Dual Contrastive Loss and Attention for GANs

Ning Yu[1,2] Guilin Liu[3] Aysegul Dundar[3,4]
Andrew Tao[3] Bryan Catanzaro[3] Larry Davis[1] Mario Fritz[5]
[1]University of Maryland [2]Max Planck Institute for Informatics [3]NVIDIA
[4]Bilkent University [5]CISPA Helmholtz Center for Information Security

https://github.com/ningyu1991/AttentionDualContrastGAN

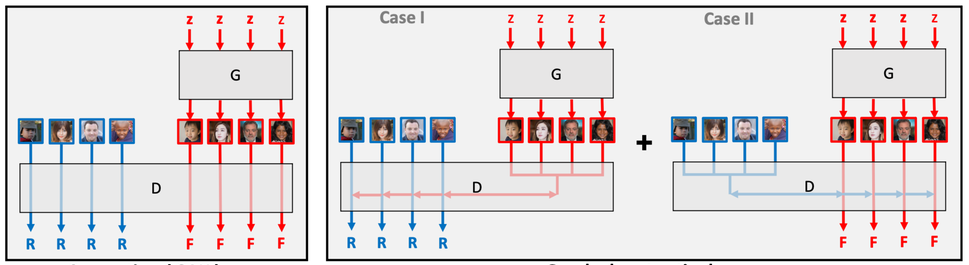
2021 ICCV OCTOBER 11-17 VIRTUAL

Motivations

- Generative Adversarial Networks (GANs) evolve fast in the past 7 years for photorealistic generation. However, uncurated generations still suffer from **artifacts** that are easy to spot.
- We improve on StyleGAN2, by revisiting its **loss** and **architectures**.
- We propose a novel **dual contrastive loss** to replace traditional cross-entropy loss in the adversarial training.
- We revisit the **self-attention** modules in the generator architecture.
- We propose a novel **reference-attention** module in the discriminator architecture.

Dual contrastive loss

- **Batch-wise pick-one-out** classification instead of sample-wise binary classification.
  - One real v.s. a batch of fakes.
  - One fake v.s. a batch of reals.



Conventional GAN loss | Our dual contrastive loss

$$L_{real}^{contr}(G,D) = \mathop{\mathbb{E}}_{\mathbf{x}\sim p(\mathbf{x})}\left[\log \frac{e^{D(\mathbf{x})}}{e^{D(\mathbf{x})} + \sum_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)} e^{D(G(\mathbf{z}))}}\right] = -\mathop{\mathbb{E}}_{\mathbf{x}\sim p(\mathbf{x})}\left[\log\left(1 + \sum_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)} e^{D(G(\mathbf{z}))-D(\mathbf{x})}\right)\right]$$

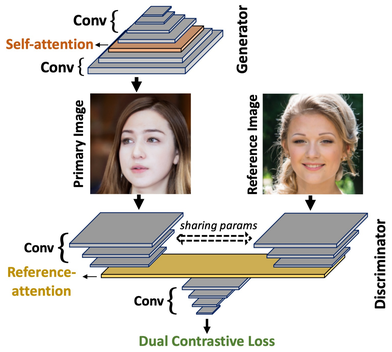
$$+ \quad L_{fake}^{contr}(G,D) = \mathop{\mathbb{E}}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)}\left[\log \frac{e^{-D(G(\mathbf{z}))}}{e^{-D(G(\mathbf{z}))} + \sum_{\mathbf{x}\sim p(\mathbf{x})} e^{-D(\mathbf{x})}}\right] = -\mathop{\mathbb{E}}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)}\left[\log\left(1 + \sum_{\mathbf{x}\sim p(\mathbf{x})} e^{D(G(\mathbf{z}))-D(\mathbf{x})}\right)\right]$$

	FFHQ	Bedroom	Church	Horse	CLEVR
Non-saturating [23] (default)	4.86	4.01	4.54	3.91	9.62
Saturating [23]	5.16	4.26	4.80	5.90	10.46
Wasserstein [24]	7.99	6.05	6.28	7.23	**5.82**
Hinge [47]	4.14	4.92	4.39	5.27	14.87
Dual contrastive (**ours**)	**3.98**	**3.86**	**3.73**	**3.70**	6.06

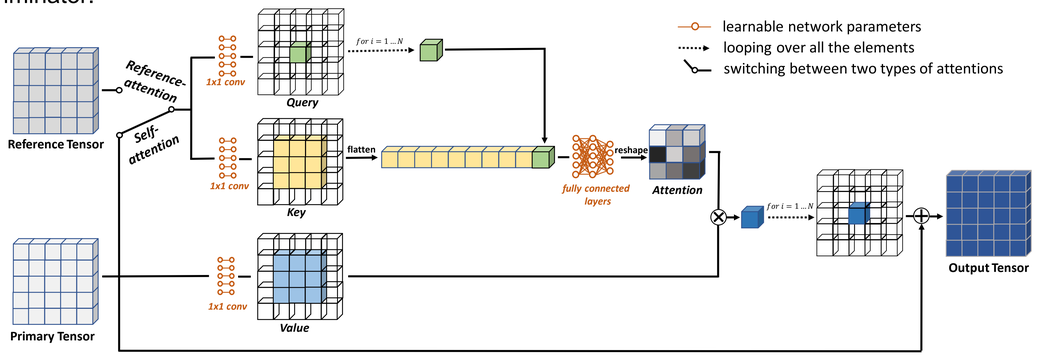
Significant FID improvements over traditional losses on several datasets

Self-attention in the generator and reference-attention in the discriminator

- In the generator, we replace one layer of convolution with the self-attention module SAN [1]: **long-range** and **spatially adaptive**.
- In the discriminator, we introduce a reference real image input and Siamese network, and merge the two branches using a novel reference-attention module.
  - **Feature augmentation** for discriminator training.
  - **Balance** between generator and discriminator.



Architecture diagram



Self-attention and reference attention modules

	CelebA	Animal Face	Bedroom	Church
StyleGAN2 [43]	9.84	36.55	19.33	11.02
+ DFN [37]	**8.41**	35.10	26.86	11.31
+ VT [85]	9.18	34.70	16.85	10.64
+ SAGAN [98]	9.35	34.83	17.94	10.65
+ SAN [103]	8.60	**32.72**	**16.36**	**9.62**

All the self-attention modules in G improve FID
SAN [1] improves the most

Method	FLOPS (G)	#parameters (M)
StyleGAN2 [41]	1.08	48.77
+ DFN [35]	4.20	177.60
+ VT [81]	7.39	240.09
+ SAGAN [94]	0.99	44.99
+ SAN [99]	1.08	48.43

SAN [1] does not increase complexity

	CelebA	Animal Face	Bedroom	Church
StyleGAN2 [41]	9.84	36.55	19.33	11.02
+ self attn in D	10.49	42.41	17.22	11.06
+ ref attn in D	**7.48**	**31.08**	**8.32**	**7.86**

Reference-attention in D improves FID
Self-attention in D does not

[1] Zhao, Hengshuang, Jiaya Jia, and Vladlen Koltun. "Exploring self-attention for image recognition." CVPR. 2020.

Combine all our contributions

Method	Loss	FFHQ	Bedroom	Church	Horse	CLEVR
BigGAN [5]	Adv	11.4	-	-	-	-
U-Net GAN [69]	Adv	7.48	17.6	11.7	20.2	33.3
StyleGAN2 [43]	Adv	4.86	4.01	4.54	3.91	9.62
StyleGAN2 w/ attn	Adv	5.13	3.48	4.38	3.59	8.96
StyleGAN2	Contr	**3.98**	3.86	3.73	3.70	6.06
StyleGAN2 w/ attn	Contr	4.63	**3.31**	**3.39**	**2.97**	**5.05**

Progressively improves FID by 17% - 48%