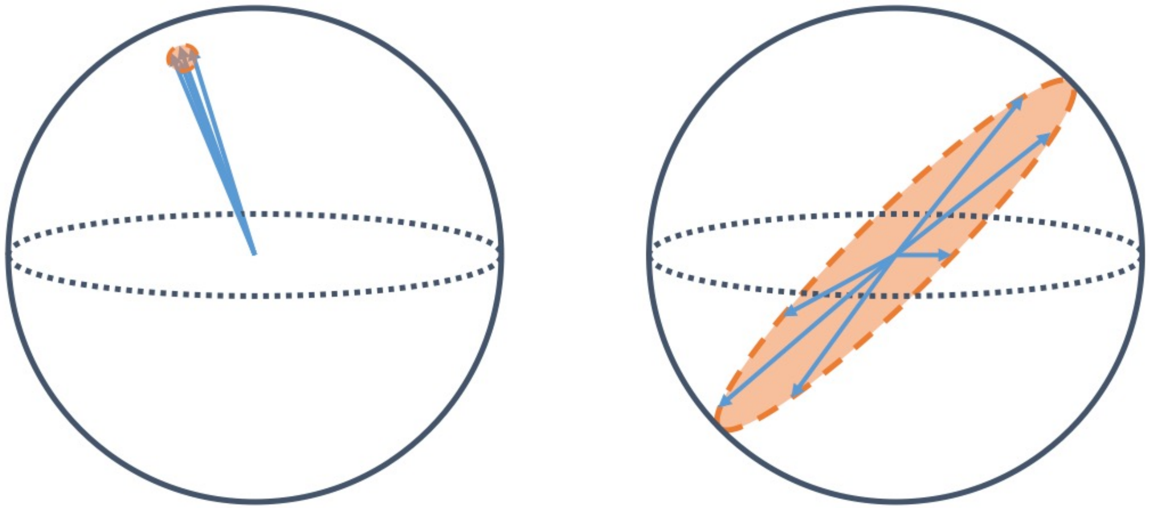
LDReg: Local Dimensionality Regularized Self-Supervised Learning

Hanxun Huang[1], Ricardo JGB Campello[2], Sarah Erfani[1], Xingjun Ma[3], Michael E Houle[4], James Bailey[1]

[1]The University of Melbourne, Australia
[2]University of Southern Denmark, Denmark
[3]Fudan University, China
[4]New Jersey Institute of Technology, USA

To appear in *International Conference on Learning Representations (ICLR) 2023*

Background: dimension collapse



(b) complete collapse (c) dimensional collapse

Jing, Li, et al. "Understanding Dimensional Collapse in Contrastive Self-supervised Learning." ICLR. 2022.

Background: dimension collapse

Representation dimension

- Total number of variables of the representation space
- E.g. 3D space of this room

Background: dimension collapse

Representation dimension

- Total number of variables of the representation space
- E.g. 3D space of this room

Intrinsic dimension

- Minimal number of variable to describe the data
- How many features (variables) are needed to describe where everyone is sitting in this room?

Background: dimension collapse

Contrastive:

- Sample contrastive: SimCLR, NNCLR, MOCO
- Asymmetrical model: BYOL, SimSiam

Generative:

- Masked image modeling: MAE

Suffer from dimension collapse

(Jing, Li, et al. 2022.)
(Zhang, Q., Wang, Y., & Wang, Y. 2022)

Background: dimension collapse

Contrastive:

- Sample contrastive: SimCLR, NNCLR, MOCO
- Asymmetrical model: BYOL, SimSiam

Generative:

- Masked image modeling: MAE

Alleviate global collapse

(Zhuo, Zhijian, et al. 2023)

Question: no more dimension collapse?

Question: no more dimension collapse?

In this room:

- How many features are needed to describe your position with the person sitting next to you?

Background: local intrinsic dimensionality (LID)

Let F be a real-valued function that is non-zero over some open interval containing $r \in \mathbb{R}$, $r \neq 0$.

Definition 1 ([33]). *The* intrinsic dimensionality of F at r *is defined as follows, whenever the limit exists:*

$$\mathrm{IntrDim}_F(r) \;\triangleq\; \lim_{\epsilon \to 0} \frac{\ln\left(F((1+\epsilon)r)/F(r)\right)}{\ln((1+\epsilon)r)/r)}\,.$$

Theorem 1 ([33]). *If F is continuously differentiable at r, then*

$$\mathrm{ID}_F(r) \triangleq \frac{r \cdot F'(r)}{F(r)} = \mathrm{IntrDim}_F(r)\,.$$

Houle, Michael E. "Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications." SISAP 2017.

Background: local intrinsic dimensionality (LID)

Theorem 2 (LID Representation Theorem [33]). *Let $F : \mathbb{R} \to \mathbb{R}$ be a real-valued function, and assume that ID_F^* exists. Let x and w be values for which x/w and $F(x)/F(w)$ are both positive. If F is non-zero and continuously differentiable everywhere in the interval $[\min\{x, w\}, \max\{x, w\}]$, then*

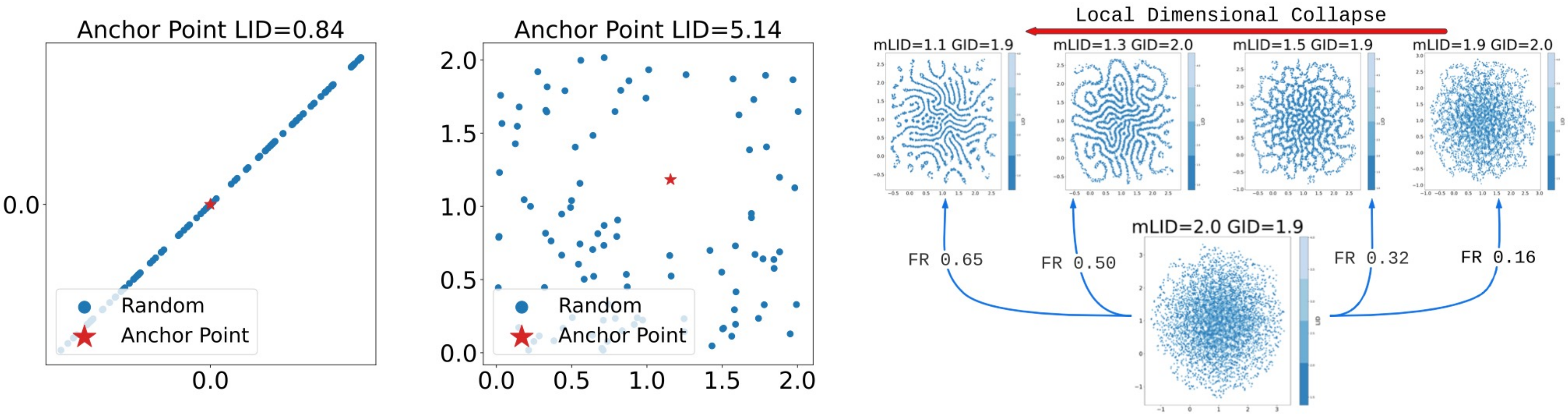
$$\frac{F(x)}{F(w)} = \left(\frac{x}{w}\right)^{\mathrm{ID}_F^*} \cdot A_F(x, w), \quad \textit{where} \quad A_F(x, w) \triangleq \exp\left(\int_x^w \frac{\mathrm{ID}_F^* - \mathrm{ID}_F(t)}{t} \, \mathrm{d}t\right),$$

whenever the integral exists.

- Space filling capability of the region surrounding an example
- Intrinsic dimensionality of the surroundings neighbors

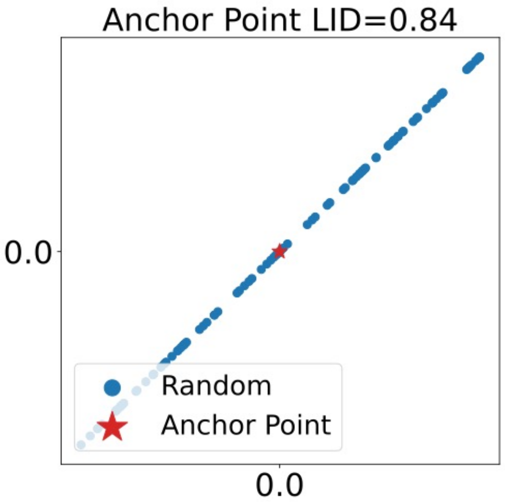
Houle, Michael E. "Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications." SISAP 2017.

Question: no more dimension collapse?

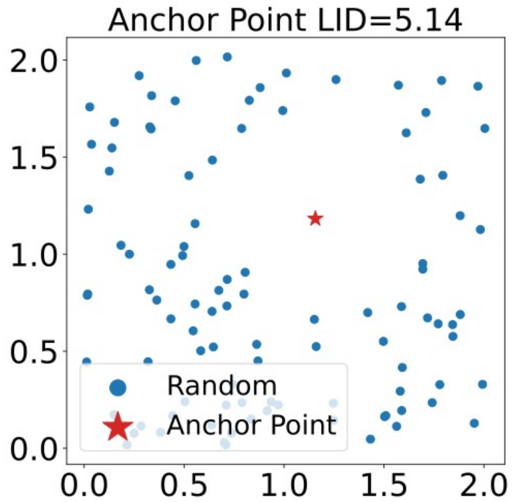


(a) Local collapse (b) No local collapse (c) Local dimensional collapse

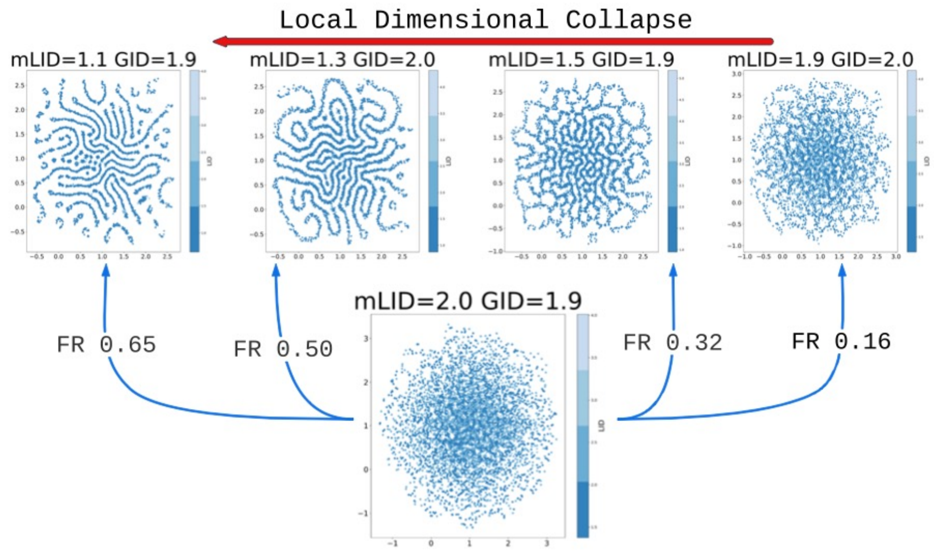
Question: no more dimension collapse?



(a) Local collapse



(b) No local collapse



(c) Local dimensional collapse

Local dimension collapse!

Question: how to regularize LID?

- Need a metric to measures the "distance" between LID distributions
- Fisher-Rao metric:

$$\mathcal{I}_w(\theta) = \int_0^w \left(\frac{\partial}{\partial \theta} \ln F'_w(x|\theta) \right)^2 F'_w(x|\theta) \, \mathrm{d}x \, .$$

Lemma 1. *Consider the family of distributions on* $[0, w]$ *parameterized by* θ*, whose CDFs are smooth growth functions of the form*

$$H_{w|\theta}(x) = \left(\frac{x}{w} \right)^{\theta} .$$

The Fisher-Rao distance d_{FR} *between* $H_{w|\theta_1}$ *and* $H_{w|\theta_2}$ *is*

$$d_{\mathrm{FR}}(H_{w|\theta_1}, H_{w|\theta_2}) = \left| \ln \frac{\theta_2}{\theta_1} \right| .$$

Local Dimensionality Regularization (LDReg)

Definition 3. *Given two smooth-growth distance distributions with CDFs F and G, their* asymptotic Fisher-Rao distance *is given by*

$$d_{\mathrm{AFR}}(F,G) \triangleq \lim_{w\to 0^+} d_{\mathrm{FR}}(H_{w|\,\mathrm{ID}_F^*}, H_{w|\,\mathrm{ID}_G^*}) = \left| \ln \frac{\mathrm{ID}_G^*}{\mathrm{ID}_F^*} \right| .$$

Local Dimensionality Regularization (LDReg)

Our dimensionality L_1 regularization for SSL corresponds to minimizing the negative log of the geometric mean of the ID values (recall Corollary 3.1). We assume that $\mathrm{ID}^*_{F^i_w}$ is desired to be ≥ 1.

$$\max \frac{1}{N} \sum_{i}^{N} \lim_{w \to 0} d_{\mathrm{AFR}}(F^i_w(r), U_{1,w}(r)) = \max \frac{1}{N} \sum_{i}^{N} \left| \ln \frac{\mathrm{ID}^*_{F^i_w}}{1} \right| = \min \left(-\frac{1}{N} \sum_{i}^{N} \ln \mathrm{ID}^*_{F^i_w} \right).$$

Make the learned representation away from uniform distance distribution (LID = 1)

Experiments

Representation quality is evaluated with linear probing on frozen encoder

Dataset: ImageNet-1k

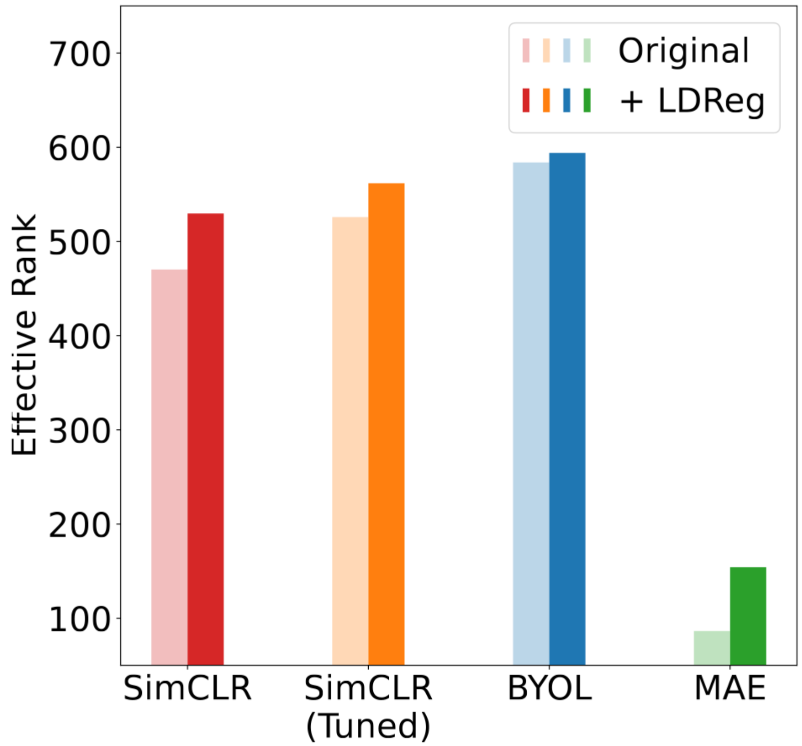
Contrastive:

- Sample contrastive: SimCLR, SimCLR-Tuned
- Asymmetrical model: BYOL

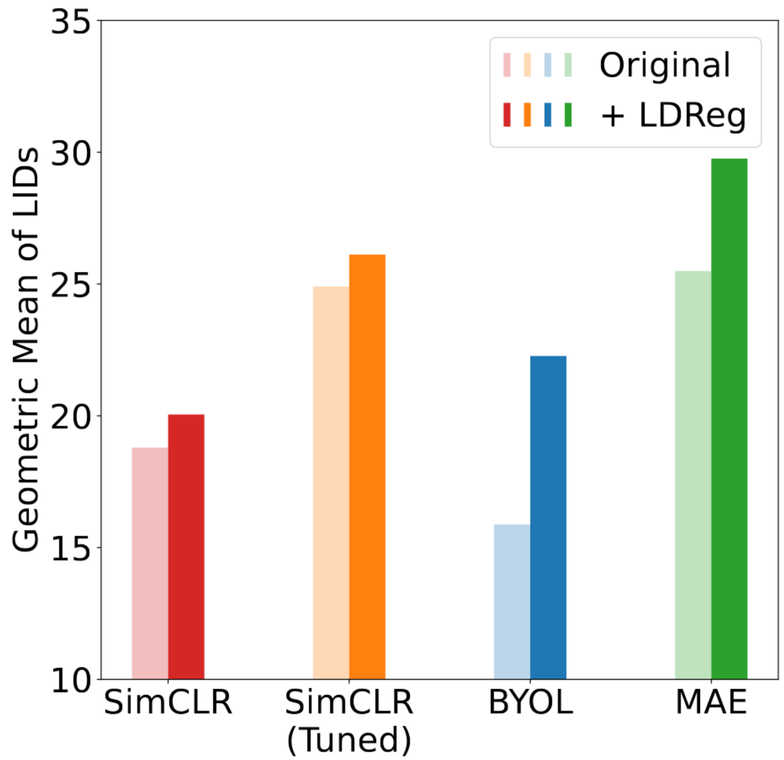
Generative:

- Masked image modeling: MAE

Increases both local and global intrinsic dimensions



(c) Effective rank



(d) Geometric mean of LID

Evaluations

Table 1: The linear evaluation results (accuracy (%)) of different methods with and without LDReg. The effective rank is calculated on the ImageNet validation set. The best results are **boldfaced**.

Model	Epochs	Method	Regularization	Linear Evaluation	Effective Rank
ResNet-50	100	SimCLR	-	64.3	470.2
			LDReg	**64.8**	529.6
		SimCLR (Tuned)	-	67.2	525.8
			LDReg	**67.5**	561.7
		BYOL	-	67.6	583.8
			LDReg	**68.5**	594.0
ViT-B	200	SimCLR	-	72.9	283.7
			LDReg	**73.0**	326.1
		MAE	-	57.0	86.4
			LDReg	**57.6**	154.1

Consistent improvement on existing SSL methods

Evaluations

Table 2: The transfer learning results in terms of linear probing accuracy (%), using ResNet-50 as the encoder. The best results are **boldfaced**.

Method	Regularization	Batch Size	Epochs	ImageNet	Food-101	CIFAR-10	CIFAR-100	Birdsnap	Cars	DTD
SimCLR	-	2048	100	64.3	69.0	89.1	**71.2**	32.0	36.7	**67.8**
	LDReg			**64.8**	**69.1**	**89.2**	70.6	**33.4**	**37.3**	67.7
	-	4096	1000	69.0	71.1	90.1	71.6	37.5	35.3	70.7
	LDReg			**69.8**	**73.3**	**91.8**	**75.1**	**38.7**	**41.6**	**70.8**

Consistent improvement on transfer learning

Evaluations

Table 3: The performance of the pre-trained models (ResNet-50) on object detection and instance segmentation tasks, when fine-tuned on COCO. The bounding-box (AP^{bb}) and mask (AP^{mk}) average precision are reported with the best results are **boldfaced**.

Method	Regularization	Epochs	Batch Size	Object Detection			Segmentation		
				AP^{bb}	AP^{bb}_{50}	AP^{bb}_{75}	AP^{mk}	AP^{mk}_{50}	AP^{mk}_{75}
SimCLR	-	100	2048	35.24	55.05	**37.88**	31.30	51.70	32.82
	LDReg			**35.26**	**55.10**	37.78	**31.38**	**51.88**	**32.90**
BYOL	-			36.30	55.64	38.82	32.17	52.53	34.30
	LDReg			**36.82**	**56.47**	**39.62**	**32.47**	**53.15**	**34.60**
SimCLR	-	1000	4096	36.48	56.22	39.28	32.12	52.70	34.02
	LDReg			**37.15**	**57.20**	**39.82**	**32.82**	**53.81**	**34.74**

Consistent improvement on finetuning tasks

Local collapse triggering complete collapse

Min LID: using Fisher-Rao metric, making LID lower

$$\min\left(\frac{1}{N}\sum_{i}^{N}\ln \mathrm{LID}^{*}_{F_w^i}\right).$$

Table 13: Comparing the results of linear evaluations of regularization terms of LDReg, MinLID and baseline. All models are trained on ImageNet for 100 epochs using ResNet-50 as encoder. The results are reported as linear probing accuracy (%) on ImageNet.

Method	Regularization	β	Linear Acc	Effective Rank
SimCLR	LDReg	0.01	64.8	529.6
	-	-	64.3	470.2
	Min LID	0.01	64.2	150.7
	Min LID	0.1	63.1	15.0
	Min LID	1.0	46.4	1.0
	Min LID	10.0	Complete collapse	-

Contributions

- Mitigating dimensional collapse in SSL through LID.
- Theory to support the formulation of LID regularization.
- Consistent empirical improvement on standard SSL evaluations

Reference

[1] Jing, Li, et al. "Understanding Dimensional Collapse in Contrastive Self-supervised Learning." ICLR. 2022.

[2] Zhang, Qi, Yifei Wang, and Yisen Wang. "How mask matters: Towards theoretical understandings of masked autoencoders." NeurIPS. 2022.

[3] Zhuo, Zhijian, et al. "Towards a Unified Theoretical Understanding of Non-contrastive Learning via Rank Differential Mechanism." ICLR. 2023.

[4] Houle, Michael E. "Local intrinsic dimensionality I: an extreme-value-theoretic foundation for similarity applications." SISAP 2017.

Reference

Code available on GitHub

https://github.com/HanxunH/LDReg

Thank you