AdaptiveMix: Improving GAN Training via Feature Space Shrinkage

Haozhe Liu [#1], Wentian Zhang [#2], Bing Li [*1], Haoqian Wu [3],
Nanjun He [2], Yawen Huang [2], Yuexiang Li [*2], Bernard Ghanem [1], Yefeng Zheng [2]

Joint First Author * Corresponding Author

[1] AI Initiative, King Abdullah University of Science and Technology (KAUST)
[2] Jarvis Lab, Tencent, [3] YouTu Lab, Tencent

Outline

AdaptiveMix

1. Generate Hard Samples

Given x_i, x_j and Feature Extractor $\mathcal{F}(\cdot)$

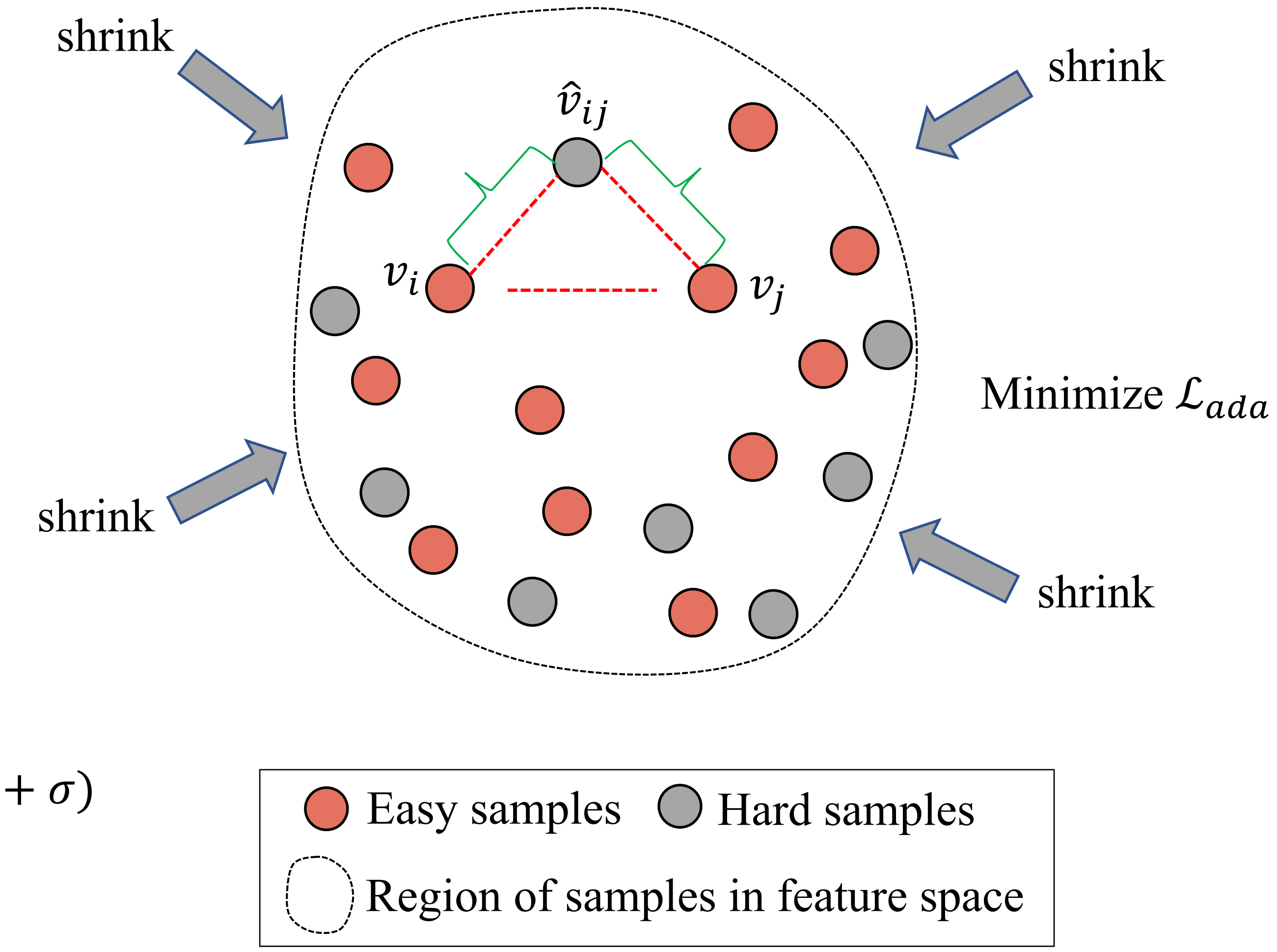
$$\hat{x}_{ij} = \lambda x_i + (1 - \lambda) x_j$$

$$v_i = \mathcal{F}(x_i) \quad v_j = \mathcal{F}(x_j) \quad \hat{v}_{ij} = \mathcal{F}(\hat{x}_{ij})$$

2. AdaptiveMix Loss

$$\mathcal{L}_{ada} = \sum_i \sum_j \mathbb{D}_v(\lambda \mathcal{F}(x_i) + (1 - \lambda)\mathcal{F}(x_j), \mathcal{F}(\hat{x}_{ij}) + \sigma)$$

$\mathbb{D}_v(\cdot)$ computes the $L_1(L_2)$ distance



AdaptiveMix on Image Generation

The **learning objective** of AdaptiveMix based image generation:

$$\min_{G} \max_{\mathcal{F},\mathcal{J}} \mathop{\mathbb{E}}_{x\sim p_r}[\mathcal{J}(\mathcal{F}(x))] - \mathop{\mathbb{E}}_{z\sim p_z}[\mathcal{J}(\mathcal{F}(G(z)))] + \min_{\mathcal{F}} \mathop{\mathbb{E}}_{x\sim p_r,p_g}[\mathcal{L}_{ada}]$$

AdaptiveMix Loss

$$\mathcal{L}_{ada} = \sum_i \sum_j \mathbb{D}_v(\lambda\mathcal{F}(x_i) + (1-\lambda)\mathcal{F}(x_j), \mathcal{F}(\hat{x}_{ij}) + \sigma),$$

Discriminator is consisting of feature extractor $\mathcal{F}(\cdot)$ and classifier head $\mathcal{J}(\cdot)$

AdaptiveMix on Image Generation

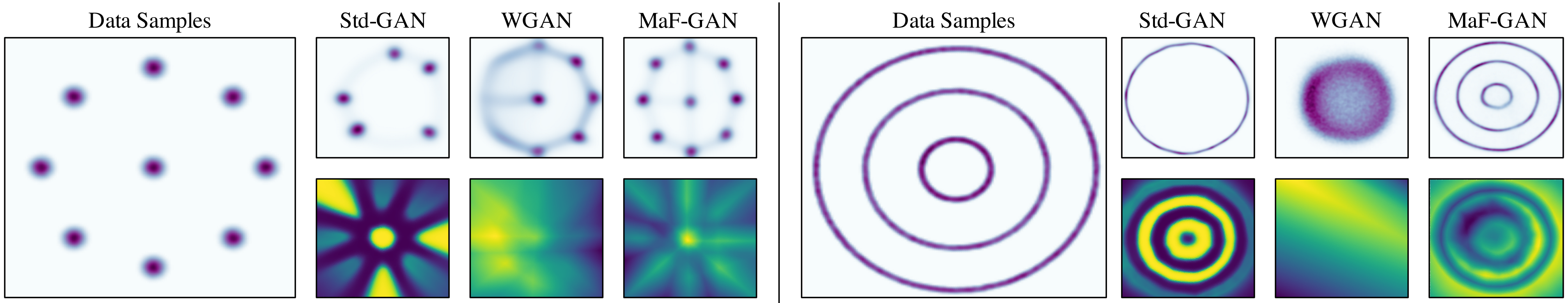
FIDs of DCGAN using various learning objectives

Learning Objective	CIFAR-10	CelebA
WGAN [1] (ICML'17)	55.96	-
HingeGAN [58] (ICLR'17)	42.40	25.57
LSGAN [33] (ICCV'17)	42.01	30.76
DCGAN [37] (ICLR'16)	38.56	27.02
WGAN-GP [6] (NIPS'17)	41.86	70.28
Re-implemented WGAN-GP	38.63	70.16
Realness GAN-Obj.1 [47] (ICLR'2020)	36.73	-
Realness GAN-Obj.2 [47] (ICLR'2020)	34.59	23.51
Realness GAN-Obj.3 [47] (ICLR'2020)	36.21	-
AdaptiveMix (Ours)	**30.85**	**12.43**

The proposed method for StyleGAN-V2

Method	AFHQ-Cat-5k		FFHQ (Full)	
	FID	IS	FID	IS
StyleGAN-V2 [20] (CVPR'20)	7.737	1.825	3.862	5.243
StyleGAN-V2 (Re-Impl.)	7.924	1.890	3.810	5.185
LC-Reg [43] (CVPR'21)	6.699	1.943	3.933	**5.312**
Style GAN-V2 + Ours	**4.477**	**1.972**	**3.623**	5.222
ADA [17] (NIPS'20)	6.053	2.119	4.018	5.329
ADA (Re-Impl.)	5.582	2.059	3.713	5.200
ADA + Ours	**4.680**	**2.069**	**3.681**	**5.335**
APA [15] (NIPS'2021)	4.876	2.156	3.678	5.336
APA (Re-Impl.)	4.645	2.093	3.752	5.281
APA+Ours	**4.148**	**2.096**	**3.609**	**5.296**

The experimental results on a synthetic data set



(a) Mixed Gaussian Distributions (b) Mixed Circle Lines

AdaptiveMix on Image Generation

FFHQ

StyleGAN-V2

StyleGAN-V2 + AdaptiveMix

AdaptiveMix on Visual Recognition

Image Classification

(1) Given samples and their labels $(x_i, y_i), (x_j, y_j) \sim (\mathcal{X}, \mathcal{Y})$

(2) Generate hard samples $\hat{x}_{ij} = g(x_i, x_j, \lambda)$ $\hat{y}_{ij} = g(y_i, y_j, \lambda)$ $(\hat{x}_{ij} = g(x_i, x_j, \lambda) = \lambda x_i + (1-\lambda) x_j)$

(3) Input samples to feature extractor $v_i, v_j, \hat{v}_{ij} \leftarrow \mathcal{F}(x_i), \mathcal{F}(x_j), \mathcal{F}(\hat{x}_{ij})$

(4) Given an orthogonal classifier $\tilde{\mathcal{J}}(\cdot)$

The **learning objective** of AdaptiveMix based image classification:

$$\mathcal{L}_t = \hat{y}_{ij} log(\tilde{\mathcal{J}}(\mathcal{F}(\hat{x}_{ij}))) + \hat{y}_{ij} log(\tilde{\mathcal{J}}(g(v_i, v_j, \lambda))) + \mathcal{L}_{ada}$$

Image OOD Detection

Given a test sample x_t, the probability ϕ_t that x_t is OOD is calculated as:

$$\phi_t = \min_k \arccos\left(\frac{|\mathcal{F}^T(x_t) v_k^*|}{||\mathcal{F}(x_t)||}\right),$$

AdaptiveMix on Visual Recognition

Robust Image Recognition

Table 7. Accuracy (%) on CIFAR-10 based on WRN-28-10 trained with the various methods with orthogonal classifier (Orth.).

CIFAR10	FGSM (8/255)	PGD-8 (4/255)	PGD-16 (4/255)	CW-100 (c=0.01)	CW-100 (c=0.05)
Baseline	38.03	0.92	0.28	11.1	0.39
Mixup [54]	60.17	3.97	1.16	30.32	2.36
Orth. + Mixup	44.80	3.99	2.66	71.12	49.47
M.-Mixup [44]	59.32	7.97	2.97	51.47	11.12
Orth. + M.-Mixup	38.76	5.77	4.38	69.08	53.98
Ours	**74.18**	**32.12**	**22.12**	**81.39**	**74.72**

Table 8. Accuracy (%) on CIFAR-100 and Tiny-ImageNet against various adversarial attacks based on WRN-28-10 [53] and PreActResNet-18 [9] respectively.

Dataset	Method	FGSM (8/255)	PGD-8 (4/255)	PGD-16 (4/255)	CW-100 (c=0.01)	CW-100 (c=0.05)
C-100	Baseline	11.71	0.79	0.42	4.42	0.23
	Mixup [54]	27.34	0.28	0.11	4.83	0.28
	M.-Mixup [44]	**29.73**	1.19	0.49	10.75	0.77
	Ours	24.28	**8.22**	**7.40**	**42.02**	**26.18**
T-ImageNet	Baseline	4.26	0.81	0.60	27.92	7.52
	Mixup [54]	4.23	0.98	0.77	29.13	15.41
	M.-Mixup [44]	3.04	0.82	0.59	29.69	16.86
	Ours	**7.10**	**4.66**	**4.98**	**35.93**	**34.22**

Clean Image Recognition

Table 11. Accuracy (%) of the proposed AdaptiveMix on varying baselines and datasets. Res. stands for resolution of the input.

Dataset	Architecture	Res.	Baseline	Ours
CIFAR-10	WRN-28-10 [53]	32^2	96.11	**96.80**
CIFAR-100	WRN-28-10 [53]	32^2	80.82	**82.02**
T-ImageNet	PreActResNet-18 [9]	64^2	57.23	**60.59**
ImageNet	ResNet-50 [8]	128^2	67.38	**68.69**

OOD Detection

Table 12. OOD detection on various OOD sets, where TIN-C, TIN-R, LSUN-C, and LSUN-R refer to the OOD set of Tiny ImageNet-Crop, Tiny ImageNet-Resize, LSUN-Crop, and LSUN-Resize, respectively. All values are F1 score (↑), † stands for the result reproduced by the open-source code.

ID Dataset	CIFAR10			
OOD Dataset	TIN-C	TIN-R	LSUN-C	LSUN-R
Methods using MC sampling				
1DS [52] (CVPR'21)	0.930	0.936	0.962	0.961
Methods which adopt OOD samples for validation and fine-tuning				
ODIN [26] (ICLR'18)	0.902	0.926	0.894	0.937
Mahalanobis [24] (NIPS'18)	0.985	0.969	0.985	0.975
Soft. Pred. [10] (ICLR'17)	0.803	0.807	0.794	0.815
Counterfactual [36] (ECCV'18)	0.636	0.635	0.650	0.648
CROSR [49] (CVPR'19)	0.733	0.763	0.714	0.731
OLTR [32] (CVPR'19)	0.860	0.852	0.877	0.877
1DS w/o MC † [52]	0.890	0.886	0.897	0.907
1DS w/o MC † +Ours	**0.922**	**0.911**	**0.934**	**0.937**

Conclusion

1. We propose a novel module, namely AdaptiveMix, to improve the training of GANs. Our AdaptiveMix is simple yet effective and plug-and-play, which is helpful for GANs to generate high-quality images.

2. We show that GANs can be stably and efficiently trained by shrinking regions of training data in image representation supported by the discriminator.

3. We show our AdaptiveMix can be applied to not only image generation, but also OOD and robust image classification tasks. Extensive experiments show that our AdaptiveMix consistently boosts the performance of baselines for four different tasks (*e.g.,* OOD) on seven widely-used datasets.

Thank you for listening!