Data InStance Prior (DISP) in Generative Adversarial Networks

Puneet Mangla* [3] **Nupur Kumari*** [2] **Mayank Singh*** [2] Balaji Krishnamurthy [1] Vineeth N Balasubramanian [3]

[1]Media and Data Science Research lab, Adobe [2]CMU [3]IIT, Hyderabad, India (* Equal Contribution)

ArXiv

Problem Definition and Contribution

Challange: In limited data regimes, GAN training typically diverges, and the generated samples are of low quality and lack diversity.

- We propose Data InStance Prior (DISP) - novel transfer learning technique for GANs in low-data setting.
- DISP achieves SOTA image quality and diversity performance in few-shot ($\sim 25-100$), limited ($\sim 2k-6k$) and large-scale ($\sim 50k-2M$) image generation benchmarks.

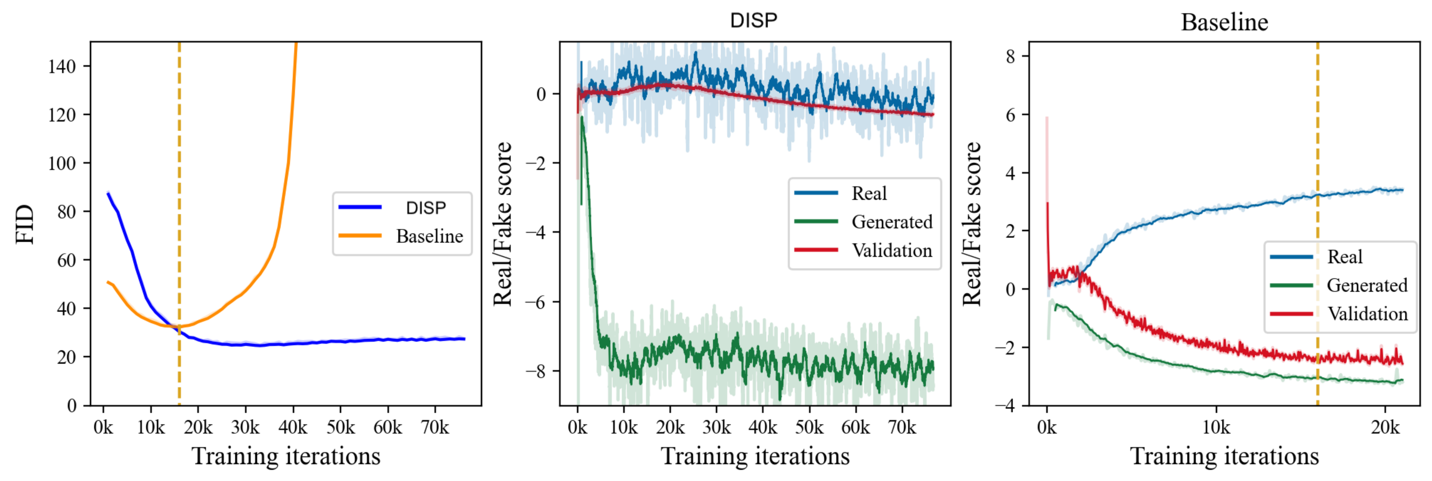


Figure 1. Comparison between DISP and Baseline when trained on 10% data of CIFAR-100.

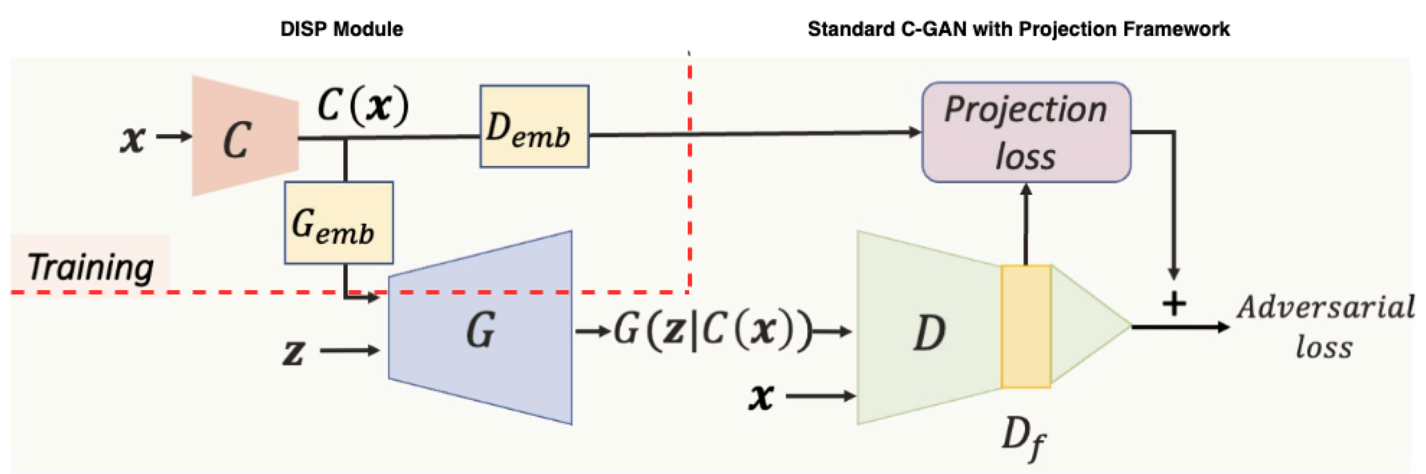
Motivation

- Knowledge transfer from self-supervised/supervised networks, pre-trained on rich source domain.
- Motivated from IMLE - propose a regularizer to prevent mode collapse and discriminator overfitting.

Our Approach: Data InStance Prior (DISP)

Training

Given a pre-trained feature extractor $C : \mathbb{R}^p \rightarrow \mathbb{R}^d$ (trained on a rich source domain using supervisory signals or self-supervision), we leverage its feature representation $C(\mathbf{x})$ as conditional information for GAN training.



$$L_D = \mathbb{E}_{\mathbf{x}\sim q(x)}[\max(0, 1 - D(\mathbf{x}, C(\mathbf{x})))] + \mathbb{E}_{\mathbf{x}\sim q(x),\mathbf{z}\sim p(\mathbf{z})}[\max(0, 1 + D(G(\mathbf{z}|C(\mathbf{x})), C(\mathbf{x})))] \quad (1)$$
$$L_G = -\mathbb{E}_{\mathbf{x}\sim q(x),\mathbf{z}\sim p(\mathbf{z})}[D(G(\mathbf{z}|C(\mathbf{x})), C(\mathbf{x}))]$$

$$D(\mathbf{x}, \mathbf{y}) = D_{emb}(\mathbf{y}) \cdot D_f(\mathbf{x}) + D_l \circ D_f(\mathbf{x}) \quad (2)$$
is the **c-GAN projection loss**

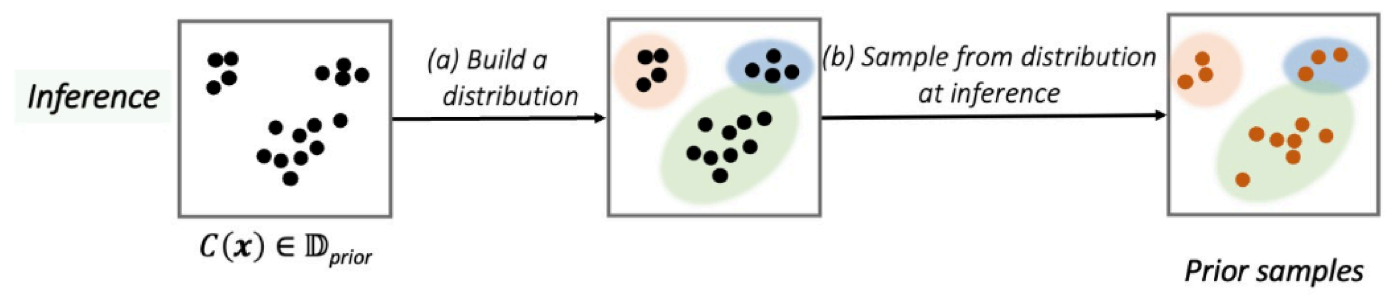
- Since $C(\mathbf{x})$ is extracted from a pre-trained network, above training objective leads to feature level knowledge distillation from C.
- It also acts as a regularizer on the discriminator to reduce overfitting.
- Enforcing feature $D_f(G(\mathbf{z}|C(\mathbf{x})))$ to be similar to $D_{emb}(C(\mathbf{x}))$ promotes mode coverage of target data distribution.

Inference

Let $\mathbb{D}_{prior} = \{C(\mathbf{x}_j)\}_{j=1}^{n}$. The generator requires access to $\mathbb{D}_{prior}$ for sample generation.

- **few-shot** and **limited** data setting - We generate images conditioned on prior samples from a mixup distribution of $\mathbb{D}_{prior}$.
- **large-scale** data setting - We learn a GMM on $\mathbb{D}_{prior}$. This enables memory efficient sampling of conditional priors.

$$G(\mathbf{z}|\mathcal{N}(\mu, \Sigma)) \quad \text{where } \mu, \Sigma \sim \text{GMM}(G_{emb}(\mathbb{D}_{prior})) \quad (3)$$

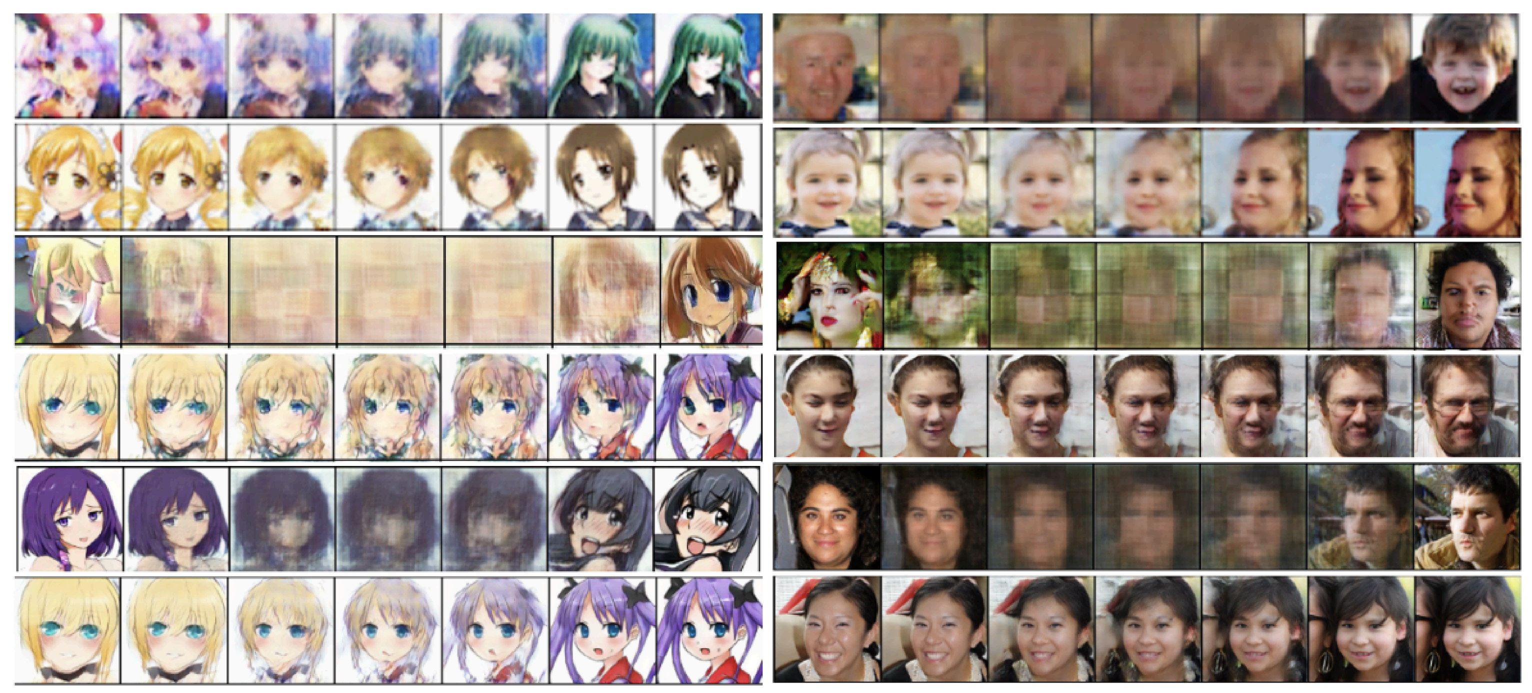


Experiments and Results

Few-shot image generation performance using 100 training images

Method	Pre-training	SNGAN (128 x 128) Anime FID↓	P↑	R↑	Faces FID↓	P↑	R↑
From scratch		120.38	0.61	0.00	140.66	0.31	0.00
+ DISP-Vgg16		**66.85**	**0.71**	**0.03**	**68.49**	**0.74**	**0.15**
TransferGAN		102.75	**0.70**	0.00	101.15	**0.85**	0.00
+ DISP-Vgg16		**86.96**	0.57	**0.02**	**75.21**	0.70	**0.10**
FreezeD		109.40	**0.67**	0.00	107.83	**0.83**	0.00
+ DISP-Vgg16		93.36	0.56	**0.03**	77.09	0.68	0.14
+ DISP-SimCLR		**89.39**	0.46	0.025	**70.40**	0.74	**0.22**
ADA		78.28	0.87	0.0	159.3	0.69	0.0
+ DISP-Vgg16		**60.8**	**0.90**	**0.003**	**79.5**	**0.85**	**0.004**
DiffAugment		85.16	**0.95**	0.00	109.25	**0.84**	0.00
+ DISP-Vgg16		**48.67**	0.82	0.03	**62.44**	0.80	0.19
+ DISP-SimCLR		52.41	0.77	**0.04**	64.53	0.78	**0.22**

Sample interpolations between two generated images for models trained in few-shot setting : Scratch *(Row 1)*, Scratch + DISP-Vgg16 *(Row 2)*, FreezeD *(Row 3)*, FreezeD + DISP-Vgg16 *(Row 4)*, DiffAugment *(Row 5)*, DiffAugment + DISP-Vgg16 *(Row 6)*



Comparison of FID on Unconditional CIFAR-10 and CIFAR-100 image generation while varying the amount of training data.

Method	CIFAR-10 100% data	CIFAR-10 20% data	CIFAR-10 10% data	CIFAR-100 100% data	CIFAR-100 20% data	CIFAR-100 10% data
BigGAN	17.22	31.25	42.59	20.37	33.25	42.43
+ DISP	9.70	16.24	27.86	12.89	21.70	31.48
+ DiffAugment	10.39	15.12	18.56	13.33	19.78	23.80
+ DiffAugment & DISP	**9.52**	**14.24**	**18.50**	**12.70**	**16.91**	**20.47**
StyleGAN2*	11.07	23.08	36.02	16.54	32.30	45.87
+ DiffAugment*	9.89	12.15	14.5	15.22	16.65	20.75
+ DiffAugment & DISP	**9.50**	**10.92**	**12.03**	**14.45**	**15.52**	**17.33**

Comparison of DISP with Baseline, SSGAN and Self-Cond GAN in large-scale image generation setting on FID, Precision and Recall metrics.

Method	CIFAR-10 FID↓	P↑	R↑	CIFAR-100 FID↓	P↑	R↑	FFHQ FID↓	P↑	R↑	LSUN-Bedroom FID↓	P↑	R↑	ImageNet32x32 FID↓	P↑	R↑
Baseline	19.73	0.64	**0.70**	24.66	0.61	**0.67**	21.67	0.77	0.47	9.89	0.58	0.42	16.19	0.60	**0.67**
SSGAN	15.65	0.67	0.68	21.02	0.61	0.65	-	-	-	7.68	0.59	0.50	17.18	0.61	0.65
Self-Cond GAN	16.72	0.71	0.64	21.8	0.64	0.60	-	-	-	-	-	-	15.56	**0.66**	0.63
DISP-Vgg16	**11.24**	**0.74**	0.64	**15.71**	**0.70**	0.62	**15.83**	0.76	**0.55**	4.99	**0.66**	**0.54**	**12.11**	0.64	0.62
DISP-SimCLR	14.42	0.68	0.65	20.08	0.67	0.62	16.62	**0.77**	0.53	**4.92**	0.62	0.53	14.99	0.60	0.63

Semantic Diffusion Exploit $C(\mathbf{x})$, to get some control over the high-level semantics (e.g. hair, gender, glasses, etc in case of faces) of generated image.

(a) Custom Editing - First column shows human-edited version where certain portion of image is substituted with another to achieve desired semantics. Rest columns correspond to images generated when Vgg16 features of human-edited version is provided as prior to DISP module.

(b) Sketch-to-Image - First column shows sketch describing desired high-level semantics. Rest columns correspond to images generated when Vgg16 features of the sketch version is provided as prior in DISP module.

References

[1] Ke Li and Jitendra Malik. Implicit maximum likelihood estimation
[2] Seonguk Seo, Yumin Suh, D. Kim, Jongwoo Han, and B.Han. Learning to optimize domain specific normalizationfor domain generalization.
[3] Zhao et al.Differentiable augmentation for data-efficient gan training.
[4] Karras et al. Analyzing and improving the image quality of stylegan.
[5] Brock et al. Largescale gan training for high fidelity natural image synthesis
[6] Chen et al. Self-supervised gans via auxiliary rotation loss.
[7] Liu et al. Diverse image generation via self-conditioned gans.