Universality of intrinsic dimension of latent representations across models

UniReps Workshop

NEURAL INFORMATION PROCESSING SYSTEMS

Teresa Scheidt, Lars Kai Hansen

Technical University of Denmark,
Department of Applied Mathematics and Computer Science

tksc@dtu.dk
Teresa Scheidt
@TeresaScheidt

Introduction

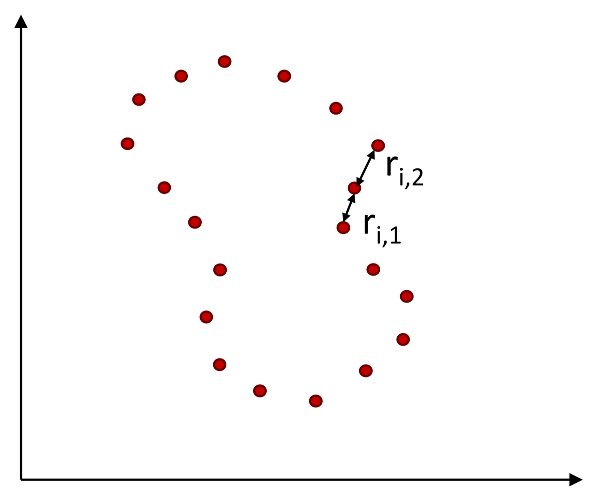
Manifold hypothesis: high-dimensional real-world datasets populate low-dimensional manifolds in neural networks
→ **Do different networks learn the same manifolds??**

Intrinsic dimension: local geometric property of latent space, minimal number of variables to represent data [1]
→ **Does intrinsic dimension depend on the class??**

Methods

Intrinsic dimension

- Estimated with TwoNN estimator [2]
- Estimated per class



1. For each point i calculate distance to first and second neighbor ($r_{i,1}$, $r_{i,2}$) and ratio μ_i
$$\mu_i = \frac{r_{i,2}}{r_{i,1}}$$
2. Probability distribution is given by:
$$P(\mu|d) = d^N \prod_{i=1}^{N} \mu_i^{-(d+1)}$$
3. Intrinsic dimension estimate d is given by:
$$\hat{d} = \frac{N}{\sum_i \log \mu_i}, \hat{\sigma}(\hat{d}) = \frac{\hat{d}}{\sqrt{N}}$$

Models and dataset

Dataset: cifar10 and cifar100 [3]
Models: data2vec [4], ViT [5] and BEiT [6]

Model	Layers	Embedding Size	Huggingface model
data2vec	12	768	facebook/data2vec-vision-base
data2vec finetuned	12	768	facebook/data2vec-vision-base-ft1k
ViT base	12	768	google/vit-base-patch16-224
ViT large	24	1024	google/vit-large-patch16-224
ViT huge	32	1280	google/vit-huge-patch14-224-in21k
ViT base finetuned	12	768	nateraw/vit-base-patch16-224-cifar10
BEiT base	12	768	microsoft/beit-base-patch16-224
BEiT large	24	1024	microsoft/beit-large-patch16-224
BEiT base finetuned	12	768	jadohu/BEiT-finetuned

Results

Correlations between models

Pearson's r (lower triangle) and Spearsman's rho (upper triangle)
For all p<0.005

	data2vec	data2vec ft	vit	vit ft	beit	beit ft	vit large	beit large	vit huge
data2vec	1.000	0.952	0.939	0.903	0.915	0.903	0.927	0.867	0.927
data2vec ft	0.948	1.000	0.988	0.939	0.988	0.964	0.903	0.891	0.927
vit	0.960	0.978	1.000	0.903	0.976	0.939	0.927	0.927	0.964
vit ft	0.951	0.967	0.995	1.000	0.952	0.976	0.891	0.867	0.867
beit	0.961	0.982	0.995	0.993	1.000	0.988	0.879	0.879	0.915
beit ft	0.964	0.968	0.988	0.991	0.994	1.000	0.867	0.855	0.891
vit large	0.876	0.875	0.928	0.908	0.889	0.887	1.000	0.954	0.976
beit large	0.908	0.950	0.985	0.980	0.968	0.961	0.954	1.000	0.964
vit huge	0.940	0.937	0.981	0.969	0.962	0.963	0.966	0.978	1.000

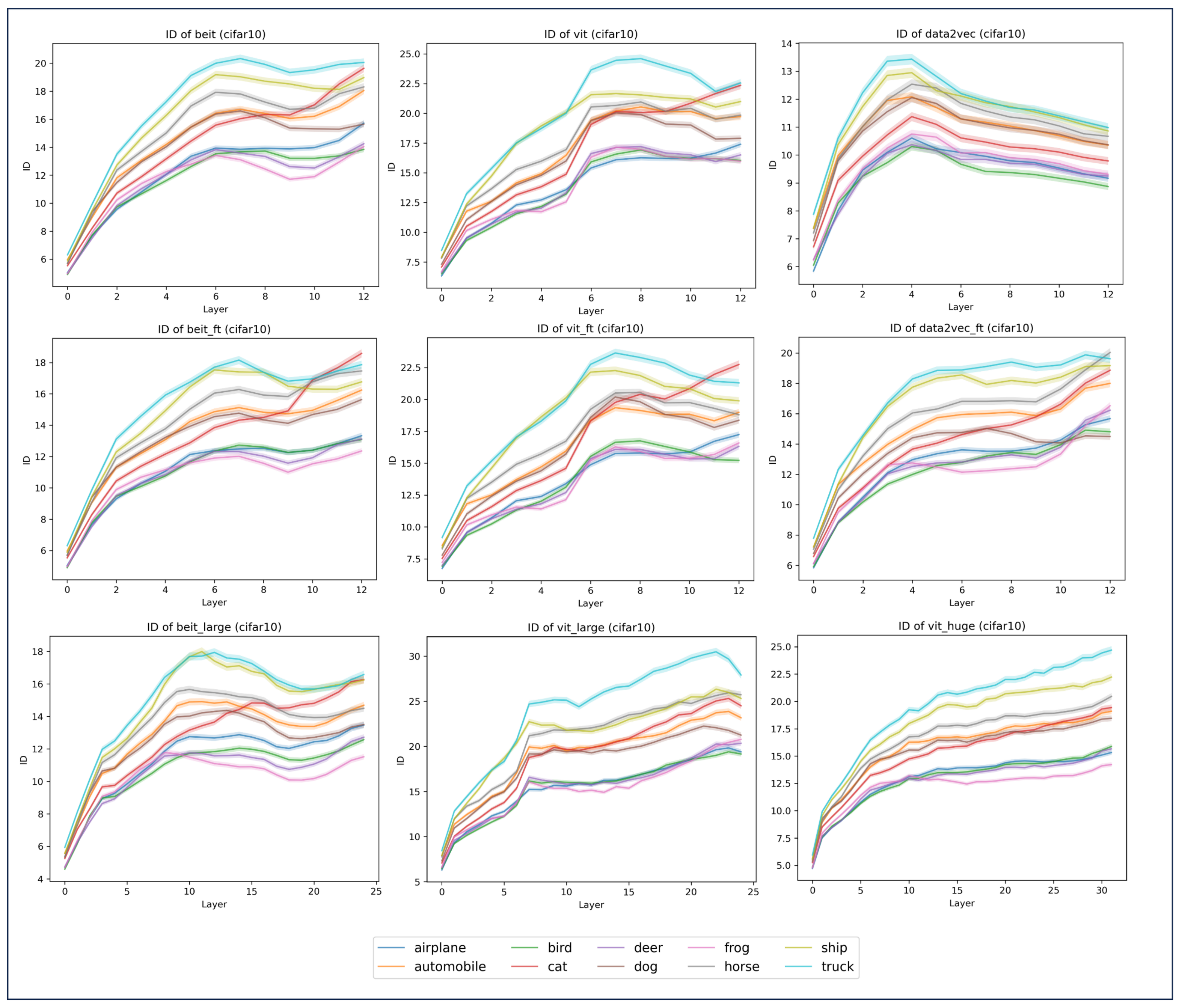
Correlations for cifar10

	data2vec	vit	vit large	beit	beit large
data2vec	1.000	0.851	0.850	0.847	0.840
vit	0.869	1.000	0.916	0.983	0.938
vit large	0.838	0.905	1.000	0.905	0.966
beit	0.855	0.984	0.888	1.000	0.929
beit large	0.844	0.926	0.962	0.920	1.000

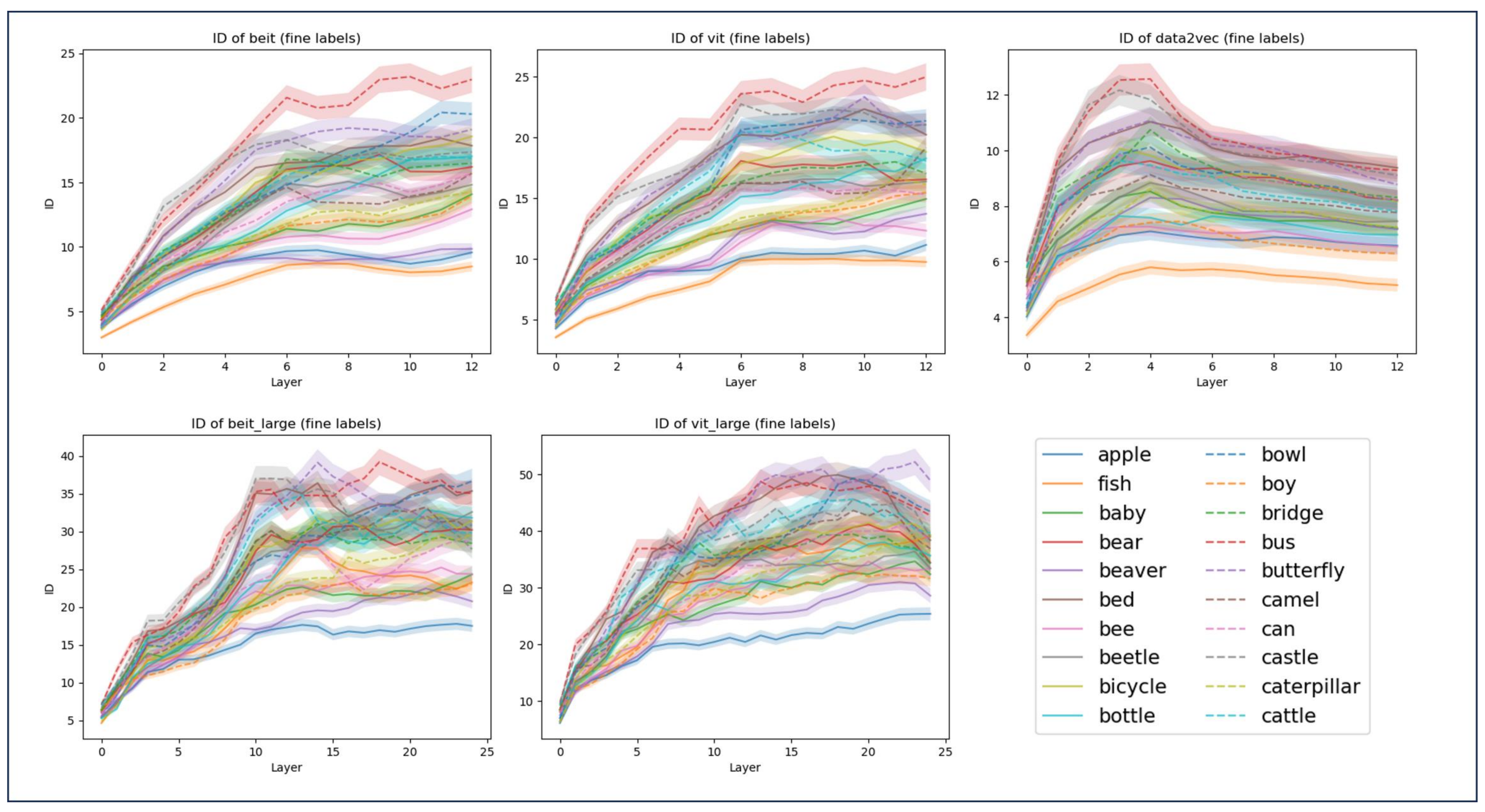
Correlations for cifar100

Results

Development of intrinsic dimension across layers for all models:
Similar patterns and ranking across models



Intrinsic dimension for classes of cifar10



Intrinsic dimension for 20 classes of cifar100

Conclusion and Outlook

- Intrinsic dimension is dependent on class
- General evolution and ordering of intrinsic dimension of classes universal across models

→ supports hypothesis that models learn similar representations

Future research:

- How can intrinsic dimension be related to the data?
- How similar are the learned manifolds?
- How can this be used for knowledge transfer and model alignment?

References

[1] A. Ansuini, A. Laio, J. H. Macke, and D. Zoccolan. Intrinsic dimension of data representations in deep neural networks. Advances in Neural Information Processing Systems, 32, 2019.

[2] E. Facco, et al. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. Scientific reports, 7(1):12140, 2017.

[3] A. Krizhevsky, et al. Learning multiple layers of features from tiny images. 2009.

[4] A. Baevski, et al. Data2vec: A general framework for self-supervised learning in speech, vision and language. In International Conference on Machine Learning, pages 1298–1312. PMLR, 2022.

[5] O. Russakovsky, et al.. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.

[6] H. Bao, et al . Beit: Bert pre-training of image transformers. arXiv preprint arXiv:2106.08254, 2021.