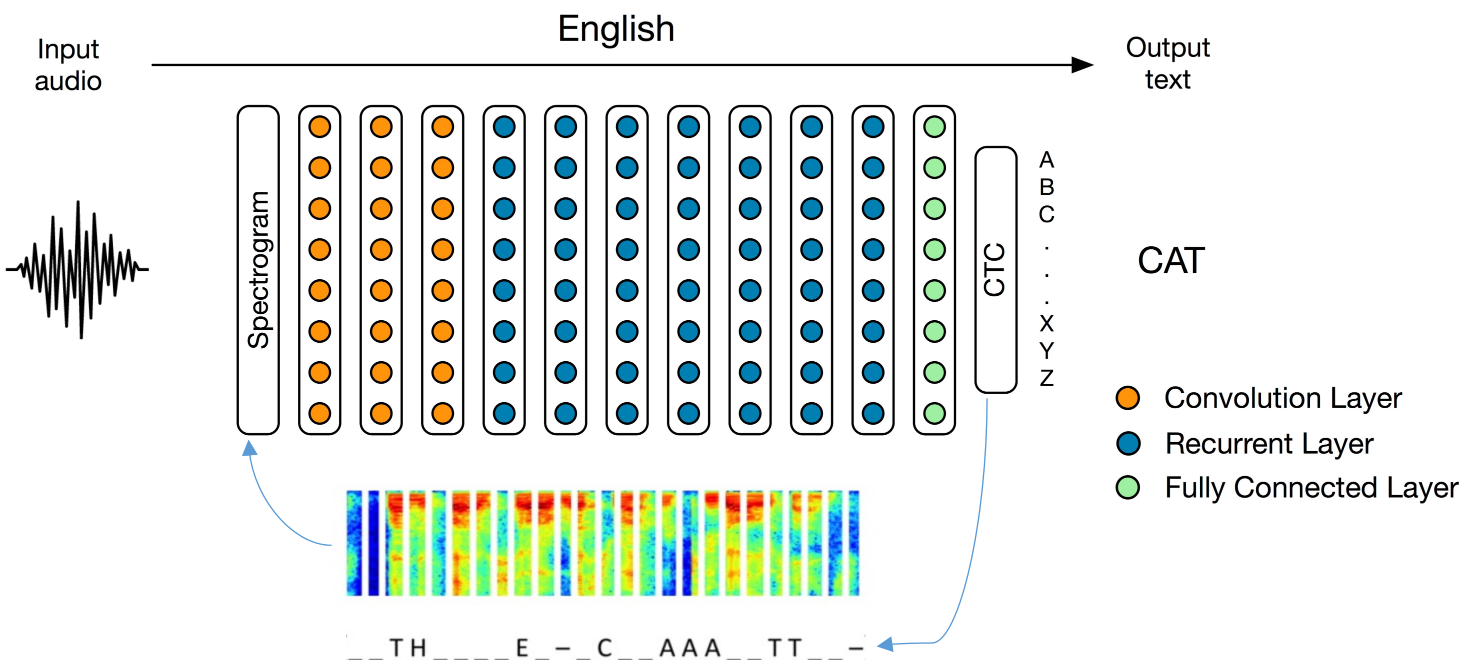
Outline

1. Introduction
2. Theory of Linear Classification of Object Manifolds
3. Object Manifolds in Visual Hierarchy
4. **Object Manifolds in Auditory Hierarchy**
5. Object Manifolds in Language Hierarchy
6. Understanding Generalization Dynamics using Object Manifolds

Cory Stephenson, Jenelle Feather, Suchismita Padhy, Oguz Elibol, Hanlin Tang, Josh McDermott, SueYeon Chung**.
Untangling in Invariant Speech Recognition. **NeurIPS 2019**

Deepspeech 2: Speech-to-text model

Training data: 1000 hours of read English speech

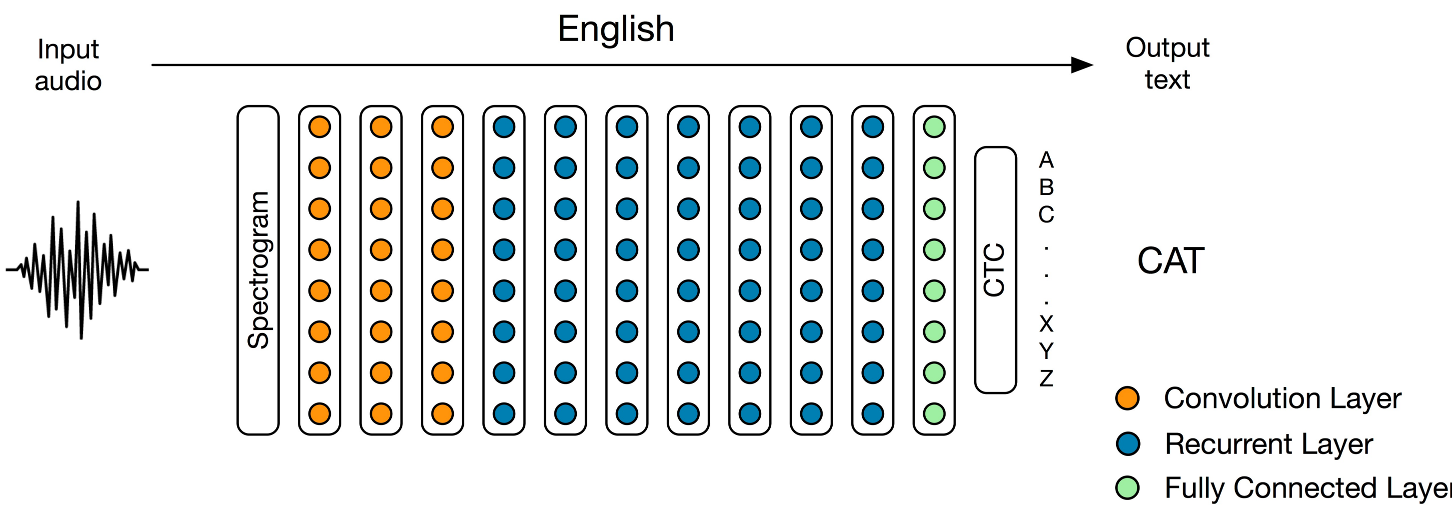


Output: THE–CAT–

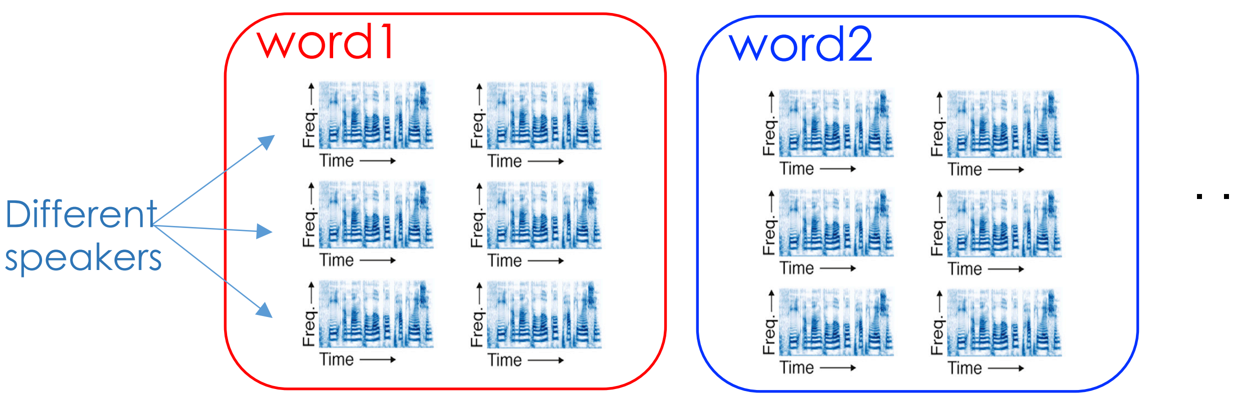
Baidu Research, NIPS (2015)
https://svail.github.io/

Deepspeech 2: Speech-to-text model

Training data: 1000 hours of read English speech

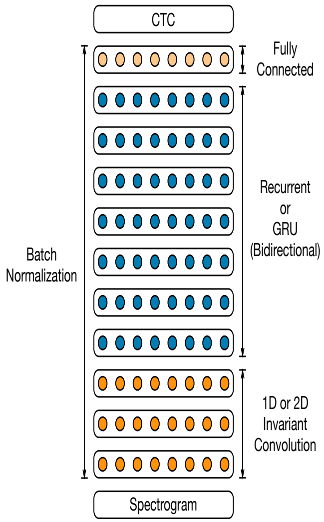


Q: Do "**word**" **manifolds** arise in DCNN and DRNN models?

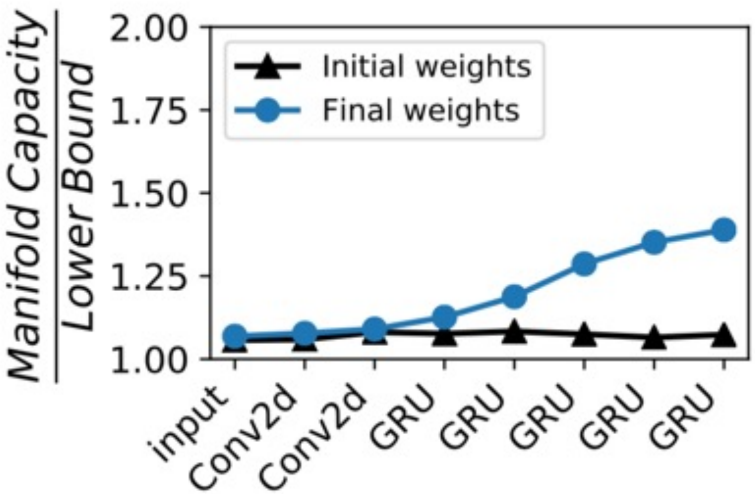


Baidu Research, NIPS (2015)
https://svail.github.io/

Word Manifolds’ capacity improves across layers



DRNN
(Speech to
Text)

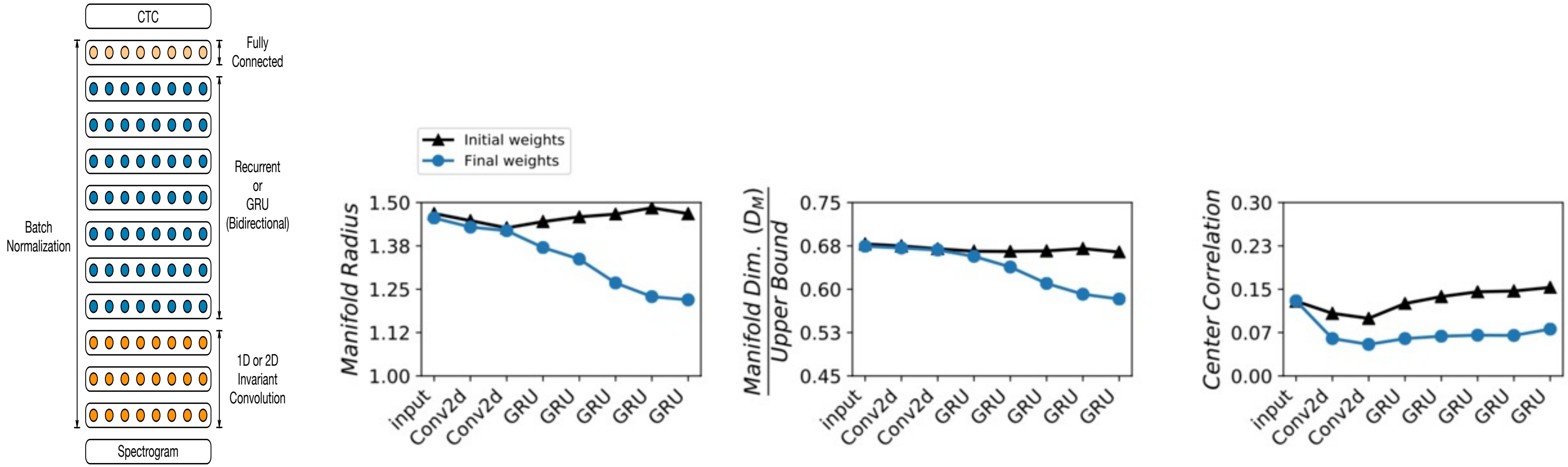


Test Data: 50 words, spoken by 50 speakers

- ➢ Untangling seen in visual systems also occurs in auditory deep networks
- ➢ Capacity is flat at initial weights, and is increasing across layers after the training

Word Manifolds' capacity improves across layers

Due to reduction in manifold dimension, radius, correlations



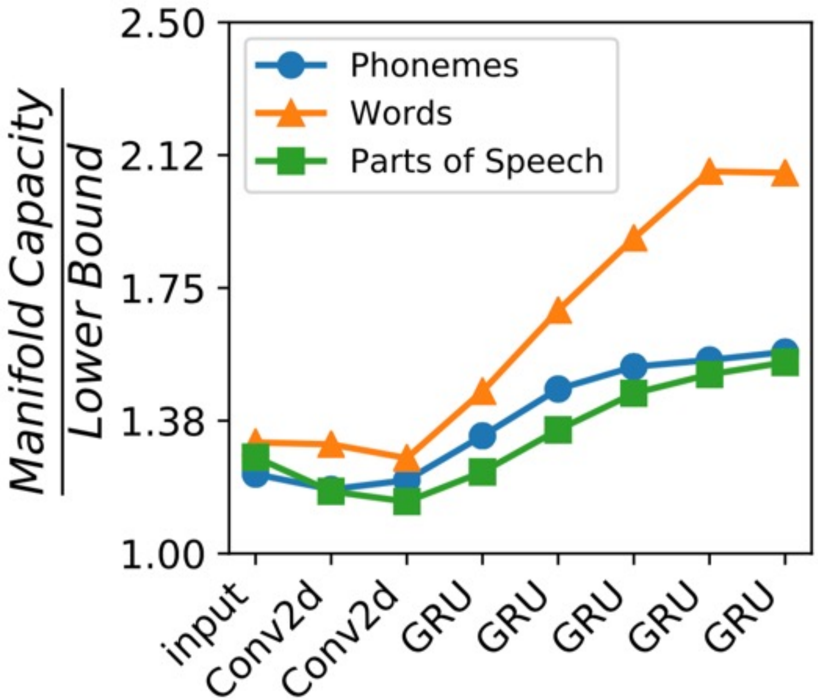
DRNN
(Speech to
Text)

- ➢ Manifold Dimension, Radius, and Correlations all decrease across layers after training (similar to Visual deep networks)

Cory Stephenson, Jenelle Feather, Suchismita Padhy, Oguz Elibol, Hanlin Tang, Josh McDermott, SueYeon Chung**.
Untangling in Invariant Speech Recognition. **NeurIPS 2019**

Untangling speech objects in multiple scales in DRNN

- Speech objects in different scales emerge across layers in Deepspeech2
  - Phonemes, Words, and Part-of-Speech (POS) Manifolds



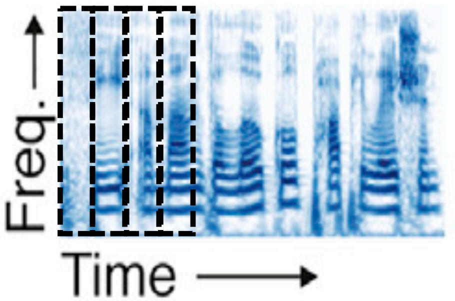
Phonemes: “aa”, “ch”, ”b”,”d”, …

Words: “carry”, “dark”, “every”,…

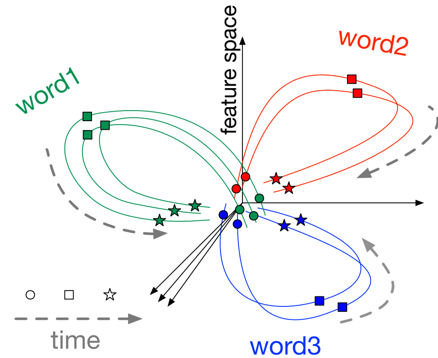
Part of speech: “Noun”, “Verb”, ”Pronoun”, …

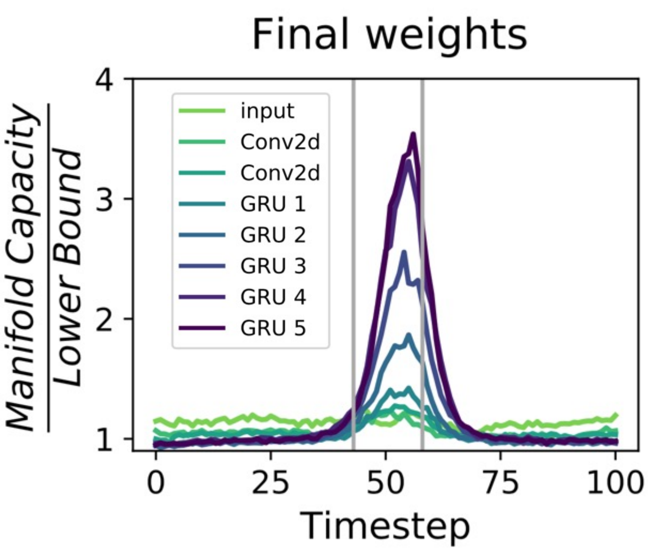
Cory Stephenson, Jenelle Feather, Suchismita Padhy, Oguz Elibol, Hanlin Tang, Josh McDermott, SueYeon Chung**. **Untangling in Invariant Speech Recognition. NeurIPS 2019**

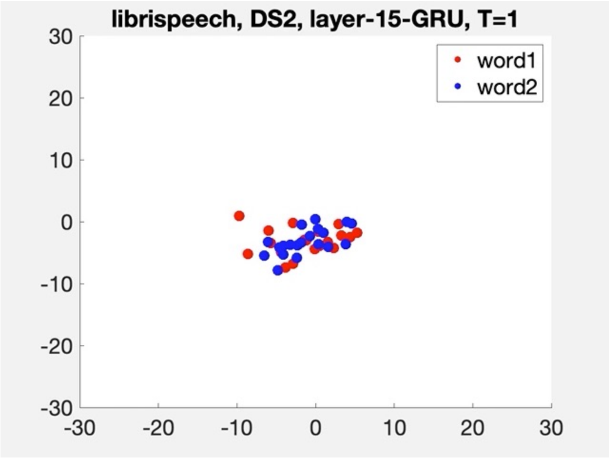
Word Untangling Across Recurrent Timesteps in DRNN



Q: What is the role of recurrent timesteps in untangling?







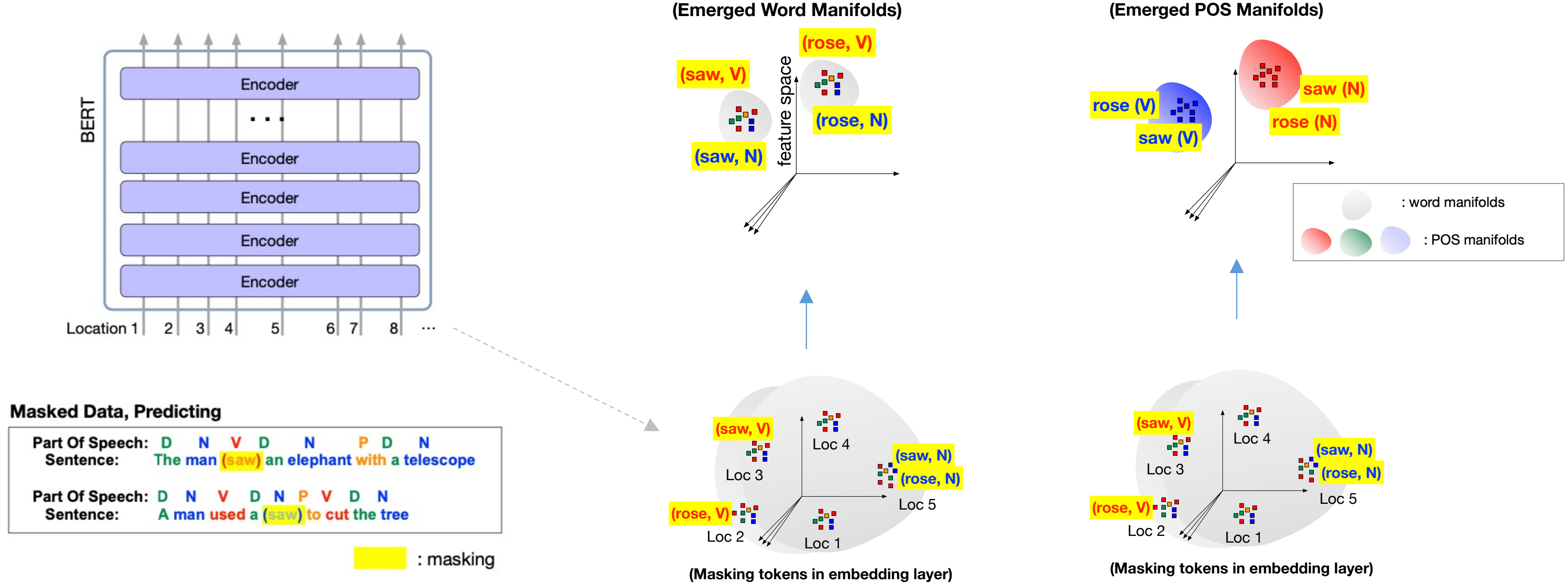
- Recurrent timesteps “untangle” word objects

Outline

1. Introduction
2. Theory of Linear Classification of Object Manifolds
3. Object Manifolds in Visual Hierarchy
4. Object Manifolds in Auditory Hierarchy
5. **Untangling in Deep Language Representations**
6. Understanding Generalization Dynamics using Object Manifolds

Jonathan Mamou*, Hang Le*, Miguel Del Rio, Cory Stephenson, Hanlin Tang, Yoon Kim, SueYeon Chung
Emergence of Separable Manifolds in Deep Language Representations. **(ICML 2020)**

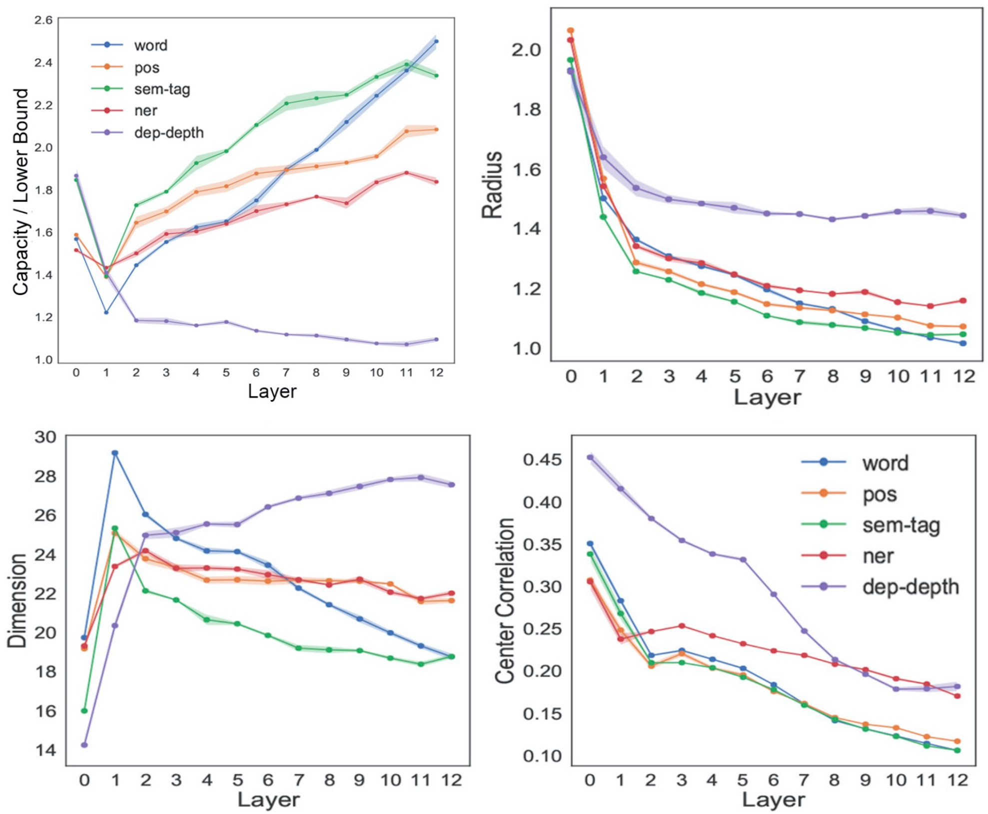
BERT: Bidirectional Encoder Representations from Transformers



Q: do language manifolds emerge across layers of BERT?

Language object class manifolds in layers in BERT

- Defined on masked tokens
- Manifolds defined with Word, POS, Semantic Tag, Named Entity Recognition (NER) improve in capacity across layers.
- Exception: dependency depth
- Improved capacity is due to reduction in radius, dimension, center correlations of manifolds



Jonathan Mamou*, Hang Le*, Miguel Del Rio, Cory Stephenson, Hanlin Tang, Yoon Kim, SueYeon Chung
Emergence of Separable Manifolds in Deep Language Representations. **(ICML 2020)**

Outline

1. Introduction
2. Theory of Linear Classification of Object Manifolds
3. Object Manifolds in Visual Hierarchy
4. Object Manifolds in Auditory Hierarchy
5. Untangling in Deep Language Representations
6. **Generalization vs. Memorization Manifolds in DNNs**

Cory Stephenson, Suchi Padhy, Abhi Ganesh, Yue Hui, Hanlin Tang, SueYeon Chung
On the geometry of generalization and memorization in deep neural networks. **(ICLR 2021)**

Probing the structure of generalization vs. memorization

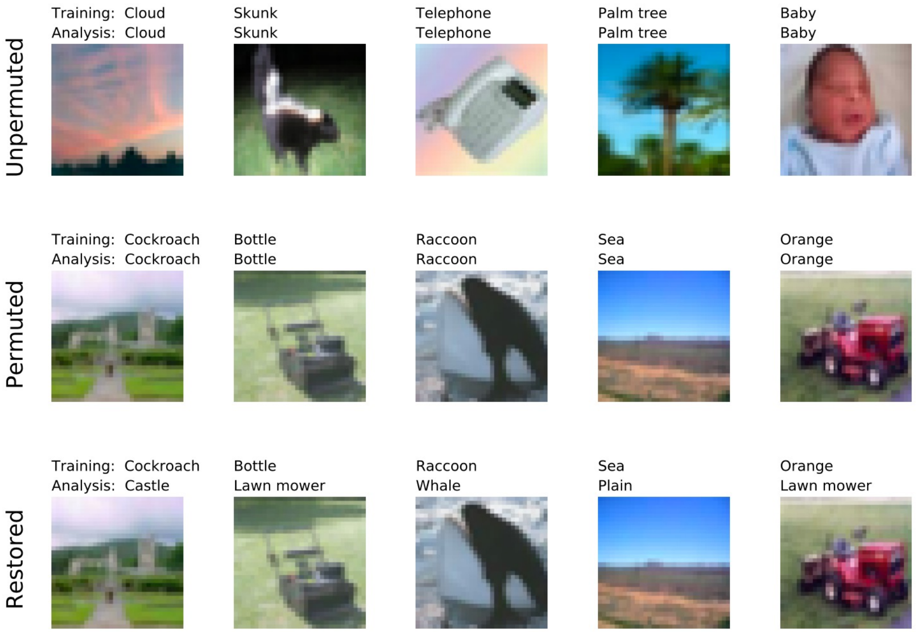
Memorization: defined as ‘behaviors exhibited by DNNs trained on noise/random labels’. (Arpit and Bengio et al, 2017)

Experiment: Train DNNs with Images where 50% of the labels are shuffled

Analysis: define object manifolds for:

(1) unpermuted labels, (2) permuted labels

(3) restored labels (while trained with permuted labels)

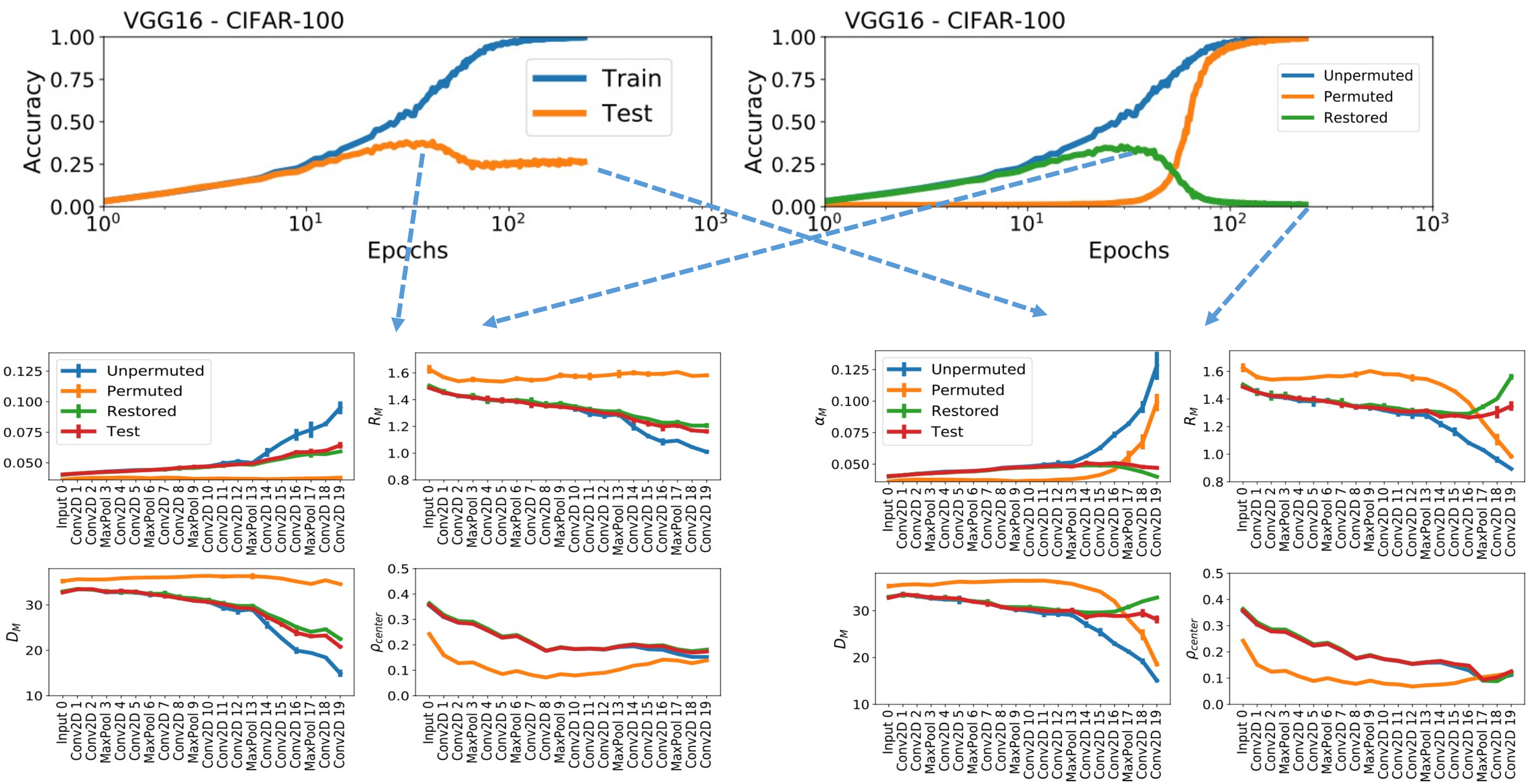


(CIFAR100)

Cory Stephenson, Suchi Padhy, Abhi Ganesh, Yue Hui, Hanlin Tang, SueYeon Chung
On the geometry of generalization and memorization in deep neural networks. **(ICLR 2021)**

Probing the structure of generalization vs. memorization

- Decrease in test performance coincide with decrease in accuracy for the ‘restored’ labels & increase in accuracy for ‘permuted’ labels
- ‘Unpermuted’ (easy) examples learn first, ‘permuted’ (hard) examples learn later



- ‘permuted’ examples haven’t been learned
- ‘restored’ manifolds similar to ‘test’ manifolds

- Most memorization occurs in the final layers
- Early layers ignore the effect of memorization

Cory Stephenson, Suchi Padhy, Abhi Ganesh, Yue Hui, Hanlin Tang, SueYeon Chung
On the geometry of generalization and memorization in deep neural networks. (ICLR 2021)

Summary

- Generalized statistical mechanical theory of linear classification of points to that of general randomly oriented manifolds.

- **Capacity of category manifolds** measures invariant object information in features

- Manifold capacity is **predicted by** the **effective size ($\mathbf{R_M}$), dimensionality ($\mathbf{D_M}$),** and **correlations** of the perceptual manifolds in neural representation

- Analysed **neural manifold geometry in deep neural networks & neural data**

  - Manifold properties change in the direction to improve capacity across visual hierarchy in **visual deep networks** and **macaque visual cortex** (reduction in manifold dimension, radius and center-center correlations)

  - **Untangling phenomenon** found in vision seems to also happen in **speech and language processing deep networks**, though not explicitly trained

  - **Structure** of features relevant for **generalization vs. memorization** can be analysed geometrically.

    - Geometry reveals that **most of the memorization occurs in the final layers, and during the final epochs**.

- **DNNs** are **only a testbed**, originally designed for **neural data**

- Many more applications: olfaction, learning dynamics, motor motifs…

Acknowledgement

- **Haim Sompolinsky** (Harvard, Hebrew Univ.)
- **Daniel D. Lee** (Cornell Tech, Samsung AI)
- **James DiCarlo** (MIT)
- **Josh McDermott** (MIT)
- **Hanlin Tang** (Intel AI)
- **Yoon Kim** (MIT-IBM Watson AI)
- **Larry Abbott** (Columbia University)

Uri Cohen (HUJI)
Cory Stephensen (Intel AI)
Suchismita Padhy(Intel AI)
Joel Dapello (Harvard, MIT)
Jenelle Feather (MIT)
Jonathan Mamou (Intel AI)
Hang Le (MIT)