ICLR' 21 Spotlight

The Intrinsic Dimension of Images and Its Impact on Learning

Phil Pope, Chen Zhu, Ahmed Abdelkader,
Micah Goldblum, Tom Goldstein

발표자: 공채린

Motivation

- Natural Images are believed to reside in lower dimension manifold.

→ How low is the intrinsic dimension?

→ How does it affect the *difficulty* in deep learning tasks?

Contribution

- Verify the reliability of MLE-based intrinsic dimension estimator using GANs

- Measure and report dimensionalities of popular datasets like MNIST, CIFAR-10, ImageNet.

- Discover that intrinsic dimensionality positively correlates with sample complexity, while extrinsic dimensionality has no influence.

Intrinsic Dimension Estimator

- Assume constant density in local neighborhood of each data point, and model # data points within certain radius with Poisson process.

Levina & Bickel (NIPS 2004)

- Maximum Likelihood Estimator of intrinsic dimension at point x:

$$\hat{m}_k(x) = \left[\frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{T_k(x)}{T_j(x)} \right]^{-1}$$

$T_k(x)$: Euclidean distance to k-th neareast data point from x

- Estimator across the data points:

$$\bar{m}_k = \left[\frac{1}{n} \sum_{i=1}^{n} \hat{m}_k(x_i)^{-1} \right]^{-1} = \left[\frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k-1} \log \frac{T_k(x_i)}{T_j(x_i)} \right]^{-1}$$

Proof Sketch

From Levina & Bickel (NIPS 2004)

$x_1, x_2, \dots, x_n \in R^p, X_i = g(Y_i), Y_i \sim i.i.d.\, f(x) \sim R^m$

$T_k(x)$: Euclidean distance to k-th neareast data point from x

$\frac{k}{n} \approx f(x)V(m)[T_k(x)]^m$ $V(m)$: Volume of unit sphere in m dimension

"The proportion of sample points falling into a ball around x is roughly f(x) times the volume of the ball."

Assume $f(x) \approx const$ in neighborhood of x, let $S_x(R)$: sphere of radius R around x

$N(t,x) = \sum_{i=1}^{n} \mathbf{1}\{X_i \in S_x(t)\}$ → Poisson process

$\lambda(t) = f(x)V(m)mt^{m-1}$

$L(m,\theta) = \int_0^R \log \lambda(t)\, dN(t) - \int_0^R \lambda(t)\, dt$ → Log-likelihood (θ: $\log f(x)$)

$$\frac{\partial L}{\partial \theta} = \int_0^R dN(t) - \int_0^R \lambda(t)dt = N(R) - e^{\theta}V(m)R^m = 0$$

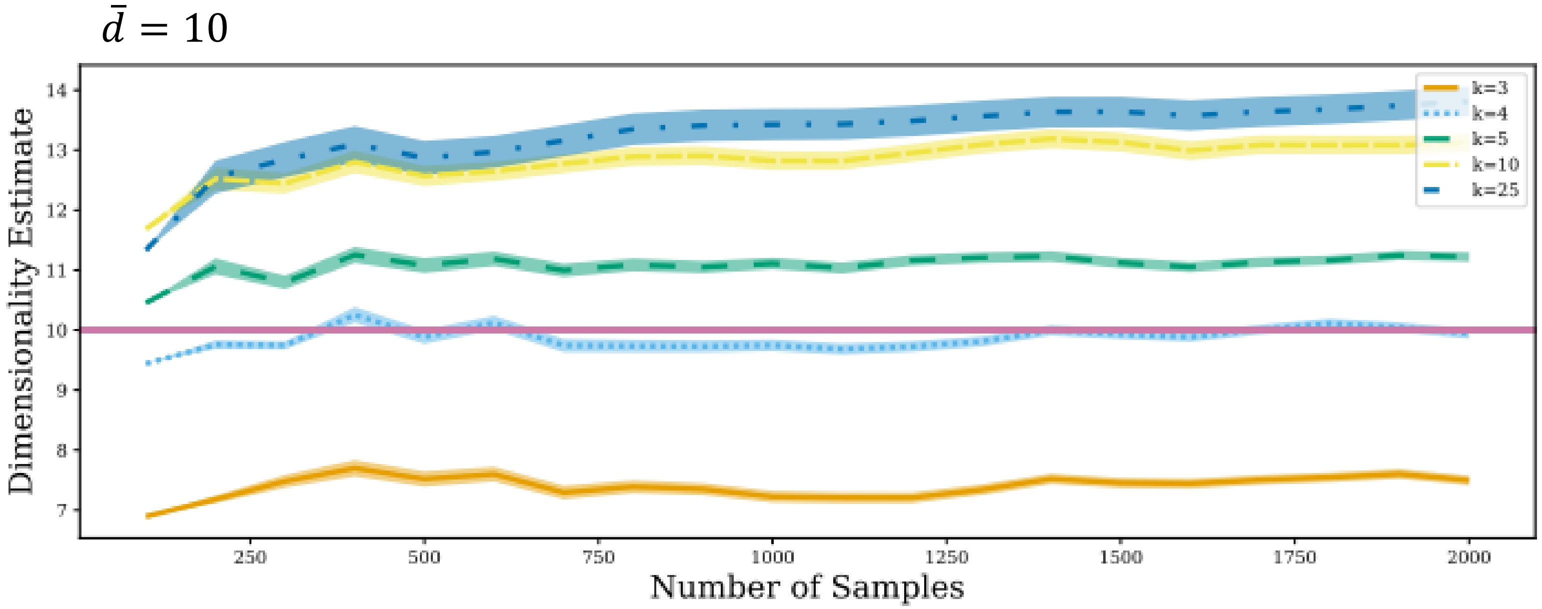
$$\frac{\partial L}{\partial m} = \left(\frac{1}{m} + \frac{V'(m)}{V(m)}\right) N(R) + \int_0^R \log t\, dN(t) - -e^{\theta}V(m)R^m\left(\log R + \frac{V'(m)}{V(m)}\right) = 0.$$

IDE Validation

We don' t know how good that estimator is since we never know the true intrinsic dimension of the dataset

→ Can upper-bound intrinsic dimensionality for GAN generated images using latent dimension degree of freedom.

BigGAN on 128x128 ImageNet (class= 'basenji') with 128-dim latent vectors, all zero except for $\bar{d}$ entries



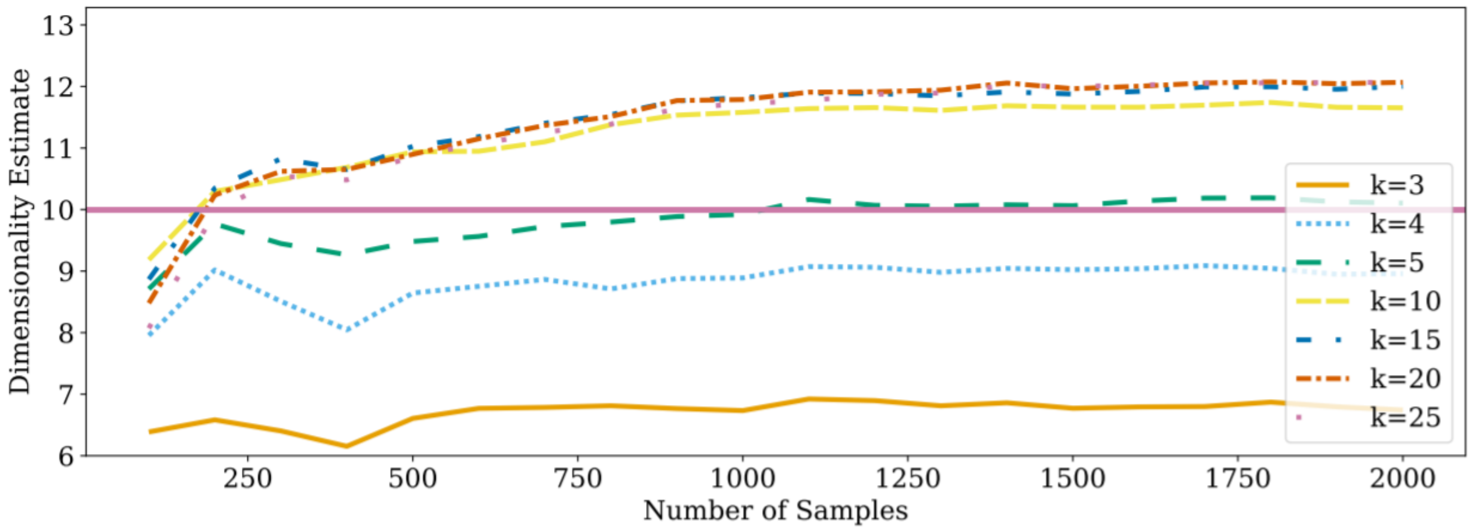
IDE Validation

Hyperparameter K affects the dimension estimate

→ Generally, high K leads to higher estimate

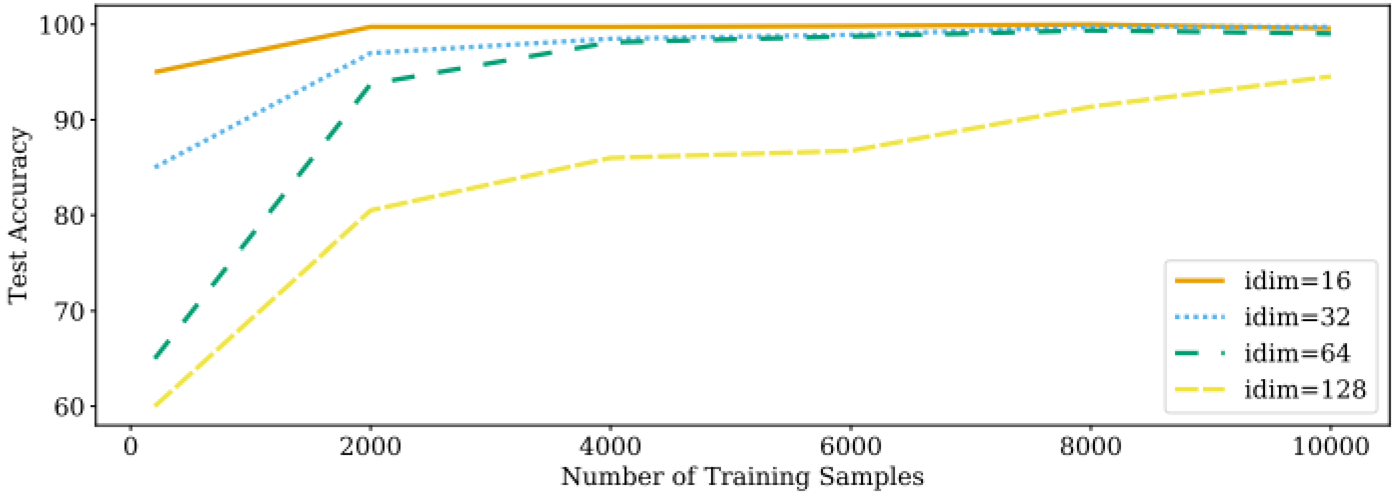
("Low K has high variance while high K has larger positive bias") → Report all values under different Ks

Class='daisy', $\bar{d} = 10$



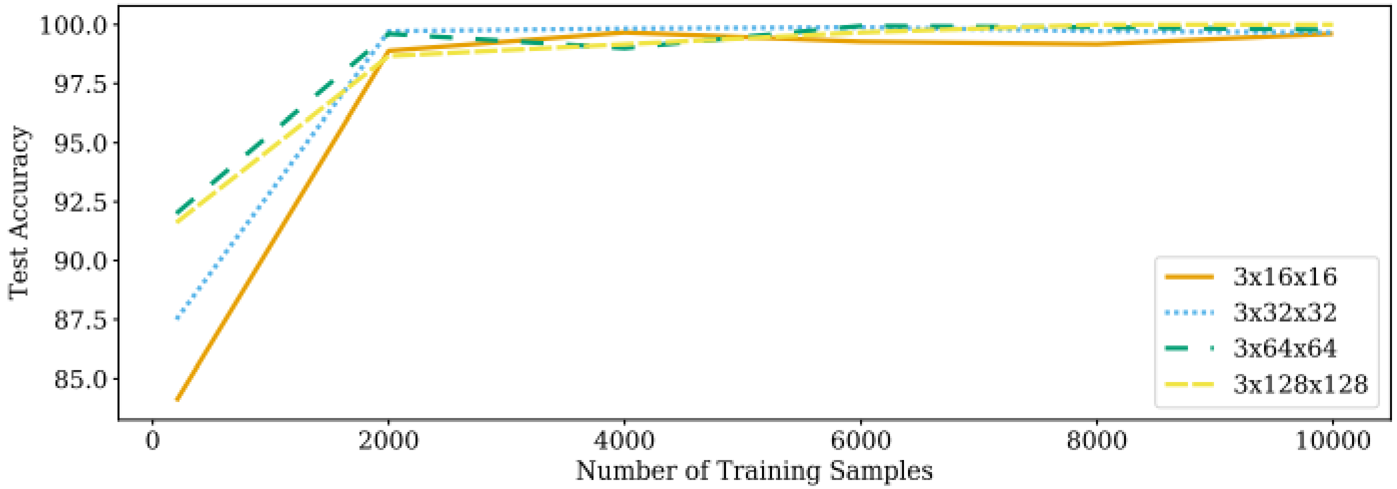
Generalization

- Construct synthetic data of class {basenji, beagle} with ResNet-18 architecture GAN
- Fix extrinsic resolution at 32x32 and vary intrinsic dim 16, 32, 64, 128
- Measure binary classification test accuracy against sample size ("Sample Complexity")



Generalization

- Construct synthetic data of class {basenji, beagle} with ResNet-18 architecture GAN
- Fix intrinsic dimension at 128 and vary extrinsic dimension by nearest neighbor interpolation → {16, 32, 64, 128, 256}



Generalization

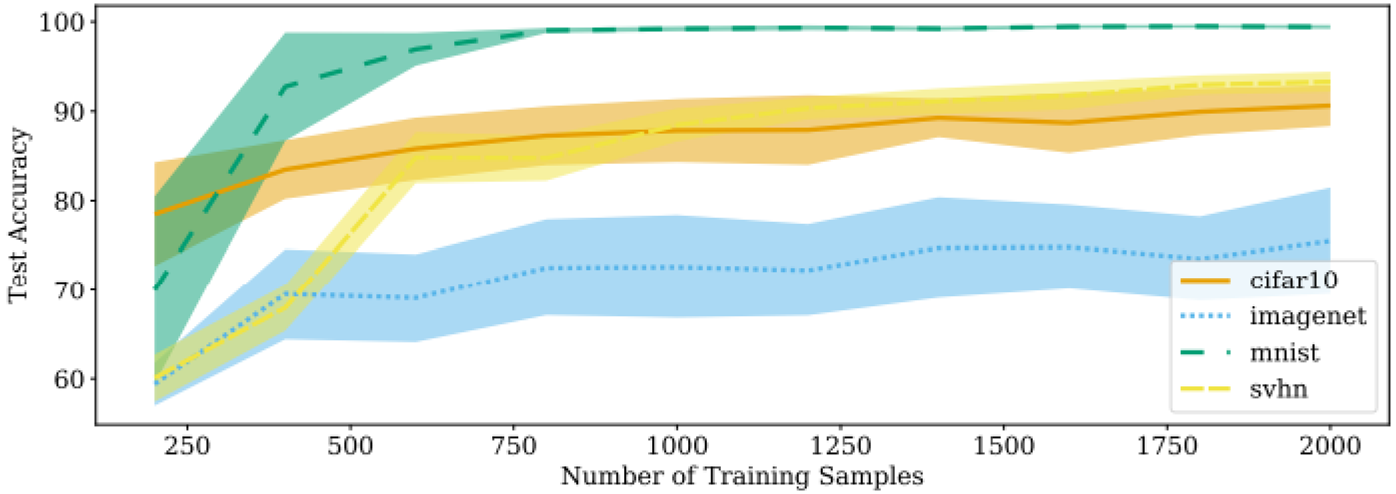
- Intrinsic Dimension Estimate for Real Data

Dataset	MNIST	SVHN	CIFAR-100	CelebA	CIFAR-10	MS-COCO	ImageNet
MLE (k=3)	7	9	11	9	13	22	26
MLE (k=5)	11	14	18	17	21	33	38
MLE (k=10)	12	18	22	24	25	37	43
MLE (k=20)	13	19	23	26	26	36	43
SOTA Accuracy	99.84	99.01	93.51	-	99.37	-	88.55

- Align image resolutions by resizing

	MNIST	SVHN	CIFAR-10	ImageNet
$k=3$	7.5 (0.2)	8.5 (0.1)	11.4 (0.2)	15.4 (0.8)
$k=4$	9.8 (0.3)	11.6 (0.1)	15.9 (0.2)	19.8 (0.9)
$k=5$	10.9 (0.4)	13.2 (0.1)	18.3 (0.3)	21.6 (1.0)

- Sample Complexity for real datasets

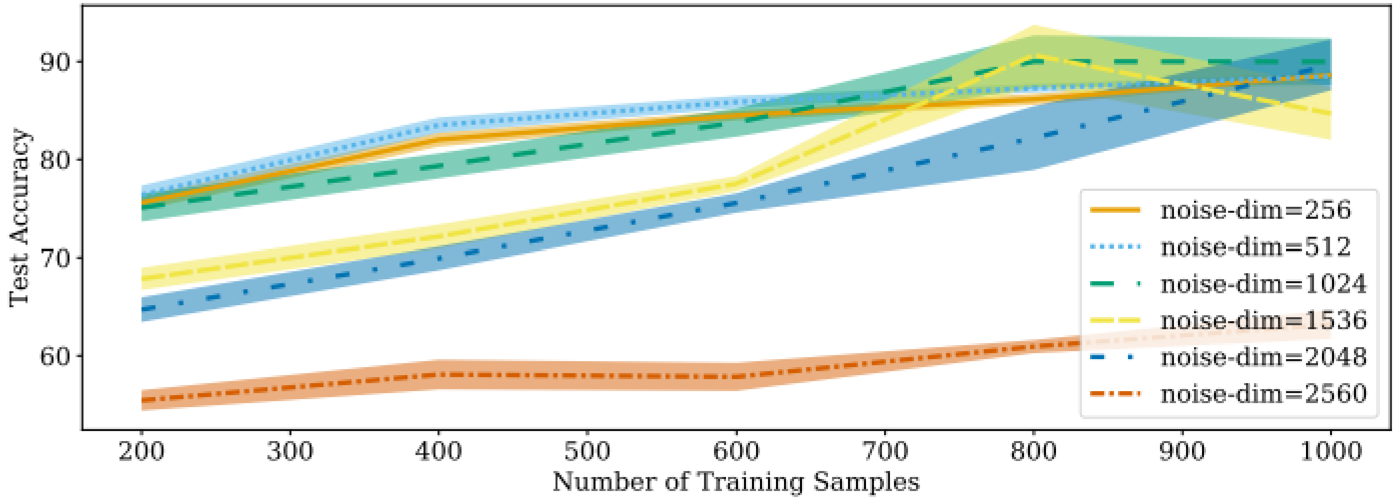


Additional Experiments

- Add uniformly sampled noise to CIFAR-10

	$\underline{d} = 256$	$\underline{d} = 512$	$\underline{d} = 1024$	$\underline{d} = 1536$	$\underline{d} = 2048$	$\underline{d} = 2560$
$k = 3$	19.7	30.9	57.1	77.8	110.0	136.1
$k = 4$	25.2	39.1	72.8	101.3	142.1	177.7
$k = 5$	27.6	42.5	78.3	110.2	153.4	196.6

- More noise leads to Higher sample complexity



Q&A / Comments