Manifold Learning Benefits GANs

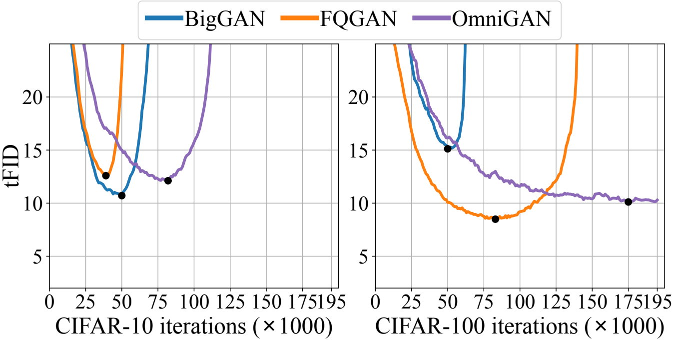
Yao Ni[*,†] Piotr Koniusz[*,§,†] Richard Hartley[†,♦] Richard Nock[♦,†]

[†]The Australian National University
[§]Data61/CSIRO
[♦]Google Research

CVPR 2022

Background: discriminator overfitting

- Training instability. (Kodali *et al.*, 2017)
- Mode collapse. (Roth *et al.*, 2017)
- **Discriminator overfitting**. (Webster *et al.*, 2019)

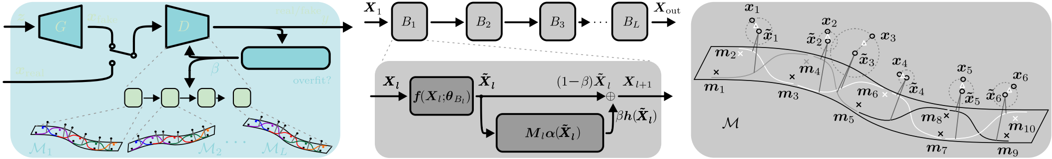


(a) Training Curves

(b) DiffAug

(c) ADA

Pipeline: main components



Dictionary Learning step: $, \boldsymbol{M} \quad \arg\min_{',\boldsymbol{M}'} ½\boldsymbol{X}-\boldsymbol{M}' \quad '½_F^2 + \quad ', \boldsymbol{M}', \boldsymbol{X}$

Intertwining step: reconstruct then intertwine

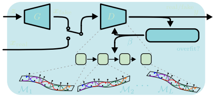
$$\boldsymbol{h}_l \; \tilde{\boldsymbol{X}}_l \quad \boldsymbol{M}_l \quad \tilde{\boldsymbol{X}}_l \quad \boldsymbol{X}_{l+1} \quad 1- \quad \tilde{\boldsymbol{X}}_l + \; \boldsymbol{h}_l \; \tilde{\boldsymbol{X}}_l$$

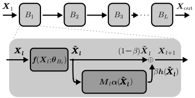
Complexity constrain: output of blocks the reconstruction

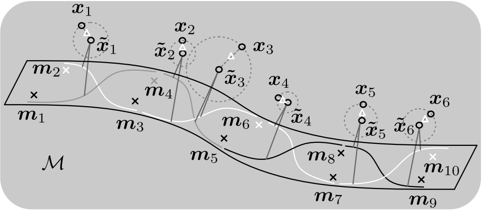
$$J_{\text{prox}} \quad \overline{L} = ½\tilde{\boldsymbol{X}}_l - \boldsymbol{h}_l \; \tilde{\boldsymbol{X}}_l \; ½_F^2$$

Meta-controller: control the intertwining step and complexity constrain by observing $r \; \boldsymbol{x}_{\text{real}} \quad \mathbb{E} \; \text{Sign} \; D \; \boldsymbol{x}_{\text{real}}$

$$_{t+1} \quad _{t} + \quad \cdot \text{Sign} \; r \; \boldsymbol{x}_{\text{real}} \; \% \quad _{t+1} \quad _{0} + \quad \cdot \quad _{t+1}$$







Coder: $\boldsymbol{x}$ arg min ½$\boldsymbol{x} - \boldsymbol{M}$ '½$_2^2$
'

HA: ' " r0, 1x^k s.t. ½ '½$_1$ 1 **LLC:** s.t. $\mathbf{1}^T$ ' 1

SA: $\boldsymbol{x}$; $\boldsymbol{M}$, S ½$\boldsymbol{x} - \boldsymbol{m}_1$½$_2$, ⋯, ½$\boldsymbol{x} - \boldsymbol{m}_k$½$_2$

S $d_1, :::, d_k$ $_j$ exp $-d_j^2$©2 ^{2} © <$_i$ exp $-d_i^2$©2 ^{2}

LCSA: differs from SA by setting $_i$ $\boldsymbol{x}$ 0 unless $\boldsymbol{m}_i$ is among the k' nearest-neighbor atoms

Theory: LCSA extends useful properties from other coders

LCSA acts as a trade-off between other coders

- **LCSA may act as HA (extreme way to guide f)**, when $\quad 0$ or $k^{l} \quad 1$.
- **LCSA may act as LLC (weak way to guide f)**.
- **Large dictionary makes LCSA a 'free learner' as f (easy to overfit)**

LCSA is locally-adaptive denoiser the Local Lipschitz constant $K \quad D^{2}© \quad ^{2}$ and D is the diameter of simplex $\quad$.

LCSA denoises by DAE-like mechanism

$$L_{\text{dae}} \quad \frac{1}{N} \sum_{n\ 1}^{N} \left\| r\ x_n - x_n \right\|_2^2 + l2 \left\| \frac{\partial r\ x}{\partial x}\Big|_{x\ x_n} \right\|_2^2$$

where $r\ x_n$ is the reconstruction of x_n akin to our $h\ x_n$

Experiment: Comparison with STOA

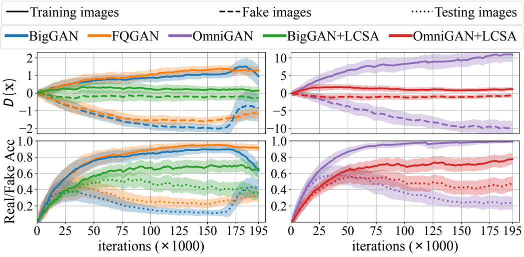
Model (d' 1024)	CIFAR-10			CIFAR-100			ImageNet (64×64)		
	IS	tFID	vFID	IS	tFID	vFID	IS	tFID	vFID
BigGAN	9.64	10.71	14.86	11.54	15.11	19.97	**34.50**	8.96	8.80
FQGAN	9.44	12.59	16.69	**11.74**	8.49	13.51	33.14	8.27	8.15
BigGAN+LCSA	**9.81**	**3.51**	**7.55**	11.60	**5.49**	**10.37**	33.29	**5.99**	**5.79**
OmniGAN	9.92	12.11	16.29	12.42	10.11	14.85	70.59	7.09	7.66
OmniGAN+LCSA	**10.21**	**2.94**	**6.98**	**13.88**	**4.97**	**9.72**	**77.45**	**4.26**	**4.94**

ImageNet (128×128)			
Model	IS	tFID	vFID
BigGAN	104.57	9.19	9.18
OmniGAN	169.13	7.11	7.30
OmniGAN+LCSA	**192.45**	**4.83**	**5.24**

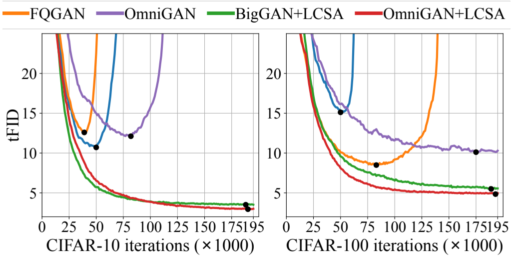
Oxford-102 Flowers (256×256)	FID
MSG-StyleGAN	18.59
MSG-StyleGAN+LCSA	**13.13**

FFHQ (256×256) FID		
Model	70k	140k
StyleGAN2	5.28	3.71
StyleGAN2+ADA	4.30	3.81
StyleGAN2+LeCam	-	3.66
StyleGAN2+bCR	3.79	3.53
StyleGAN2+LCSA	3.83	3.32
StyleGAN2+LCSA+bCR	**3.74**	**3.20**

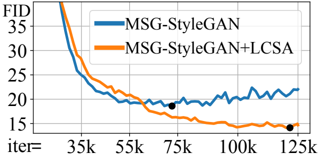
Experiment: Avoid overfitting



(a) Avoiding overfitting

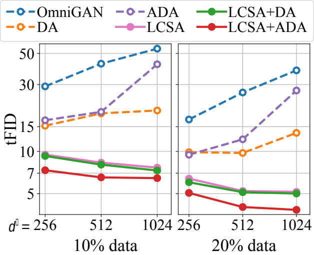


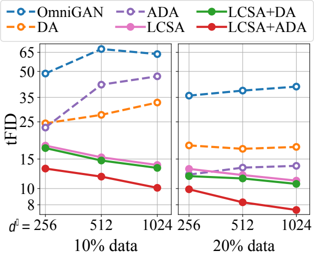
(b) Preventing diverge



(c) Continuing Learning

Experiment: Comparison with data augmentation

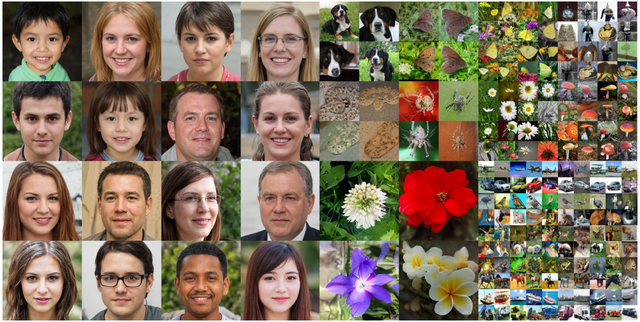




CIFAR-10 CIRAR-100

DA (leak cutout) DA+LCSA ADA (leak rotation) ADA+LCSA

Generated images & conclusions



Conclusions:

- Discriminator overfitting constrains discriminator complexity
- Manifold learning step & intertwining step & meta-controller
- LCSA extends useful properties from other coders stability & preventing overfitting, LCSA improves STOA baselines & avoid overfitting/leaking augmentation clues.