Augmentation-Aware Self-Supervision for Data-Efficient GAN Training

Liang Hou[1,2,3] Qi Cao[1] Yige Yuan[1,2] Songtao Zhao[3] Chongyang Ma[3]
Siyuan Pan[3] Pengfei Wan[3] Zhongyuan Wang[3] Huawei Shen[1,2] Xueqi Cheng[1,2]

[1]Institute of Computing Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences
[3]Kuaishou Technology

Neural Information Processing Systems, 2023

GANs under Limited Data

Generative adversarial networks (GANs) struggle to produce diverse and high-quality samples under limited training data because the discriminator is prone to over-fitting.

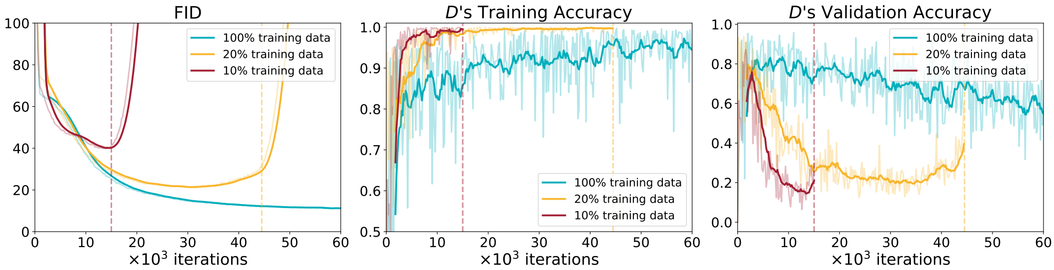


Figure: Image credit: DiffAugment (Shengyu Zhao, et al. NeurIPS 2020)

Differentiable Augmentation for GANs

color: $\boldsymbol{\omega}_{\text{color}} = (\lambda_{\text{brightness}}, \lambda_{\text{saturation}}, \lambda_{\text{contrast}}) \in \mathbb{R}^3$
translation: $\boldsymbol{\omega}_{\text{translation}} = (x_{\text{translation}}, y_{\text{translation}}) \in \mathbb{R}^2$
cutout: $\boldsymbol{\omega}_{\text{cutout}} = (x_{\text{offset}}, y_{\text{offset}}) \in \mathbb{R}^2$

Original Data $\boldsymbol{\omega}_{\text{color}} = (0.2, 0.3, 0.5)$ $\boldsymbol{\omega}_{\text{translation}} = (0.1, 0.8)$ $\boldsymbol{\omega}_{\text{cutout}} = (0.4, 0.6)$

$$\mathcal{L}_D^{\text{da}} = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x}), \boldsymbol{\omega}\sim p(\boldsymbol{\omega})}[f(D(T(\mathbf{x}; \boldsymbol{\omega})))] + \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z}), \boldsymbol{\omega}\sim p(\boldsymbol{\omega})}[h(D(T(G(\mathbf{z}); \boldsymbol{\omega})))]$$
$$\mathcal{L}_G^{\text{da}} = \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z}), \boldsymbol{\omega}\sim p(\boldsymbol{\omega})}[g(D(T(G(\mathbf{z}); \boldsymbol{\omega})))]$$

Invariance to Augmentations

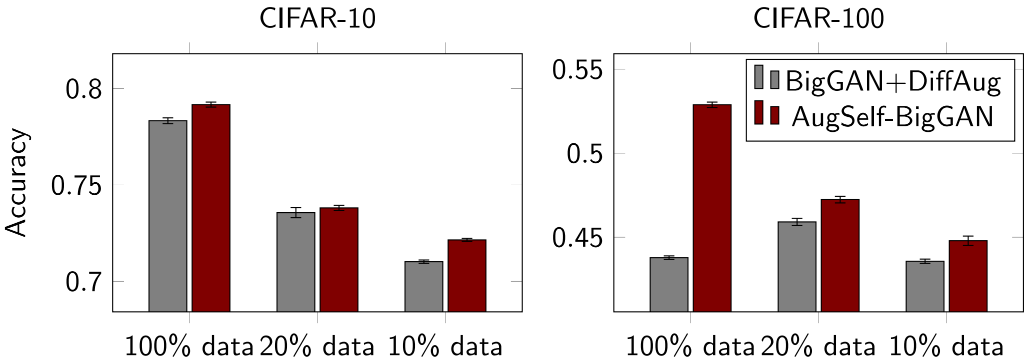
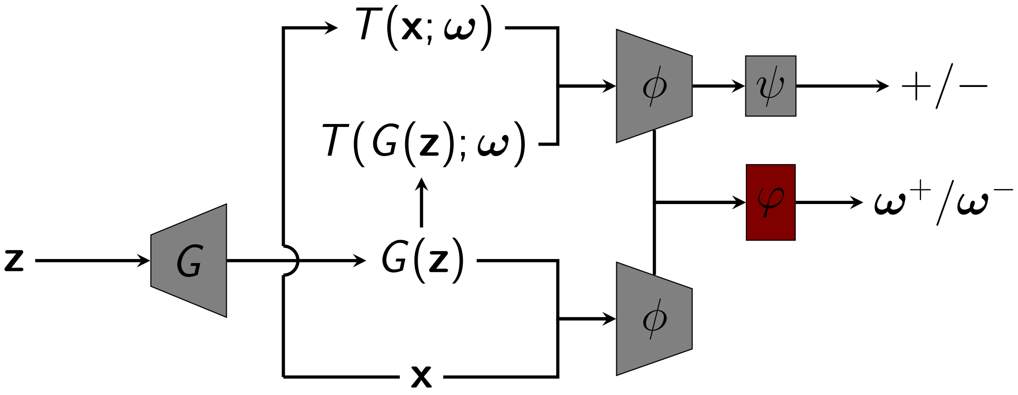


Figure: Comparison of representation learning ability of discriminator between BigGAN + DiffAugment and AugSelf-BigGAN on CIFAR-10 and CIFAR-100 using linear logistic regression.

Proposed Method

The original discriminator: $D(\hat{\mathbf{x}}) = \psi(\phi(\hat{\mathbf{x}})) : \hat{\mathcal{X}} \to \{+, -\}$
The self-supervised discriminator: $\hat{D}(\hat{\mathbf{x}}, \mathbf{x}) = \varphi(\phi(\hat{\mathbf{x}}) - \phi(\mathbf{x})) : \hat{\mathcal{X}} \times \mathcal{X} \to \Omega \times \{+, -\}$



Loss Functions

Self-Supervised Discriminator: predicts the augmentation and authenticity of data
Generator: generates augmentation-predictable real rather than not fake data

$$\mathcal{L}_{\hat{D}}^{\mathrm{ss}} = \mathbb{E}_{\mathbf{x},\boldsymbol{\omega}} \left[\| \hat{D}(T(\mathbf{x};\boldsymbol{\omega}), \mathbf{x}) - \boldsymbol{\omega}^{+} \|_2^2 \right] + \mathbb{E}_{\mathbf{z},\boldsymbol{\omega}} \left[\| \hat{D}(T(G(\mathbf{z});\boldsymbol{\omega}), G(\mathbf{z})) - \boldsymbol{\omega}^{-} \|_2^2 \right]$$
$$\mathcal{L}_{G}^{\mathrm{ss}} = \mathbb{E}_{\mathbf{z},\boldsymbol{\omega}} \left[\| \hat{D}(T(G(\mathbf{z});\boldsymbol{\omega}), G(\mathbf{z})) - \boldsymbol{\omega}^{+} \|_2^2 \right] - \mathbb{E}_{\mathbf{z},\boldsymbol{\omega}} \left[\| \hat{D}(T(G(\mathbf{z});\boldsymbol{\omega}), G(\mathbf{z})) - \boldsymbol{\omega}^{-} \|_2^2 \right]$$

Total loss functions of our proposed AugSelf-GAN are

$$\min_{D,\hat{D}} \mathcal{L}_{D}^{\mathrm{da}} + \lambda_d \cdot \mathcal{L}_{\hat{D}}^{\mathrm{ss}}$$
$$\min_{G} \mathcal{L}_{G}^{\mathrm{da}} + \lambda_g \cdot \mathcal{L}_{G}^{\mathrm{ss}}$$

where $\lambda_d \in \mathbb{R}^{+}$ and $\lambda_g \in \mathbb{R}^{+}$ are two hyper-parameters.

Analysis

Theorem

Assume that $\boldsymbol{\omega}^{+} = -\boldsymbol{\omega}^{-} = \boldsymbol{c}$ *is constant, under the optimal self-supervised discriminator, optimizing the self-supervised task for the generator is equivalent to*

$$\min_G 4c \cdot M_{\mathrm{AH}}(p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}}) \| p_G(\mathbf{x}, \hat{\mathbf{x}}))$$

where $c = \|\boldsymbol{c}\|_2^2$ *is constant and* M_{AH} *is the AHM divergence*[1]*, of which the minimum is achieved if and only if* $p_G(\mathbf{x}, \hat{\mathbf{x}}) = p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}}) \Rightarrow p_G(\mathbf{x}) = p_{\mathrm{data}}(\mathbf{x})$*.*

Corollary

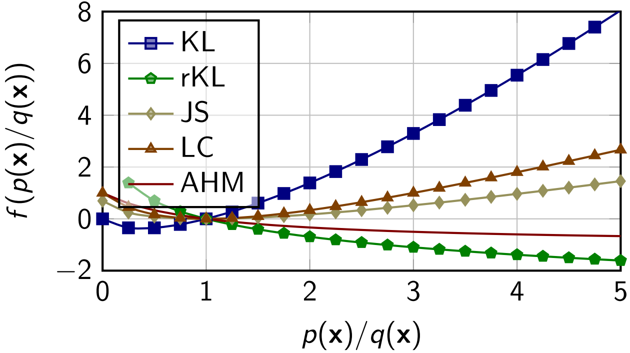
$$M_{\mathrm{AH}}(p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}}) \| p_G(\mathbf{x}, \hat{\mathbf{x}})) + M_{\mathrm{AH}}(p_G(\mathbf{x}, \hat{\mathbf{x}}) \| p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}})) = \Delta(p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}}) \| p_G(\mathbf{x}, \hat{\mathbf{x}}))$$
$$M_{\mathrm{AH}}(p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}}) \| p_G(\mathbf{x}, \hat{\mathbf{x}})) = 1 - W(p_{\mathrm{data}}(\mathbf{x}, \hat{\mathbf{x}}) \| p_G(\mathbf{x}, \hat{\mathbf{x}})) \leq 1$$

[1] M_{AH}: arithmetic−harmonic mean divergence; Δ: Le Cam divergence; W: harmonic mean divergence

Analysis: Robustness of f-function of f-divergence

Definition (f-divergence)

The f-divergence between two probability distributions $p(\mathbf{x})$ and $q(\mathbf{x})$ is defined as $D_f(p(\mathbf{x}) \| q(\mathbf{x})) = \int_{\mathcal{X}} q(\mathbf{x}) f(p(\mathbf{x})/q(\mathbf{x})) \mathrm{d}\mathbf{x}$ with a convex function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ satisfying $f(1) = 0$.



The function f of the AHM divergence yields more robust value than other common f-divergences (e.g., JS and LC) for large inputs $p(\mathbf{x})/q(\mathbf{x})$, resulting in more stable gradient feedback for updating the generator.

Results: CIFAR-10 and CIFAR-100

	Method	100% training data		20% training data		10% training data	
		IS (↑)	FID (↓)	IS (↑)	FID (↓)	IS (↑)	FID (↓)
CIFAR-10	BigGAN	9.07	9.59	8.52	21.58	7.09	39.78
	DiffAugment	9.16	8.70	8.65	14.04	8.09	22.40
	CR-GAN	9.17	8.49	8.61	12.84	8.49	18.70
	LeCam-GAN	**9.43**	8.28	8.83	12.56	8.57	17.68
	DigGAN	<u>9.28</u>	8.49	8.89	13.01	8.32	17.87
	MaskedGAN	-	8.41$\pm$.03	-	12.51$\pm$.09	-	15.89$\pm$.12
	GenCo	-	7.98$\pm$.02	-	12.61$\pm$.05	-	18.10$\pm$.06
	AugSelf-BigGAN	**9.43**$\pm$.14	<u>7.68</u>$\pm$.06	<u>8.98</u>$\pm$.09	<u>10.97</u>$\pm$.09	<u>8.76</u>$\pm$.05	<u>15.68</u>$\pm$.26
	AugSelf-BigGAN+	9.27$\pm$.05	**7.54**$\pm$.04	**9.08**$\pm$.04	**9.95**$\pm$.17	**8.79**$\pm$.04	**12.76**$\pm$.14
CIFAR-100	BigGAN	10.71	12.87	8.58	33.11	6.74	66.71
	DiffAugment	10.66	12.00	9.47	22.14	8.38	33.70
	CR-GAN	10.81	11.25	9.12	20.28	8.70	26.90
	LeCam-GAN	11.05	11.20	9.81	18.03	9.27	27.63
	DigGAN	**11.45**	11.63	9.54	19.79	8.98	24.59
	MaskedGAN	-	11.65$\pm$.03	-	18.33$\pm$.09	-	24.02$\pm$.12
	GenCo	-	10.92$\pm$.02	-	18.44$\pm$.04	-	25.22$\pm$.06
	AugSelf-BigGAN	<u>11.19</u>$\pm$.09	**9.88**$\pm$.07	**10.25**$\pm$.06	<u>16.11</u>$\pm$.25	<u>9.78</u>$\pm$.08	<u>21.30</u>$\pm$.15
	AugSelf-BigGAN+	11.12$\pm$.10	<u>10.09</u>$\pm$.05	<u>10.14</u>$\pm$.11	**15.33**$\pm$.20	**9.93**$\pm$.06	**18.64**$\pm$.09

Results: FFHQ and LSUN-Cat, and AFHQ

Method	FFHQ				LSUN-Cat			
	30K	10K	5K	1K	30K	10K	5K	1K
StyleGAN2	6.16	14.75	26.60	62.16	10.12	17.93	34.69	182.85
+ ADA	5.46	8.13	10.96	21.29	10.50	13.13	16.95	43.25
+ DiffAugment	5.05	7.86	10.45	25.66	9.68	12.07	16.11	42.26
AugSelf-StyleGAN2	**4.95**	6.98	9.69	23.38	**9.22**	**11.98**	14.86	36.76
AugSelf-StyleGAN2+	5.82	**6.65**	**9.15**	**20.39**	9.43	12.00	**14.12**	**26.52**

Method	Cat	Dog	Wild
StyleGAN2	5.13	19.4	3.48
+ DiffAugment	3.49	8.75	2.69
AugSelf-StyleGAN2	**3.23** ($\downarrow$ 7.45%)	**8.17** ($\downarrow$ 6.63%)	**2.48** ($\downarrow$ 7.81%)

Results: Low-shot Image Generation

Method	Obama	Grumpy cat	Panda	Cat	Dog
StyleGAN2	80.20	48.90	34.27	71.71	130.19
+ AdvAug	52.86	31.02	14.75	47.40	68.28
+ ADA	45.69	26.62	12.90	40.77	56.83
+ APA	42.97	28.10	19.21	42.60	81.16
+ DiffAugment	46.87	27.08	12.06	42.44	58.85
AugSelf-StyleGAN2	**26.00**	**19.81**	**8.36**	**30.53**	**48.19**

Conclusion

We propose AugSelf-GAN, a novel data-efficient GAN training method, by utilizing the adversarial self-supervision from differentiable augmentation.

We analyze a connection between the theoretical learning objective of AugSelf-GAN and the AHM divergence, which can provide robust gradient feedback to update the generator under limited data.

Thank you for your attention!
Please check out our paper and code for more details.
Paper: https://arxiv.org/abs/2205.15677
Code: https://github.com/liang-hou/augself-gan