The Intrinsic Manifolds of Radiological Images and Their Role in Deep Learning

Nick Konz

Mazurowski Lab, Duke University

Duke

TAG-DS

October 6th, 2022

MICCAI2022 Singapore

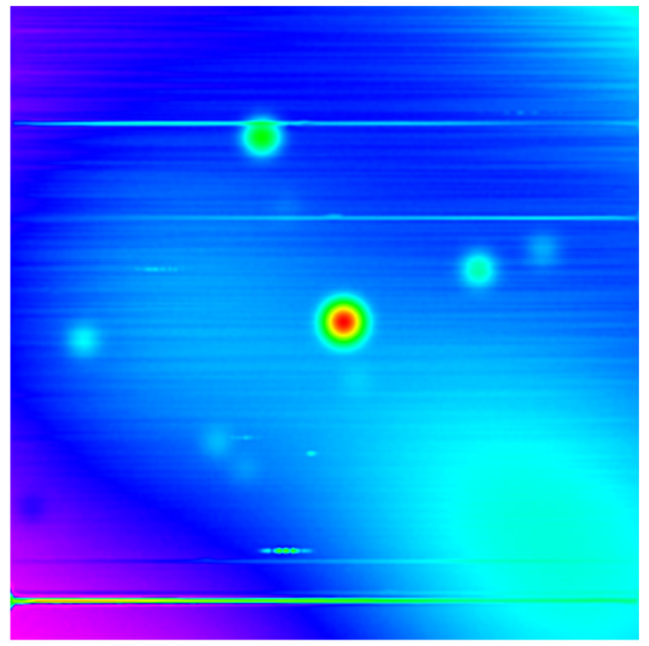
25th International Conference on Medical Image Computing and Computer Assisted Intervention

September 18–22, 2022

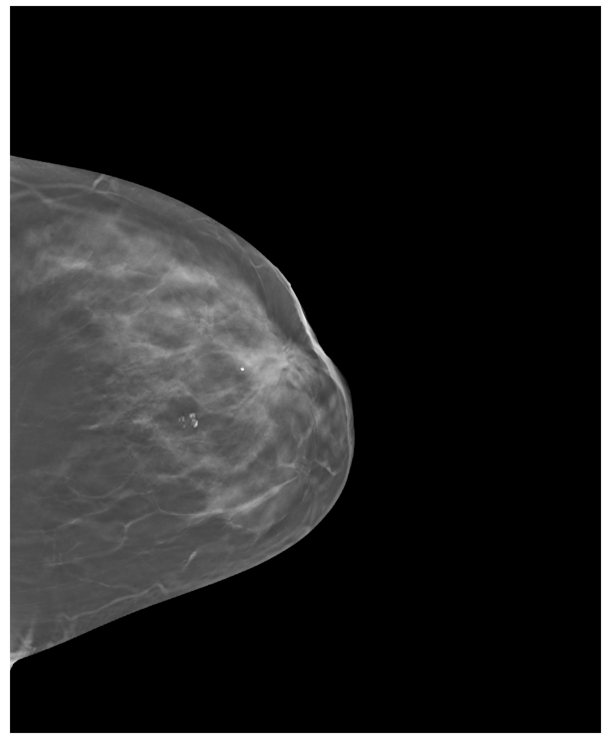
Resorts World Convention Centre Singapore

About me

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Duke



Lightning Review of Deep Learning

Dataset

$D = \{(x, y)\}_{n=1}^{N},$

$x \in \mathbb{R}^d$

$y \in \{0,1\}$

Neural network

$f : \mathbb{R}^d \to \{0,1\}$

$\hat{y} = f(x \mid \theta)$

- $\hat{y}$ ← output prediction
- x ← input data
- θ ← trainable network parameters

Loss function (prediction accuracy metric for f)

$L(\hat{y}, y)$

Higher for more inaccurate $\hat{y}$

Differentiable

Goal of training f: find

$\theta^* = \mathrm{argmin}_\theta \, \mathbb{E}_{(x,y)\in D} \, L(\hat{y}, y)$

How to train a neural net (simplified)

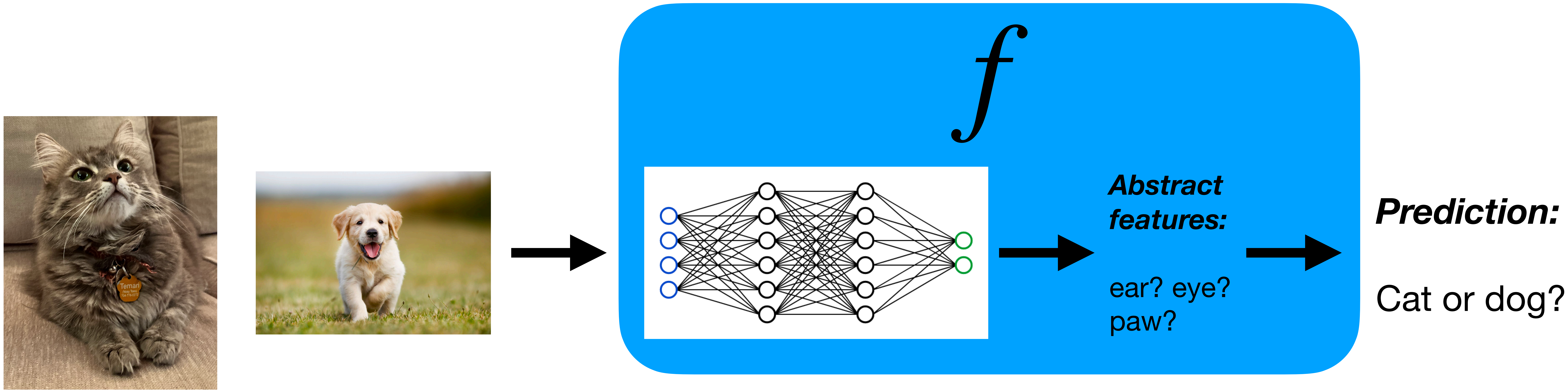
$\mathbb{E}_{(x,y)\in D} L(\hat{y}, y) = \mathbb{E}_{(x,y)\in D} L(f(x\,|\,\theta), y)$ can be minimized w.r.t θ with **gradient descent** as f is differentiable!

For i iterations (epochs):

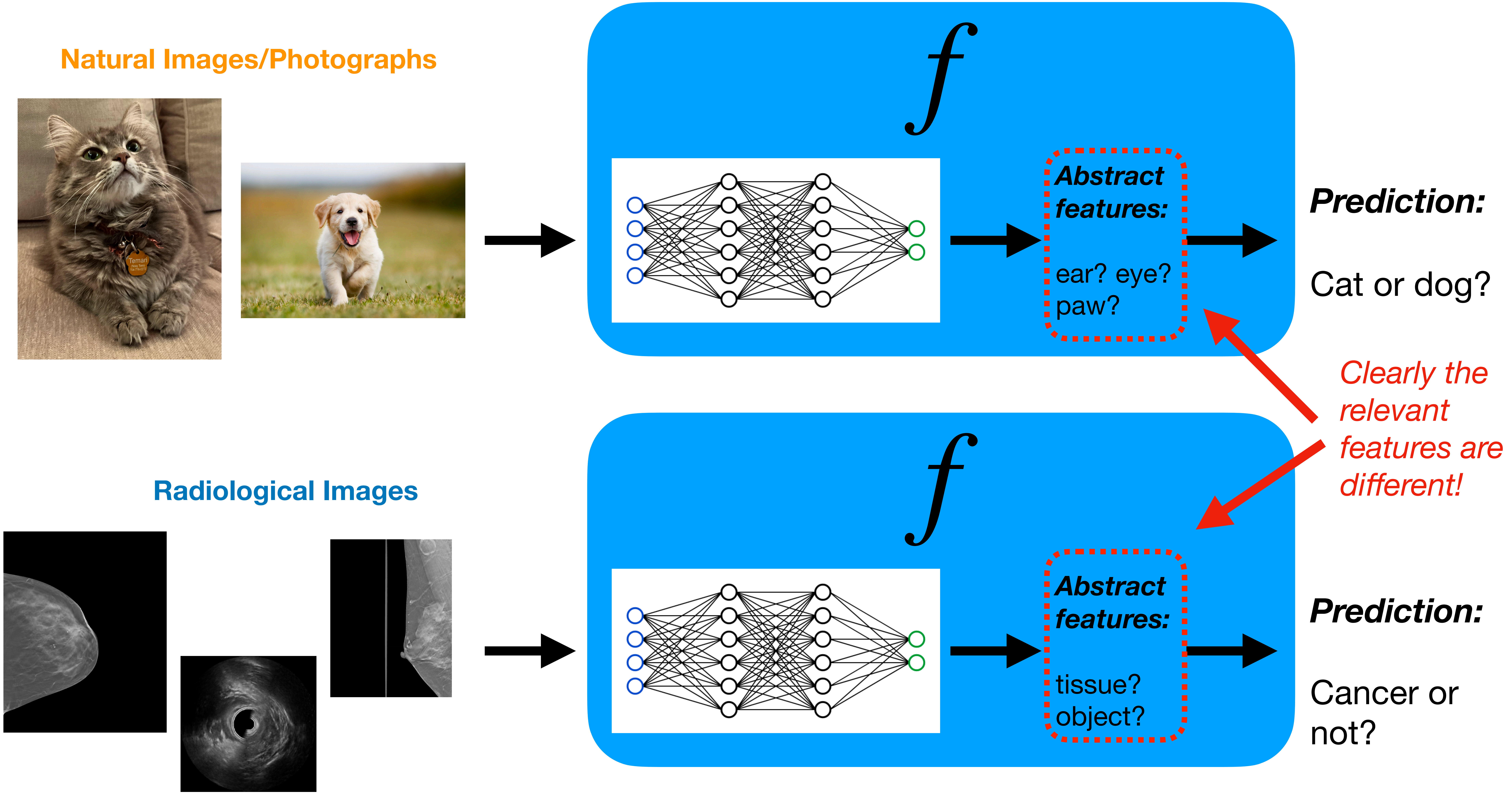
1. $\theta_{i+1} = \theta_i - \alpha \nabla_\theta L(f(x\,|\,\theta), y)$

Neural nets learn to map raw images

to abstract features that are useful for predictions.



Shifting the Data Domain to Radiology



Motivation

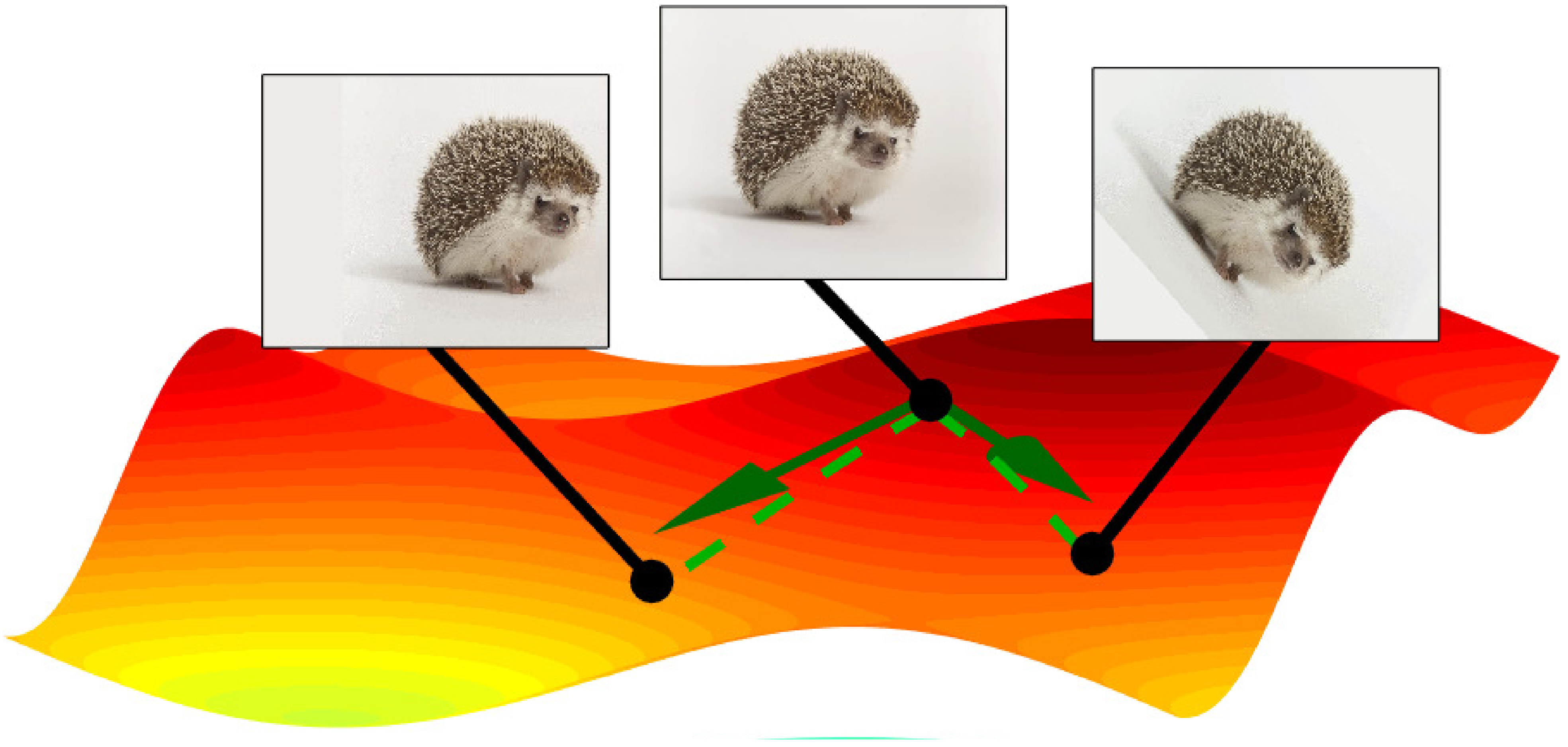
- The shift between these domains is intuitively obvious, but the difference in how networks learn from them is poorly understood

- Is there a way to **quantify** this shift?

- Then, maybe a more formal footing could be found for developing or adopting computer vision methods specifically for radiology (and medical image analysis at large).

The Manifold Hypothesis

- ***Manifold hypothesis (MH):*** Image datasets $\{x\}_{n=1}^{N} \subset \mathbb{R}^d$ lie close to some m-dimensional manifold $M \subset \mathbb{R}^d$ with $m \ll d$.
- In other words, the data can be well-described by m *abstract intrinsic* features
- *Assumption:* neural networks work by learning to map images to this abstract representation
- Therefore, the *intrinsic dimension* (ID) of the manifold m should relate to how networks learn datasets



Buchanan et al., ICLR 2020

Fefferman, C., Mitter, S. & Narayanan, H. Testing the manifold hypothesis. J. Amer. Math. Soc. 29, 983–1049 (2016).

Inspiring Literature: Pope et al. (ICLR 2021)

THE INTRINSIC DIMENSION OF IMAGES AND ITS IMPACT ON LEARNING

- Studied the ID of natural image datasets, found $m \ll d$.
- Found dataset ID to be related to network **generalization ability** (GA) for a fixed training set size.
- Supported the use of ID for generative modeling (using GANs).

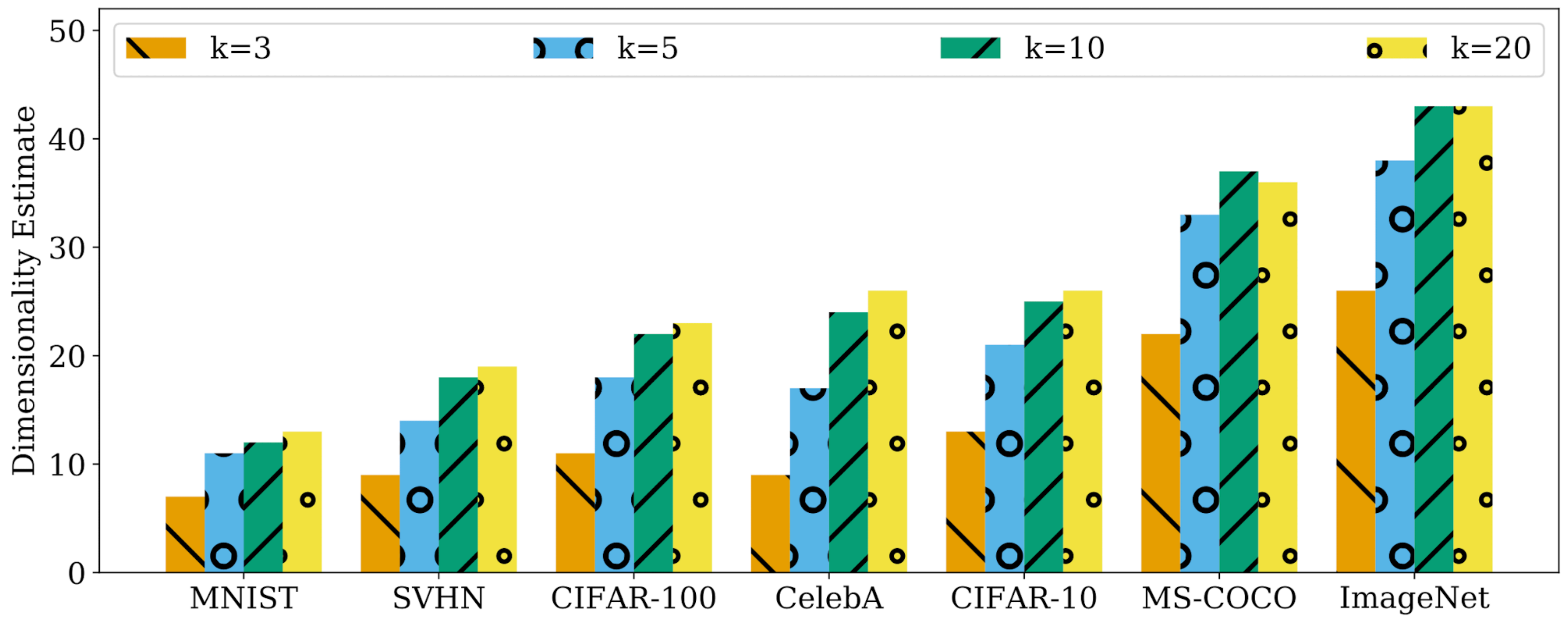


Figure 1: Estimates of the intrinsic dimension of commonly used datasets obtained using the MLE method with $k = 3, 5, 10, 20$ nearest neighbors (left to right). The trends are consistent using different k's.

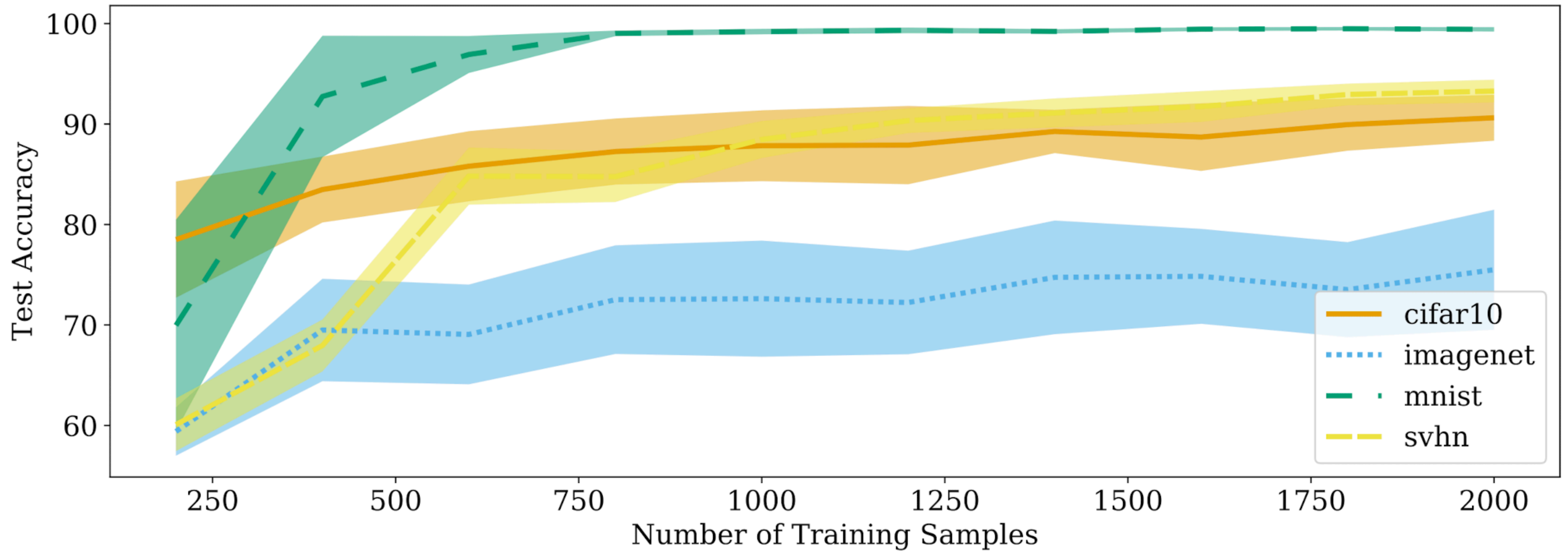


Figure 6: Sample complexity of real datasets. Standard errors are shown $N = 5$ class pairs.

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. ICLR 2021

Inspiring Literature: Ansuini et al. (NeurIPS 2021)

Intrinsic dimension of data representations in deep neural networks

- Studied the ID of *internal network representation* of natural image datasets, also found $m \ll d$.
- Found *representation* ID to be **linearly correlated with network GA** for a fixed training set size.

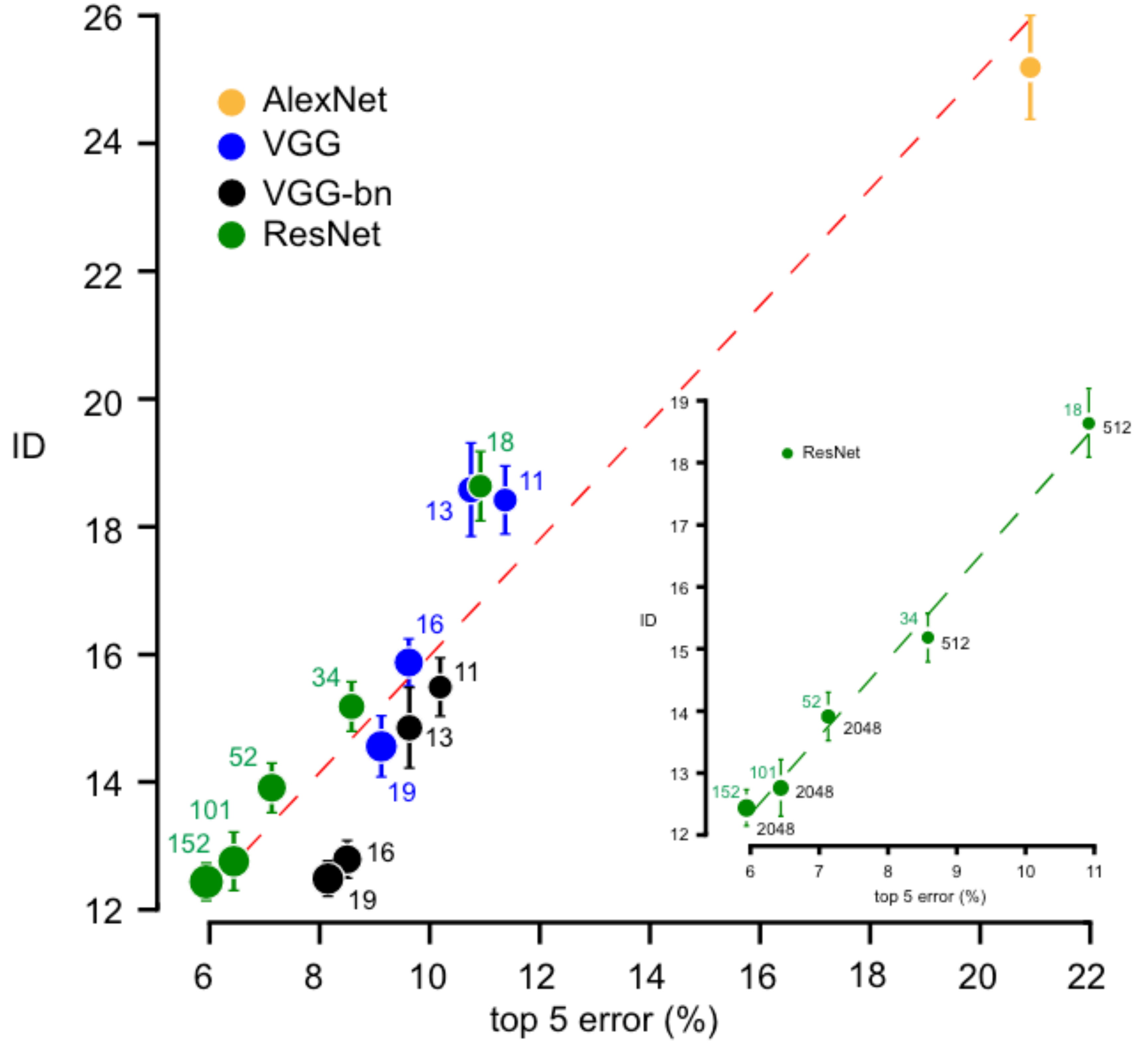


Figure 4: **ID of the last hidden layer predicts performance.** The ID of data representations (training set) predicts the top 5-score performance on the test set. **Inset** Detail for the ResNet class.

Ansuini, A., Laio, A., Macke, J. H. & Zoccolan, D. Intrinsic dimension of data representations in deep neural networks. in NeurIPS 2019.

The goals of this research

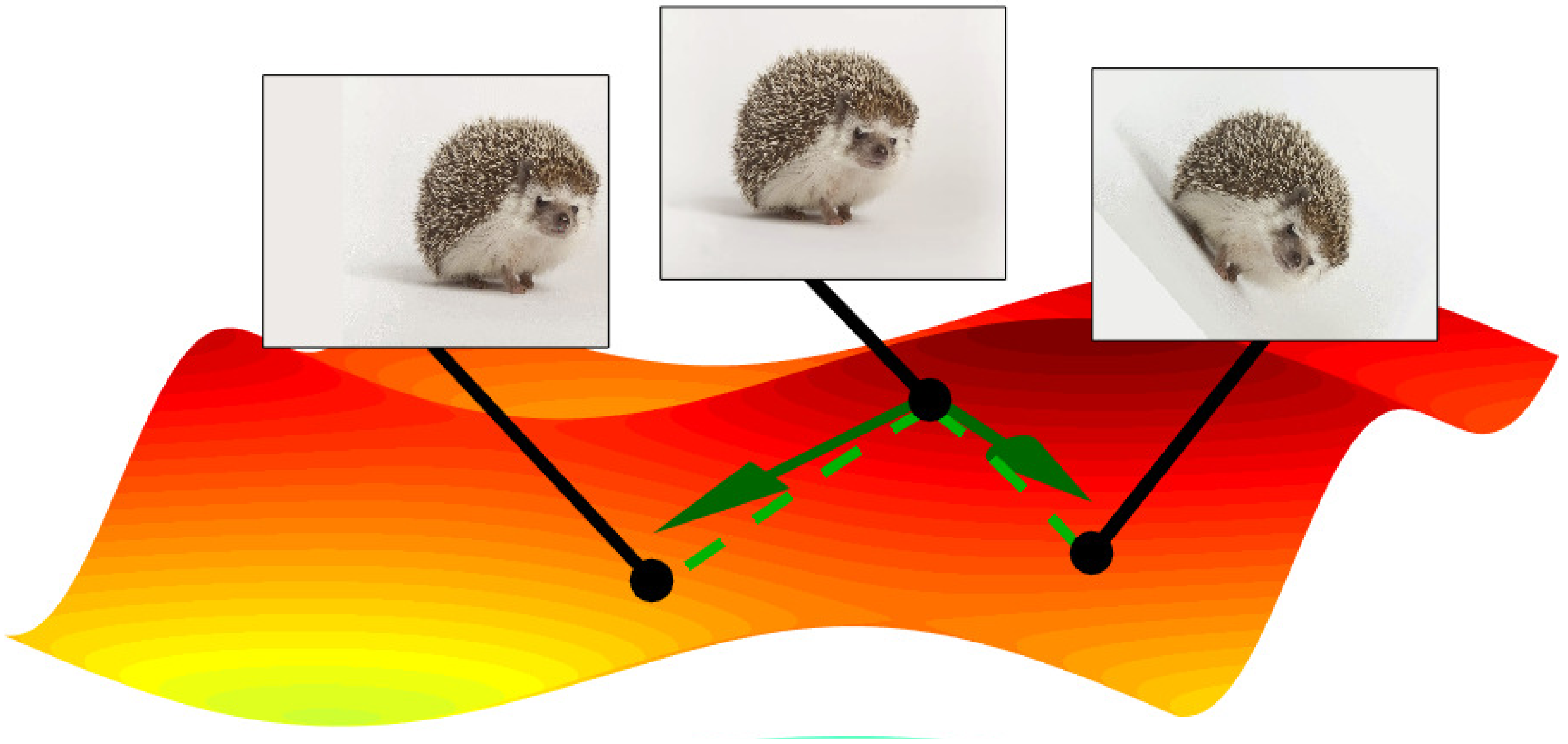
- Radiological vs. natural image datasets: different relevant semantics/abstract features.
- In order to quantify this domain shift and its relation to learning, I wanted to investigate:

    1. Does dataset **intrinsic dimension** generally differ between these two domains? **Yes!**
    2. Is dataset **ID** *linearly correlated* to network generalization ability *within* these domains? **Yes!**
        1. Does this relationship differ *between* them? **Yes!**

One goal of this talk: work towards a *formal model* that explains my empirical findings.

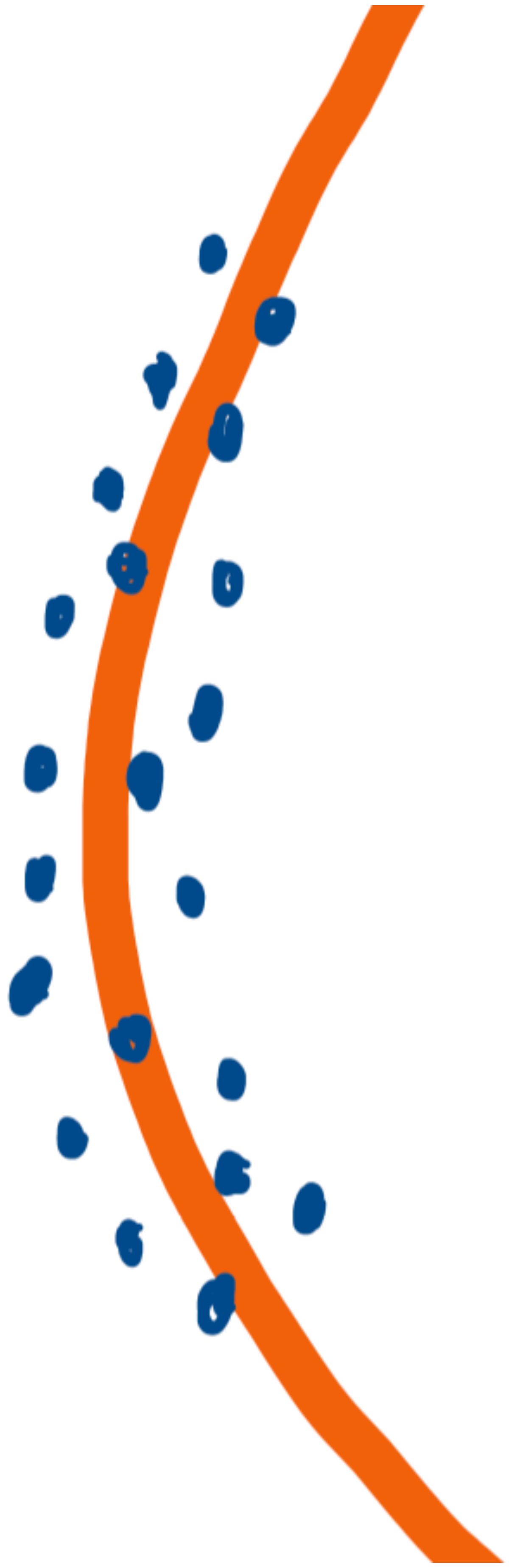
Estimating the intrinsic dimension of a manifold

2-dimensional manifold



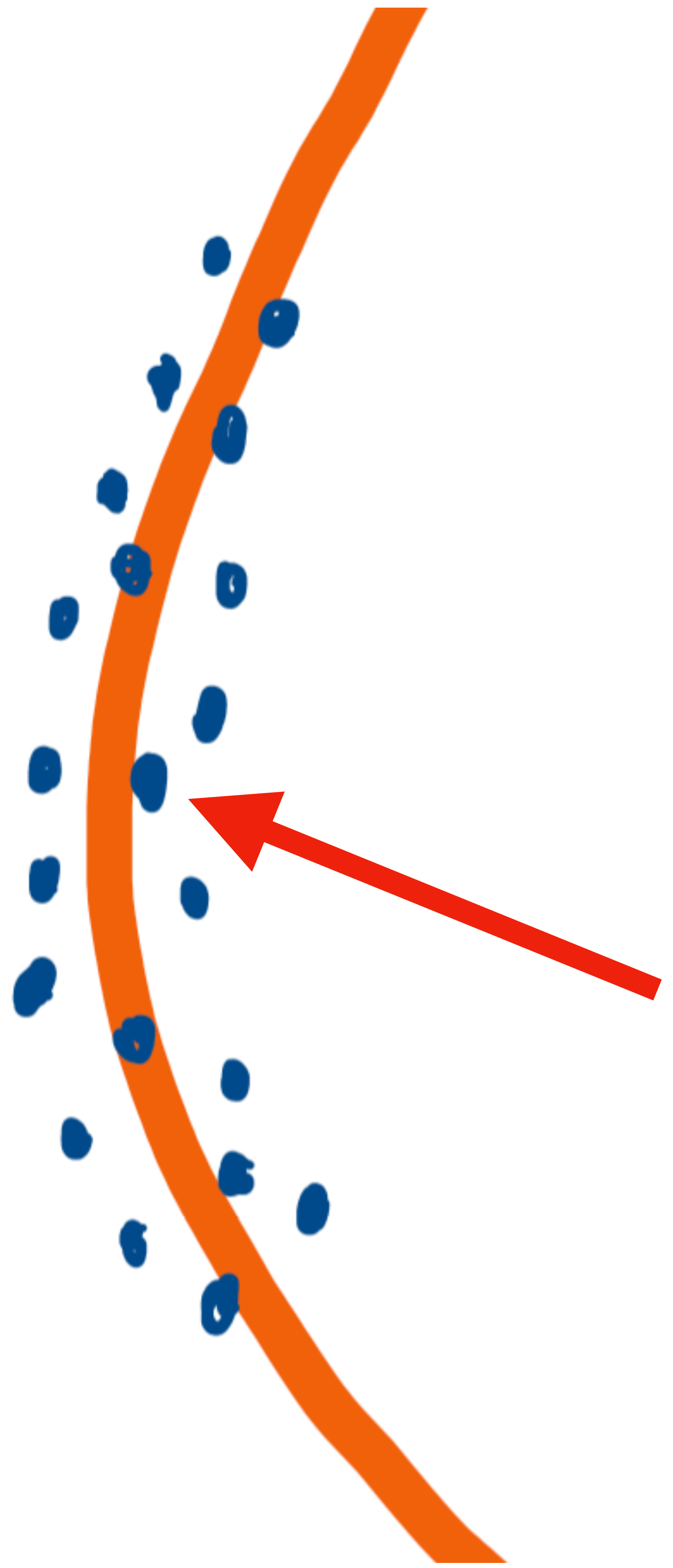
Buchanan et al., ICLR 2020

1-dimensional manifold



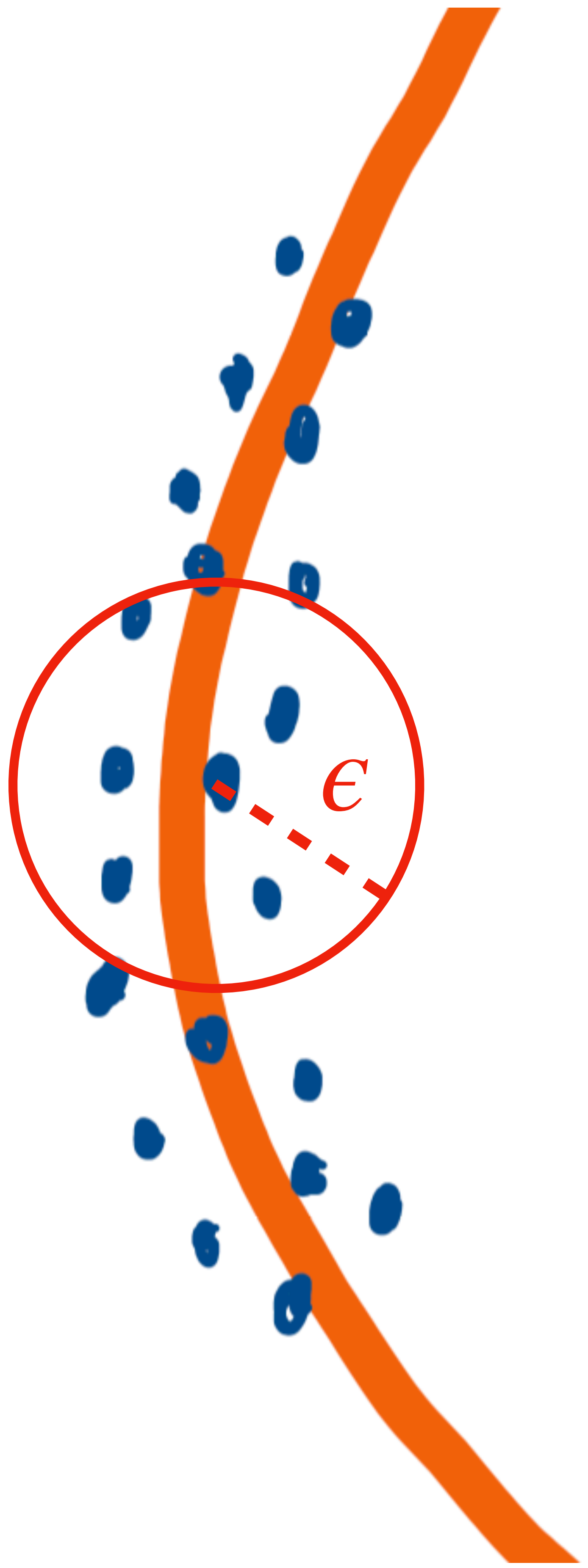
Estimating the intrinsic dimension of a manifold

- Center some ϵ-ball on a datapoint:

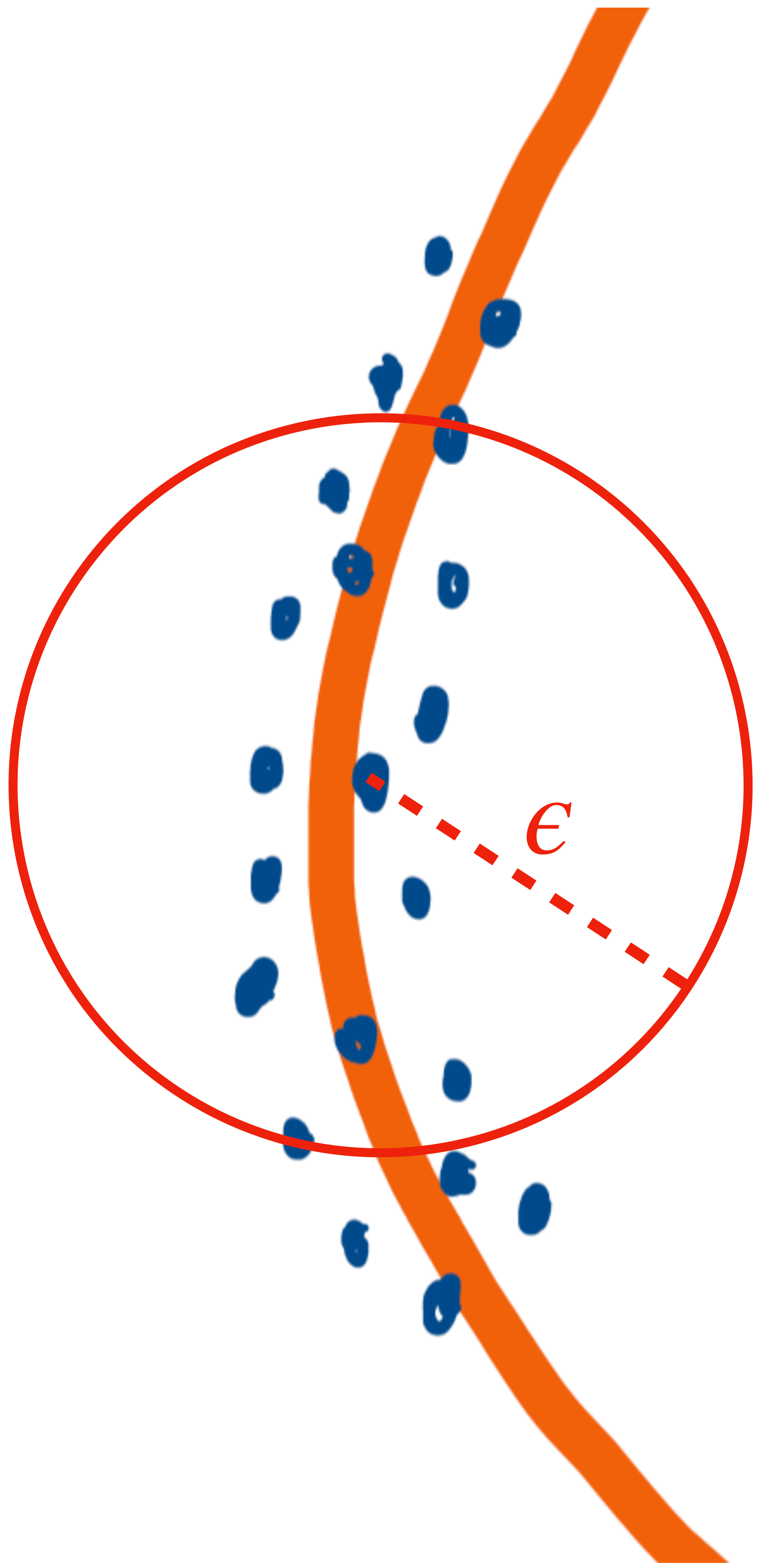


Estimating the intrinsic dimension of a manifold

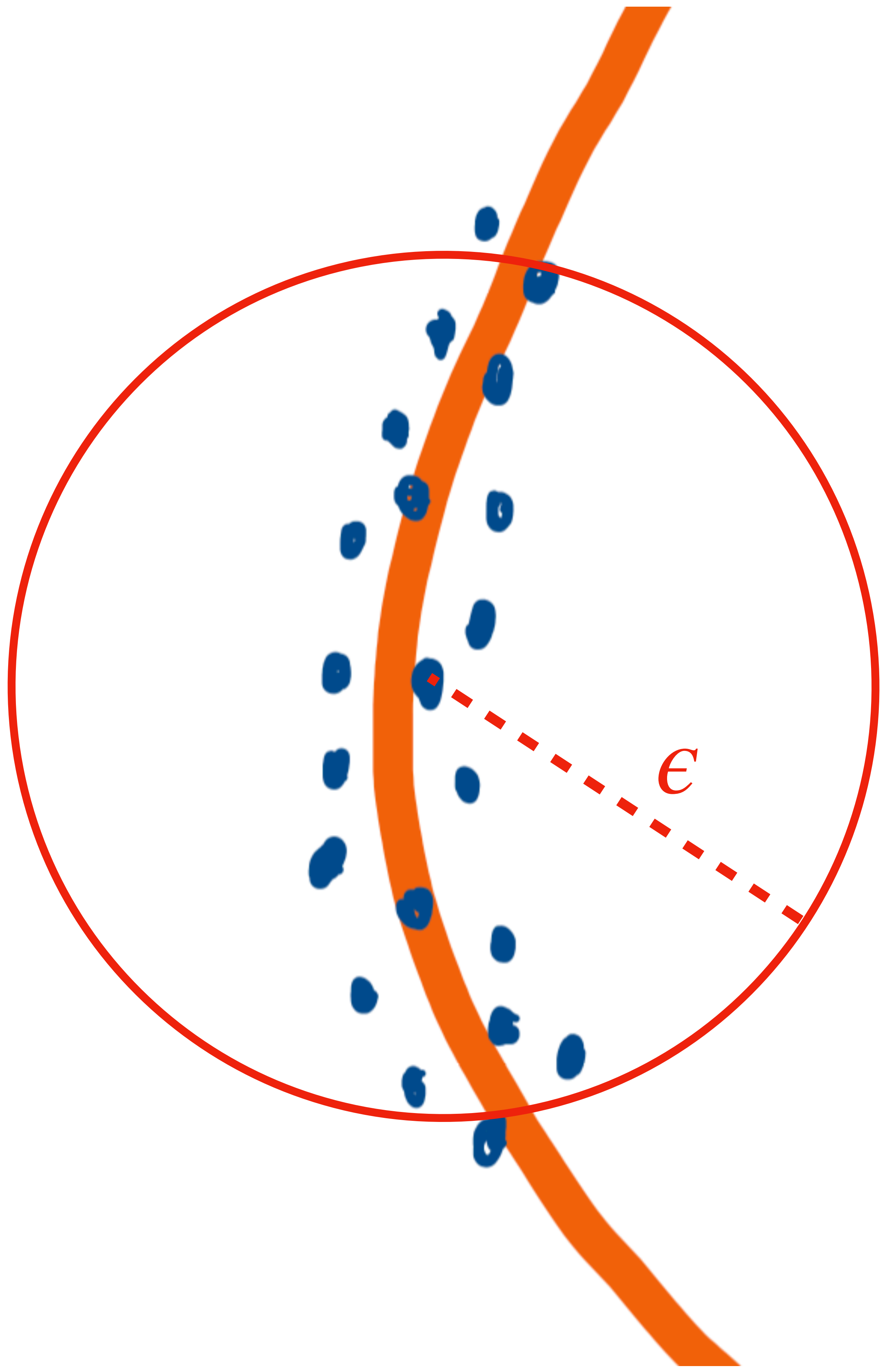
- Center some ϵ-ball on a datapoint:



Estimating the intrinsic dimension of a manifold



Estimating the intrinsic dimension of a manifold

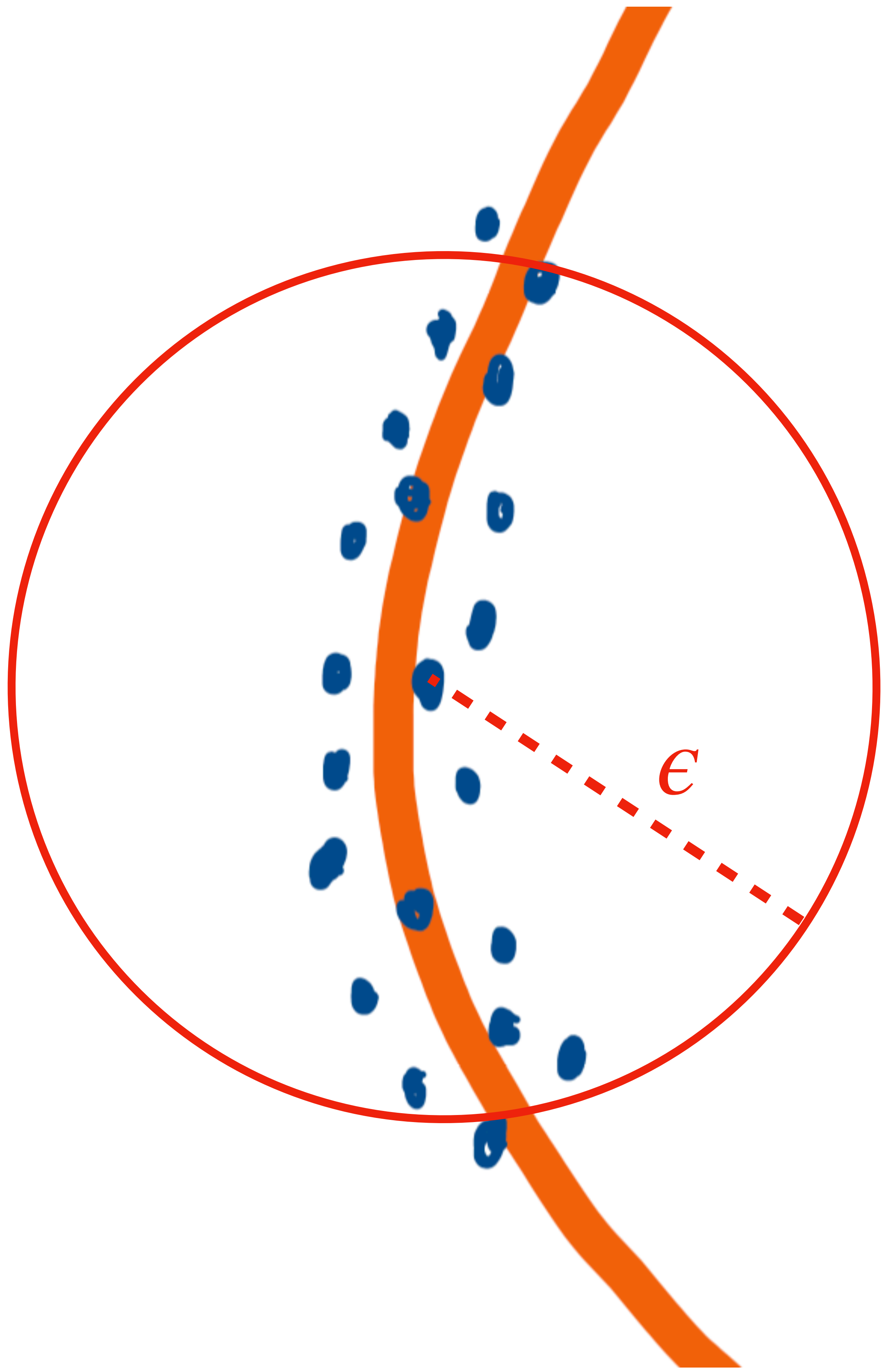


Estimating the intrinsic dimension of a manifold

- How does the volume of M enclosed within the ϵ-ball, i.e. *the number of datapoints within it*, **scale with ϵ?**

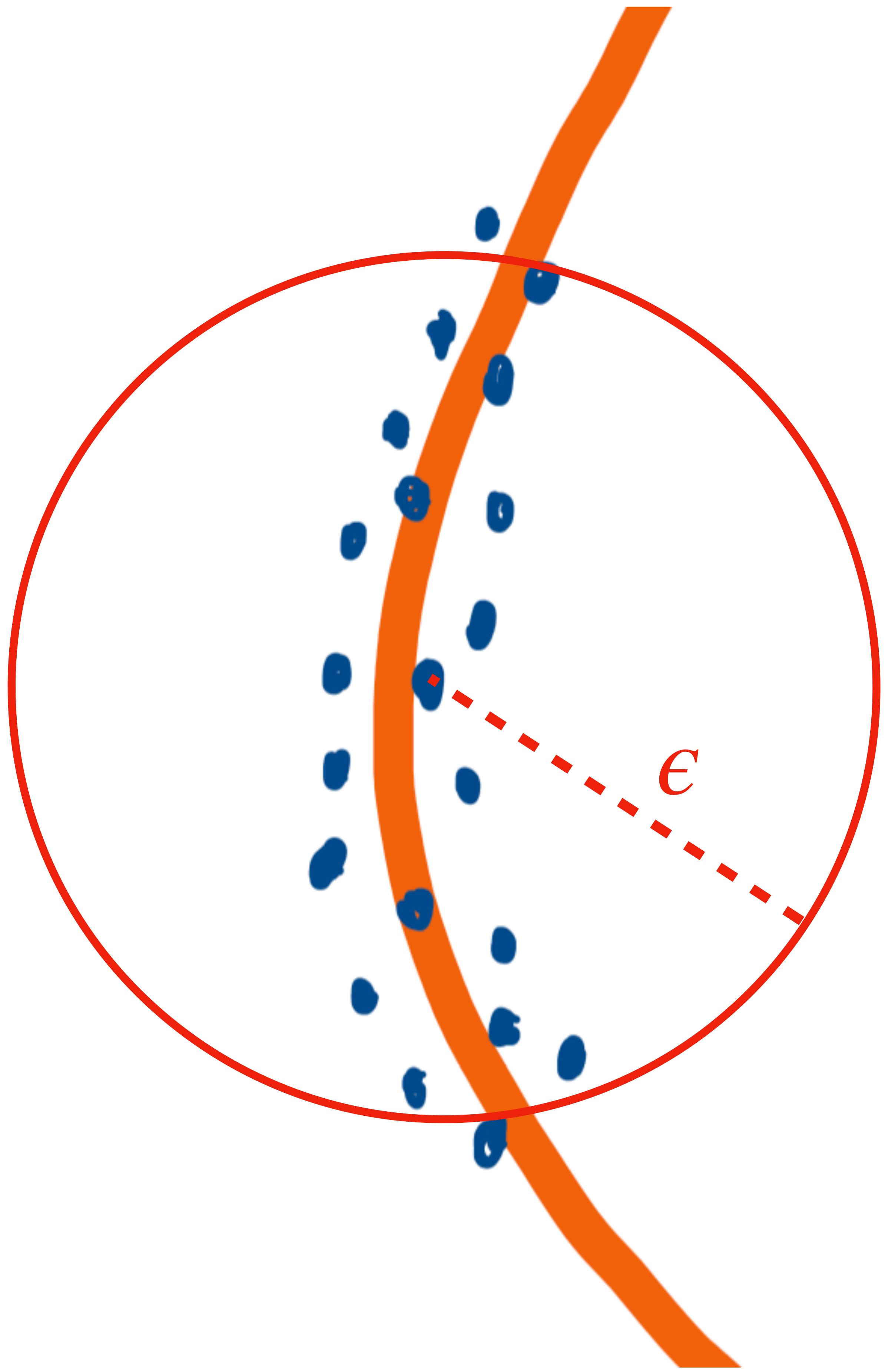
- This example: $\mathrm{vol} \sim O(\epsilon)$

- In general: $\mathrm{vol} \sim O(\epsilon^m)$



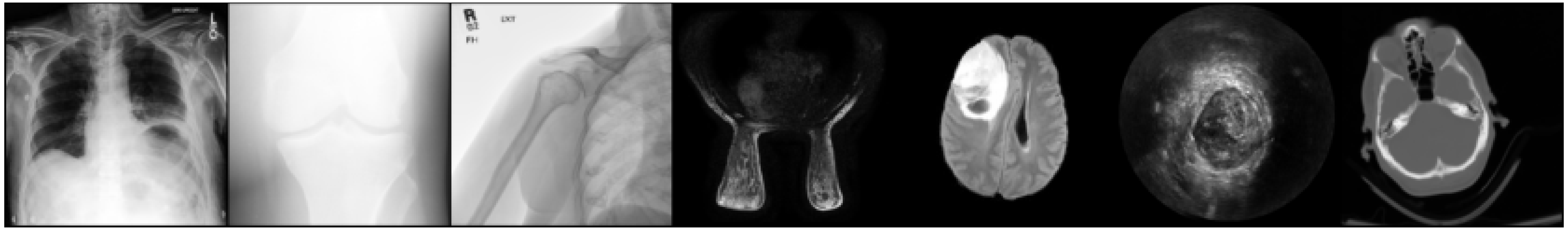
Estimating the intrinsic dimension of a manifold

- Levina et al. 2004:
  - Model data within the ϵ-ball as sampled via Poisson process.
  - Then, obtain m that maximizes the likelihood of the model generating *the entire dataset*
  - Hyperparameter to choose: ϵ to count datapoints within. In practice, the number of neighbors k to include is used instead.
  - Note: no image *labels* used, only the image data itself.



Levina, E. & Bickel, P. Maximum Likelihood Estimation of Intrinsic Dimension. in NeurIPS 2004.
MacKay, D. & Ghahramani, Z. Comments on'maximum likelihood estimation of intrinsic dimension'by e. levina and p. bickel (2004).

My Datasets



- Seven commonly used public radiology datasets
    - Seven anatomies and three modalities
- Binary classification labels assigned to each dataset for various radiological tasks
- For all experiments, datasets used had an even class balance

Table 1. Summary of datasets explored in this work.

Dataset	Anatomy/Modality	Binary Classification Task
CheXpert [14]	Chest X-ray	Pleural Effusion
OAI [23]	Knee X-ray	Osteoarthritis
MURA [26]	Musculoskeletal X-ray	Abnormality
DBC [27]	Breast MRI	Breast Cancer
BraTS 2018 [19]	Brain MRI	Glioma
Prostate-MRI [28]	Prostate MRI	Prostate Cancer
RSNA-IH-CT [11]	Brain CT	Intracranial Hemorrage

Experiment 1: The Intrinsic Dimension of Radiology Datasets

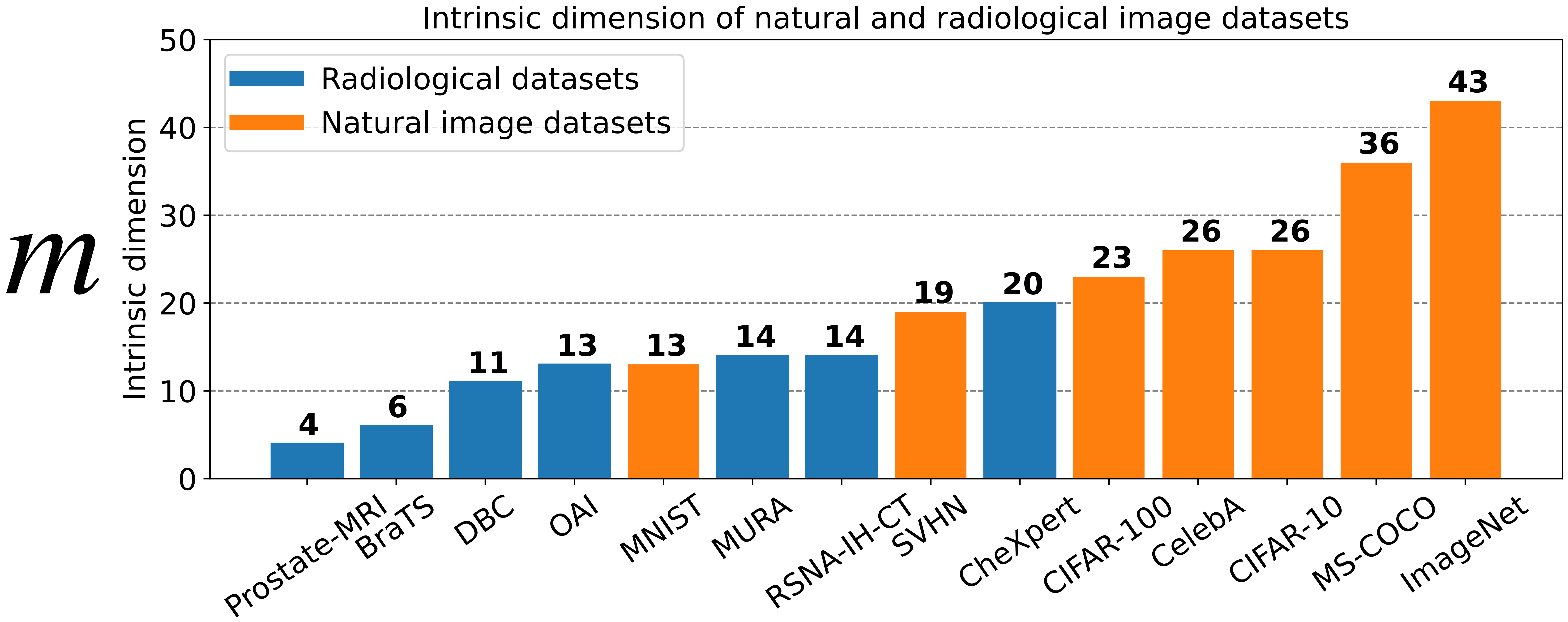
- Experimental method:
  - Use the maximum likelihood method to estimate intrinsic dimension m of each radiology dataset
  - Compare to the results found for natural images with the same method from Pope et al.

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. ICLR 2021

Results for Experiment 1: The Intrinsic Dimension of Radiology Datasets

1. Like natural image (NI) datasets, radiological image (RI) datasets have $m \ll d$. Also, modifying d (resizing, modifying number of pixel channels, etc.) had no effect on m.

2. **But**, *radiology datasets tend to have lower intrinsic dimension than natural image datasets.*



Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. ICLR 2021

Experiment 2: Generalization Ability, Learning Difficulty and Dataset Intrinsic Dimension

- Central question: *How does the intrinsic dimension of a training dataset affect the difficulty of learning to generalize to new samples?*

- Recall Pope et al.'s results on natural images: increased intrinsic dimension leads to worse generalization ability:

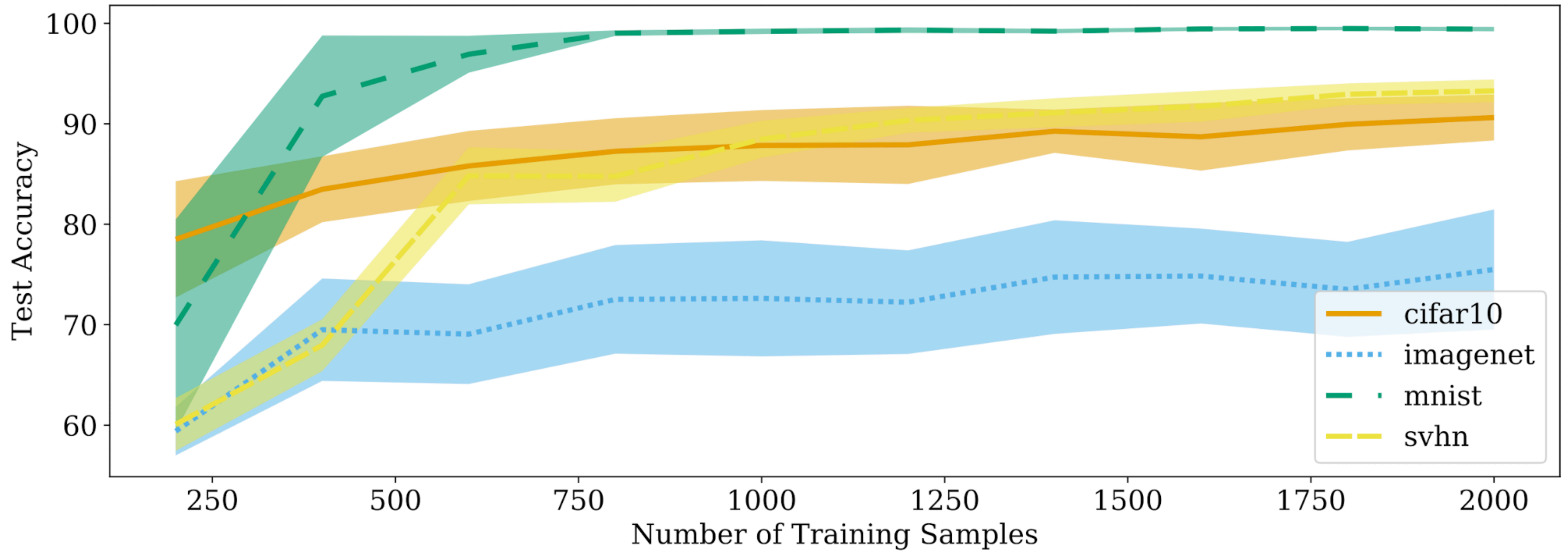


Figure 6: Sample complexity of real datasets. Standard errors are shown $N = 5$ class pairs.

- But what is the explicit form of this relationship? Do radiological datasets behave similarly?

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. ICLR 2021

Experiment 2: Generalization Ability, Learning Difficulty and Dataset Intrinsic Dimension

- To try to find a broad result for such a general question, I ran many experiments with settings as controlled as possible.

- For each of the seven datasets:

  1. Sample $N_{\text{train}} \in \{250, 500, \ldots, 2000\}$ training images, and 750 test images, with even class balancing.

  2. Train each of 7 neural net models on each training set size, for the dataset's corresponding classification task.

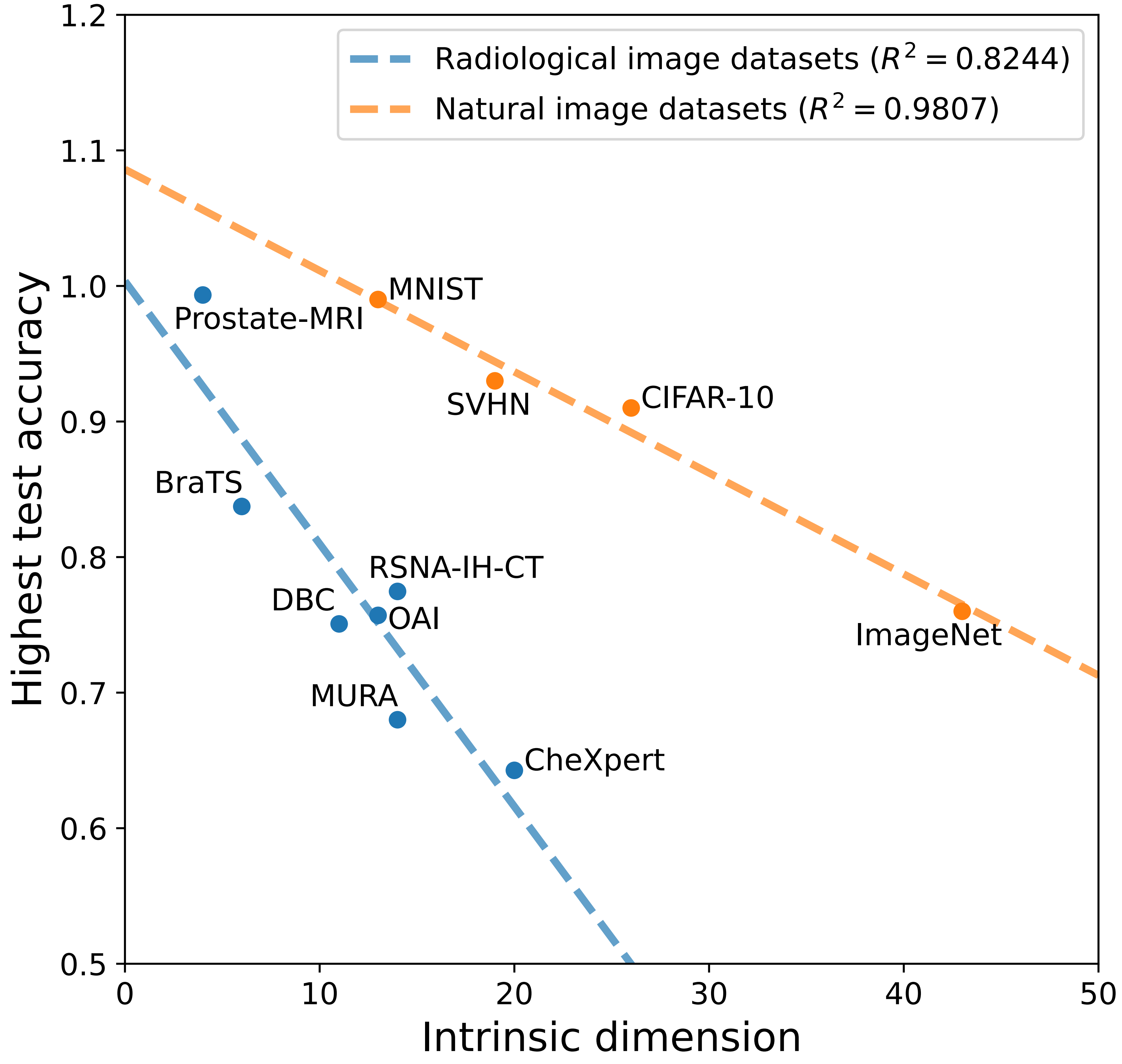
  3. Evaluate the trained network's performance on the test set.

Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. ICLR 2021

Results for Experiment 2: Generalization Ability, Learning Difficulty and Dataset Intrinsic Dimension

- For *both domains*, we found GA to be sharply linearly correlated with dataset ID:
  - **Natural image datasets:** averaged over all 9 training set sizes for ResNet-18:
    - $R^2 = 0.91 \pm 0.12$ and
    - slope = -0.0077 ± 0.0004
  - **Radiological datasets:** averaged over ***all 7 models*** *and 9 training set sizes:*
    - $R^2 = 0.70 \pm 0.08$ and
    - slope = -0.019 ± 0.001
- **Key result:** The negative correlation of GA with dataset ID *within* a domain is strong, but *much sharper* for radiology datasets!

ResNet-18 example results



Results for Experiment 2: Generalization Ability, Learning Difficulty and Dataset Intrinsic Dimension

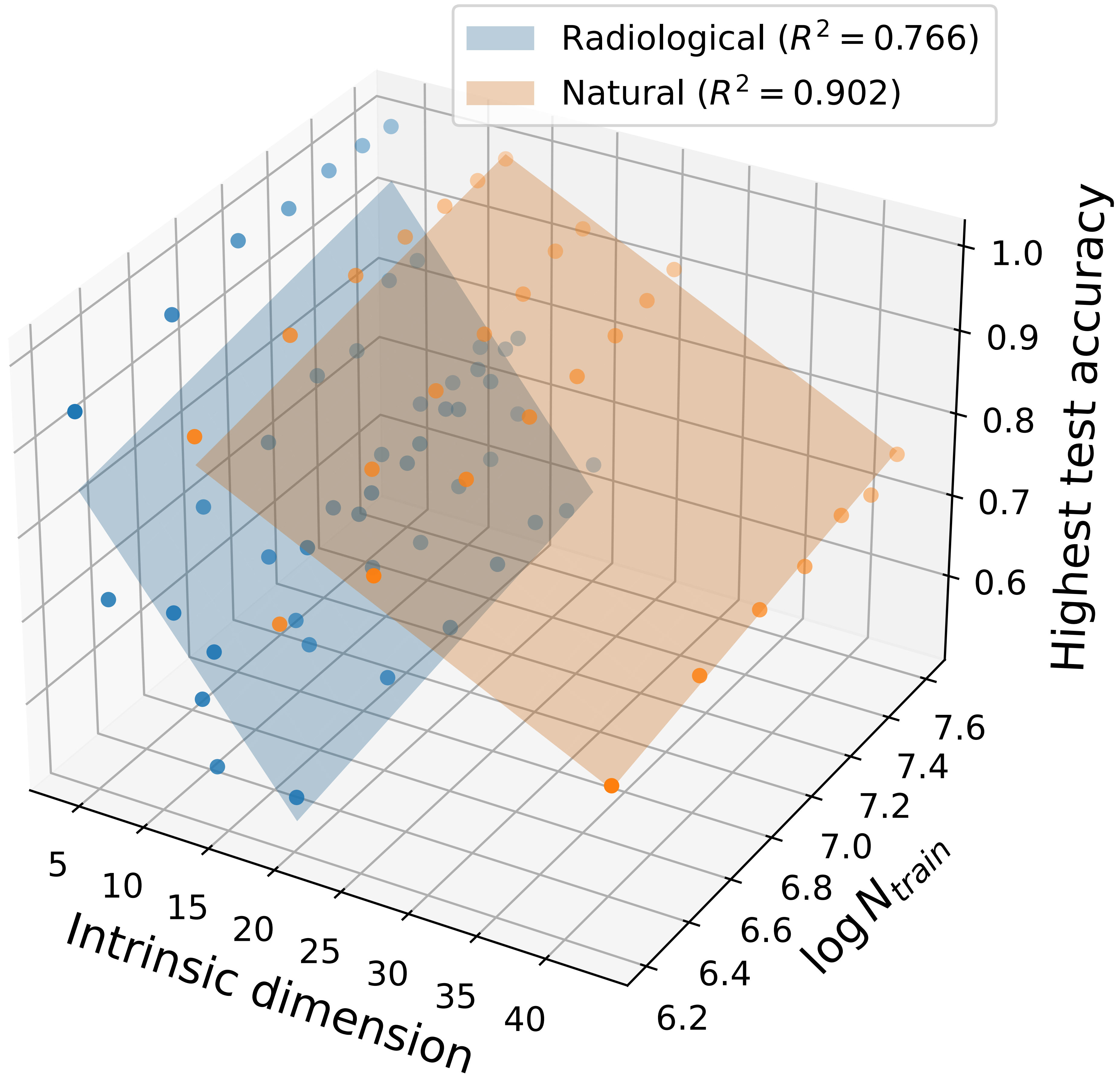
Table 2. Model-Specific Details and Results (averaged over N_{train} with std. dev.)

Model	Training batch size	R^2 (GA vs. ID)	slope $a_{\text{GA,ID}}$ (GA vs. ID)
ResNet-18 [12]	200	0.756 ± 0.052	-0.0199 ± 0.0009
ResNet-34 [12]	128	0.772 ± 0.071	-0.0193 ± 0.0012
ResNet-50 [12]	64	0.781 ± 0.066	-0.0207 ± 0.0010
VGG-13 [27]	32	0.646 ± 0.048	-0.0194 ± 0.0009
VGG-16 [27]	32	0.623 ± 0.066	-0.0184 ± 0.0008
VGG-19 [27]	32	0.597 ± 0.100	-0.0168 ± 0.0031
Squeezenet 1.1 [14]	32	0.580 ± 0.073	-0.0173 ± 0.0011
DenseNet-121 [13]	32	0.770 ± 0.073	-0.0190 ± 0.0009
DenseNet-169 [13]	32	0.765 ± 0.061	-0.0189 ± 0.0008

Results for Experiment 2: Generalization Ability, Learning Difficulty and Dataset Intrinsic Dimension

- My experiments also verified the sample complexity results of Narayanan et al. (NeurIPS 2010): generalization ability $\sim O(\log N_{\text{train}})$

- I also tried modifying the assigned classification task/labeling for certain datasets; the overall results were hardly changed.



Narayanan, H. & Mitter, S. Sample Complexity of Testing the Manifold Hypothesis. in NeurIPS 2010.

Conclusions from Results

- **Quantified empirical evidence of the difference in learning** from the imaging domains of **natural** and **radiological** images, in terms of:

    1. Intrinsic dataset feature dimension (ID).

    2. Sharpness of the relationship between dataset ID and the difficulty of a trained network to generalize to new data.

- Possible qualitative take-aways:

    1. Despite numbering fewer than that of natural image datasets, the intrinsic features of radiological datasets are more difficult to learn from/complex.

    2. Assumptions about natural images and models designed for them should not be naively extended to radiological images.

The Interesting Part: Towards Theoretical Reasons for the Results

To my knowledge, there is no rigorous mathematical explanation in the literature for:

1. The linear relationship between network generalization ability (GA) and dataset intrinsic manifold dimension (ID), beyond qualitative intuitions of correlation.

2. The noticeable difference in sharpness of the GA vs. ID slope *between* the two domains, despite the tightness of the correlation *within* each domain.

Possible Theoretical Explanation 1: Something Trivial

1. I attempted to rule out trivial reasons for my results by using a range of models, training set sizes, ablation studies, etc over very fixed experimental settings.

2. Could this be due to my choice of estimator?

    1. But Ansuini et al. had similar results (for network internal data representations) with a *different* ID estimator

    2. The validity of the MLE estimator was supported with various experiments for natural images in Pope et al.

3. Despite these, a trivial explanation is still possible.

Ansuini, A., Laio, A., Macke, J. H. & Zoccolan, D. Intrinsic dimension of data representations in deep neural networks. in NeurIPS 2019.
Pope, P., Zhu, C., Abdelkader, A., Goldblum, M. & Goldstein, T. The Intrinsic Dimension of Images and Its Impact on Learning. ICLR 2021

Possible Theoretical Explanation 2: Relating Model Manifold Fitting Error to Data Manifold Intrinsic Dimension

- I wanted to relate the error in fitting an empirical risk-minimized manifold to the dataset manifold, to the intrinsic dimension of the data manifold.
- Narayanan et al. found that to achieve some fixed GA, the number of training set samples needed is exponential with dataset ID.
  - **But, what about the ID needed to achieve some GA for a *fixed* training set size?** And why did I find this relationship to be *linear*?
- Completely unclear why the GA vs. ID slope should differ so much *between* domains.

Narayanan, H. & Mitter, S. Sample Complexity of Testing the Manifold Hypothesis. in NeurIPS 2010.

Possible Theoretical Explanation 3: Relating Effective Model Capacity to Dataset Intrinsic Dimension

- Plenty of recent studies (e.g. Birdal et al.) have tried to establish measures of *effective* learning capacity of neural nets, to mitigate limitations of classical methods like VC-dimension.
- For example, these studies model neural network parameters as *fractal structures* evolving through training. They then measure the intrinsic *Hausdorff dimension* or *persistent homology dimension* of the parameters.
- Finally, these studies provide theoretical bounds on generalization ability based on the intrinsic dimension of the model parameters:

Proposition 1. *Let $\mathcal{W} \subset \mathbb{R}^d$ be a (random) compact set. Assume that* **H**1 *holds,* ℓ *is bounded by* B *and L-Lipschitz continuous in* w*. Then, for* n *sufficiently large, we have*

$$\sup_{w \in \mathcal{W}} |\hat{\mathcal{R}}(w, S) - \mathcal{R}(w)| \leq 2B\sqrt{\frac{[\dim_{\mathrm{PH}}\mathcal{W} + 1]\log^2(nL^2)}{n} + \frac{\log(7M/\gamma)}{n}}, \qquad (4)$$

with probability at least $1 - \gamma$ *over* $S \sim \mathcal{D}^{\otimes n}$.

Birdal, T. et al. Intrinsic Dimension, Persistent Homology and Generalization in Neural Networks. in NeurIPS 2021.

Possible Theoretical Explanation 3: Relating Effective Model Capacity to Dataset Intrinsic Dimension

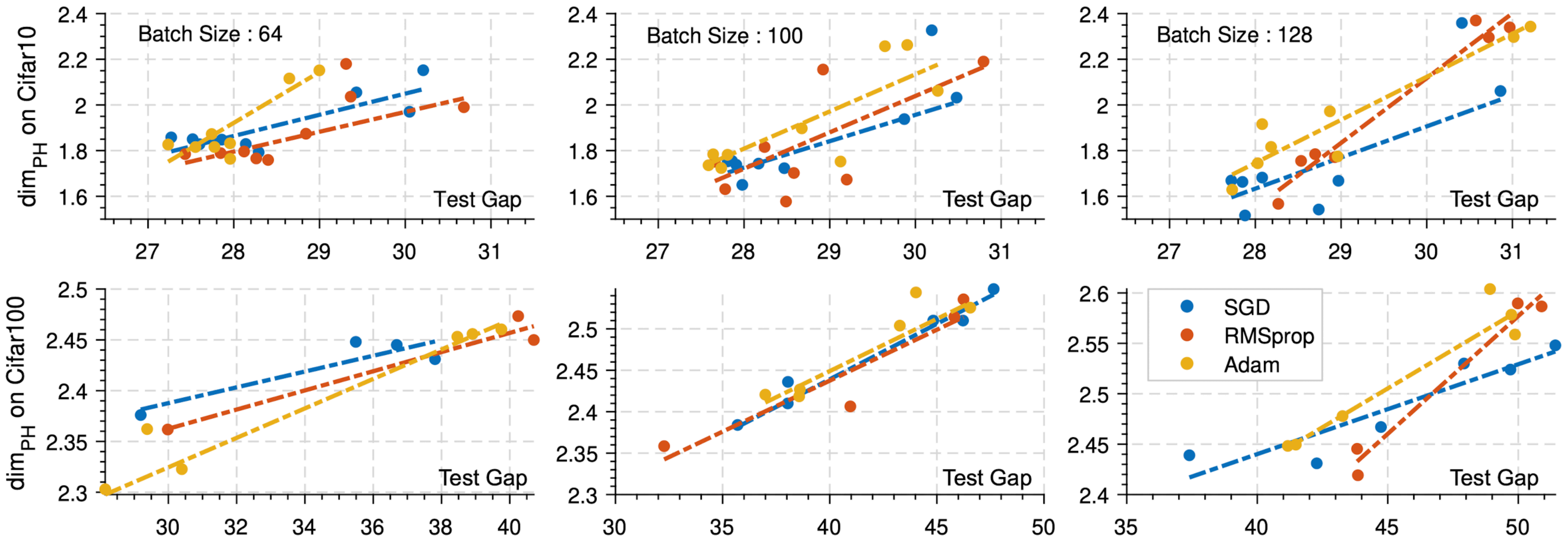


Figure 3: (Estimated) persistent homology dimension vs generalization error (training accuracy - test accuracy) for different datasets (top row CIFAR10, bottom row CIFAR100) and optimizers on AlexNet. We plot the data points and lines of best fit. Note that the PH dimension is *directly* correlated with the generalization error and is consistent across datasets and optimizers.

- The same study found supporting experimental evidence for **linear correlation between generalization ability and network parameter intrinsic dimension.**

Birdal, T. et al. Intrinsic Dimension, Persistent Homology and Generalization in Neural Networks. in NeurIPS 2021.

Possible Theoretical Explanation 3: Relating Effective Model Capacity to Dataset Intrinsic Dimension

- **However**, it's unclear how the ID of a trained network's parameters can be mathematically related to the ID of the training/testing dataset and/or the ID of the network's internal representations.

- *One "hand-wavey" explanation:* neural network training is very data driven, so the effective complexity/intrinsic dimension of a neural network is shaped by the effective complexity/intrinsic dimension of the dataset that it is fit to.

    - However, I'm not sure how to formalize this intuition, or if it's even correct.

- Finally, one should be careful when trying to develop causal relationships with generalization… (Jiang et al. 2019).

Jiang, Y., et al. Fantastic Generalization Measures and Where to Find Them. arXiv.1912.02178 (2019).

Possible future experimental studies

1. Other supervised tasks beyond binary classification, and potentially semi-supervised, self-supervised and unsupervised training methods

2. Other dataset domains in medical imaging, e.g. pathology

3. Explore further practical uses of dataset intrinsic dimension (e.g. to guide generative modeling), and visualization/understanding of what these intrinsic features of radiology datasets *are*.

In Conclusion

- I present what is (hopefully) a nontrivial, interesting mathematical problem, in explaining my empirical results.

- A formal model that explains my observed behavior both within and between natural and radiological imaging domains could lead the way towards the more principled development of methods specifically tailored for medical image analysis.

Questions?

More on this research

arXiv Preprint

Code

Published paper

Contact Me

E-Mail

nicholas.konz@duke.edu

Twitter

LinkedIn

My Lab