Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis

Qi Mao*[1], Hsin-Ying Lee*[2], Hung-Yu Tseng*[2], Siwei Ma[1,3], Ming-Hsuan Yang[2,4]

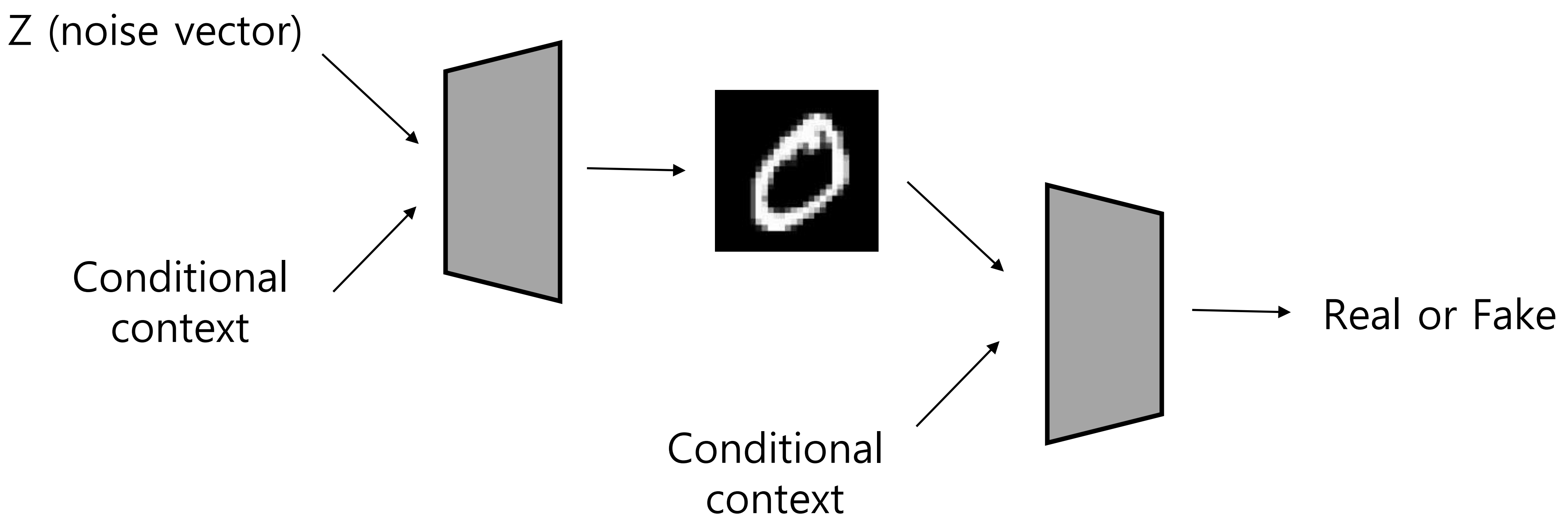
[1]Institute of Digital Media, Peking University [2]University of California, Merced
[3]Peng Cheng Laboratory [4]Google Cloud

CVPR 2019

19/07/25, Yonggyu Kim

cGAN

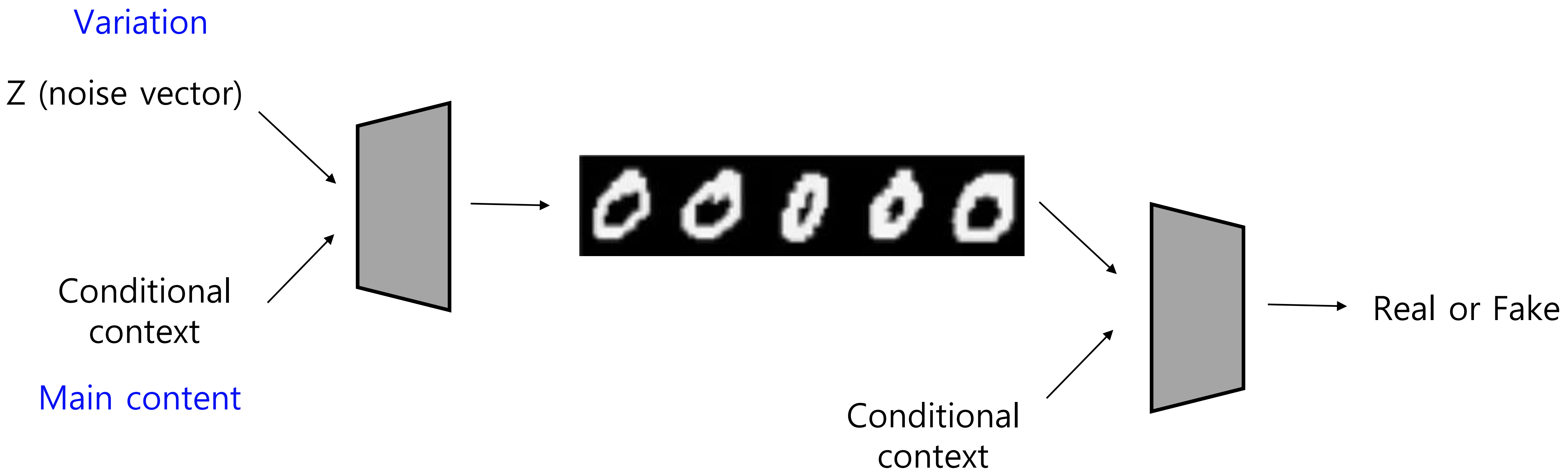
이 상



Z (noise vector)
Conditional context
Conditional context
Real or Fake

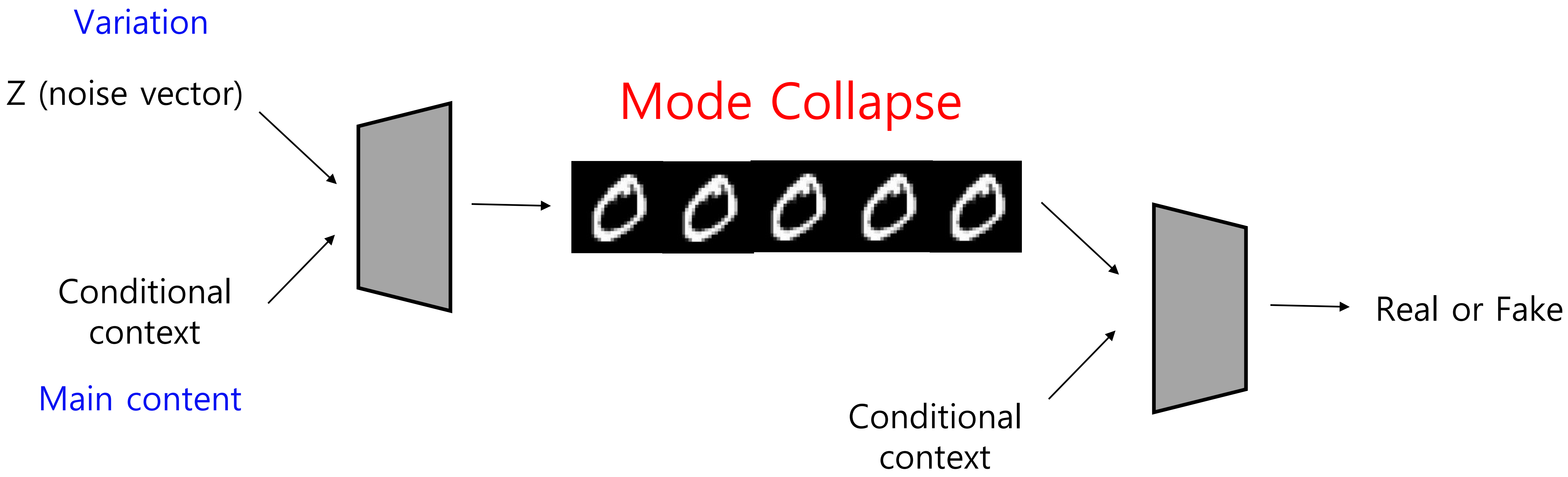
cGAN

이 상


Variation
Z (noise vector)
Conditional context
Main content
Conditional context
Real or Fake

cGAN

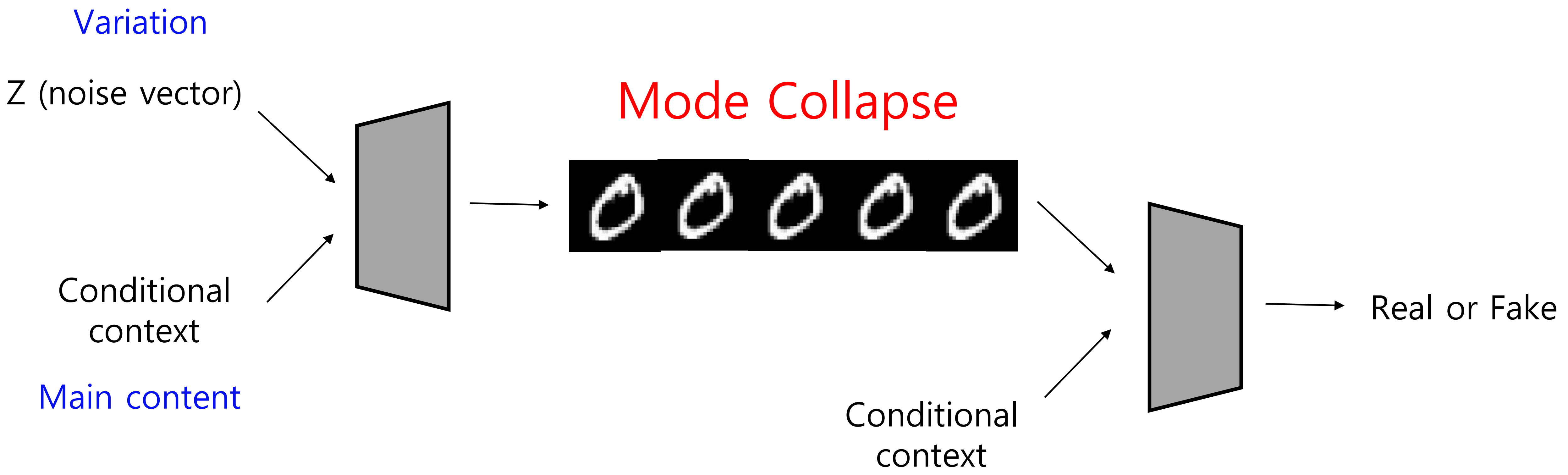
현 실



Variation
Z (noise vector)
Conditional context
Main content
Mode Collapse
Conditional context
Real or Fake

cGAN

현 실



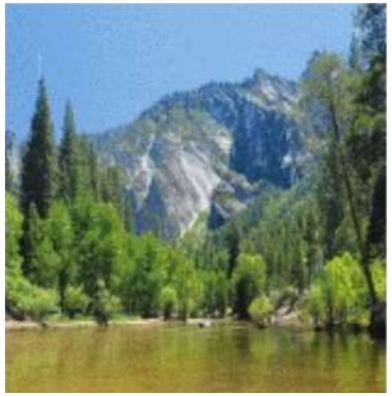
Prior conditional information

cGAN

현 실

Variation

Z (noise vector)



Main content

High dimentional context

Mode Collapse

Divergence metrics

- WGAN (M. Arjovsky. et al.)
- LSGAN (X. Mao. et al.)

Multple Generator

- Coupled generative adversarial networks (M.-T. Liu et al.)
- Multi-agent diverse generative adversaral networks. (A. Ghosh et al.)

I2I translation task

- MUNIT (X.Huang et al.)
- DRIT (H.-T. Lee et al.)
- BicycleGAN (J.-Y. Zhu et al.)

Motivation & Contribution

- They propose a simple yet effective mode seeking regularization method to address the mode Collapse problem in cGANs. This regularization scheme can be readily extended into existing frameworks with marginal training overheads and modifications.

- They demonstrate the generalizability of the proposed regularization method on three different conditional generation tasks : categorical generation, image-to-image translation, and text-to-image synthesis.

- Extensive experiments show that the proposed method can facilitate existing models from different tasks achieving better diversity without sacrificing visual quality of the generated images.

Motivation & Contribution

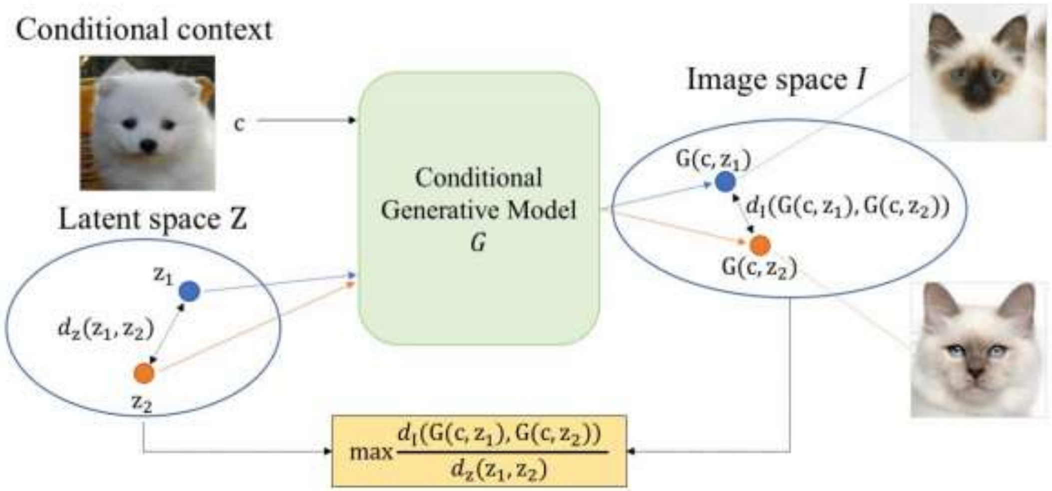
Mode Seeking Regularization

$$\mathcal{L}_{\mathrm{ms}} = \max_{G}\left(\frac{d_{\mathbf{I}}(G(c,\mathbf{z}_1), G(c,\mathbf{z}_2))}{d_{\mathbf{z}}(\mathbf{z}_1,\mathbf{z}_2)}\right), \qquad (1)$$

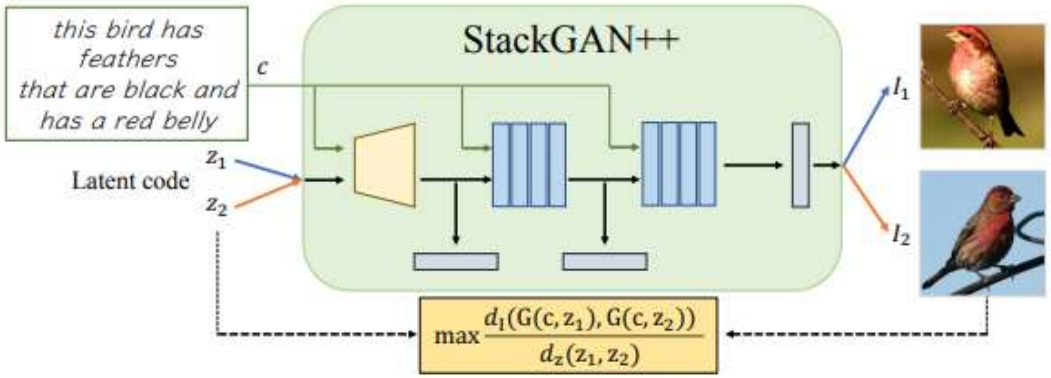
$$\mathcal{L}_{\mathrm{new}} = \mathcal{L}_{\mathrm{ori}} + \lambda_{\mathrm{ms}}\mathcal{L}_{\mathrm{ms}}, \qquad (2)$$

The mode collapse problem with GANs is well known in the literature. Several methods [2, 26, 27] attribute the missing mode to the lack of penalty when this issue occurs. Since all modes usually have similar discriminative values, larger modes are likely to be favored through the training process based on gradient descent. On the other hand, it is difficult to generate samples from minor modes.

model collapse in cGANs. By maximizing the distance between generated images with respect to that between the corresponding latent codes, the regularization term forces the generators to explore more minor modes. The proposed



(a) Proposed regularization



(b) Applying proposed regularization on StackGAN++

Motivation & Contribution

Mode Seeking Regularization

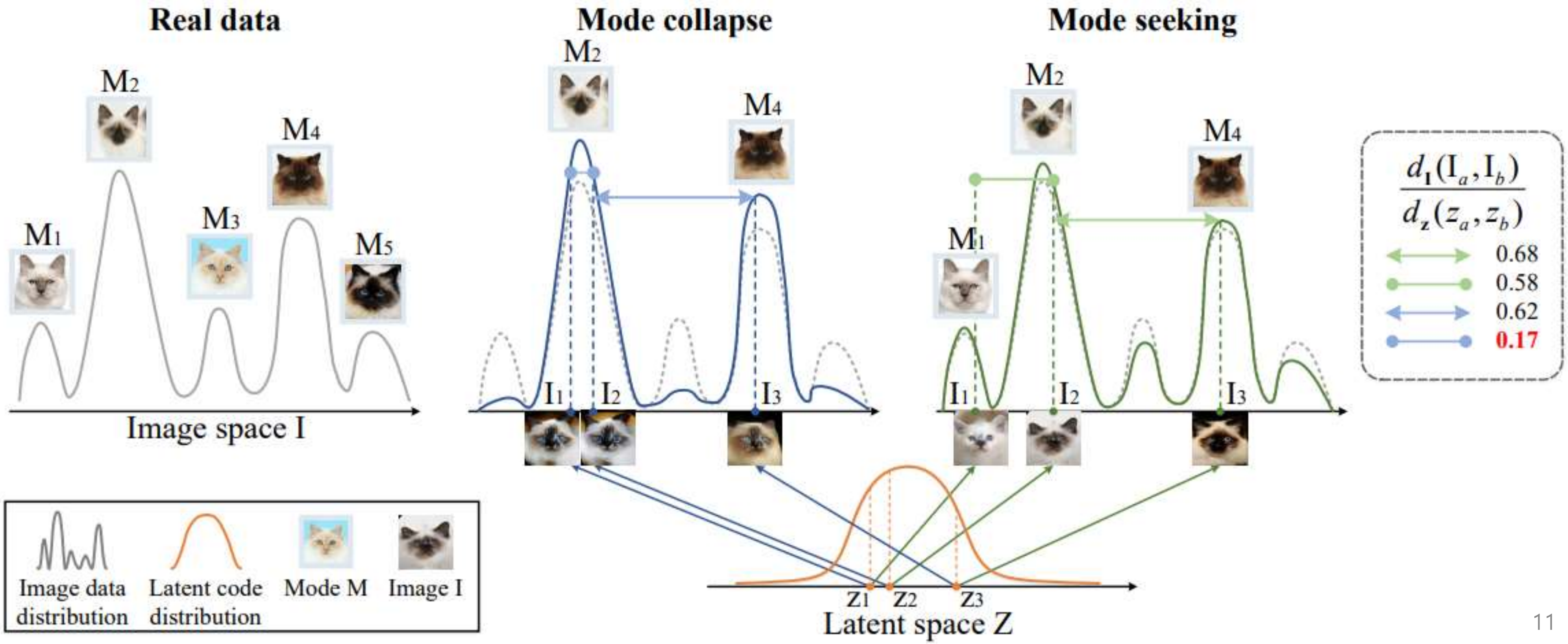
```
# mode seeking loss for A-->B and B-->A
lz_AB = torch.mean(torch.abs(self.fake_B_random2 - self.fake_B_random)) / torch.mean(torch.abs(self.z_random2 - self.z_random))
lz_BA = torch.mean(torch.abs(self.fake_A_random2 - self.fake_A_random)) / torch.mean(torch.abs(self.z_random2 - self.z_random))
eps = 1 * 1e-5
loss_lz_AB = 1 / (lz_AB + eps)
loss_lz_BA = 1 / (lz_BA + eps)
```

Motivation & Contribution

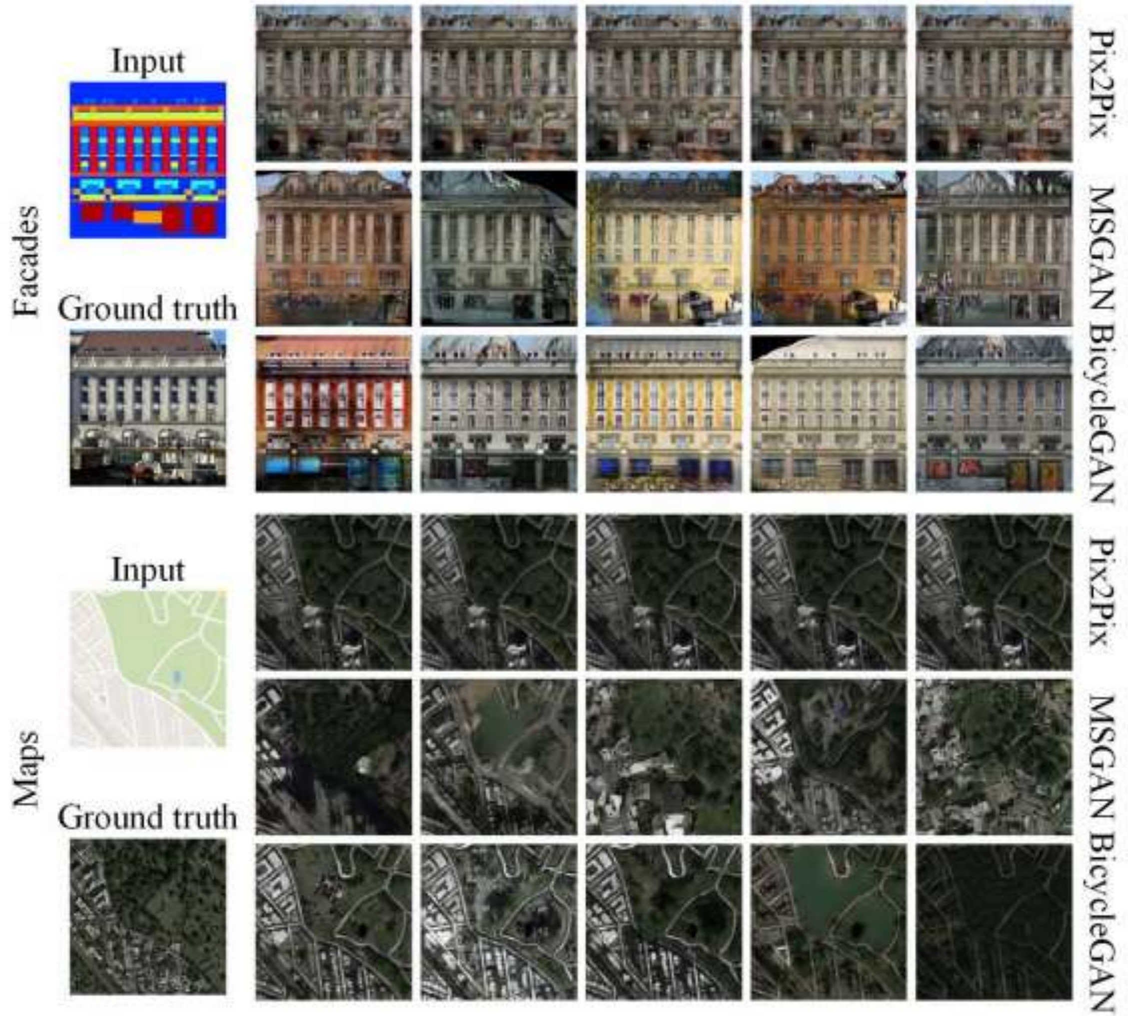
Mode Seeking Regularization

$$\mathcal{L}_{\text{ms}} = \max_{G}\left(\frac{d_{\mathbf{I}}(G(c,\mathbf{z}_1), G(c,\mathbf{z}_2))}{d_{\mathbf{z}}(\mathbf{z}_1,\mathbf{z}_2)}\right), \qquad (1)$$

$$\mathcal{L}_{\text{new}} = \mathcal{L}_{\text{ori}} + \lambda_{\text{ms}}\mathcal{L}_{\text{ms}},$$

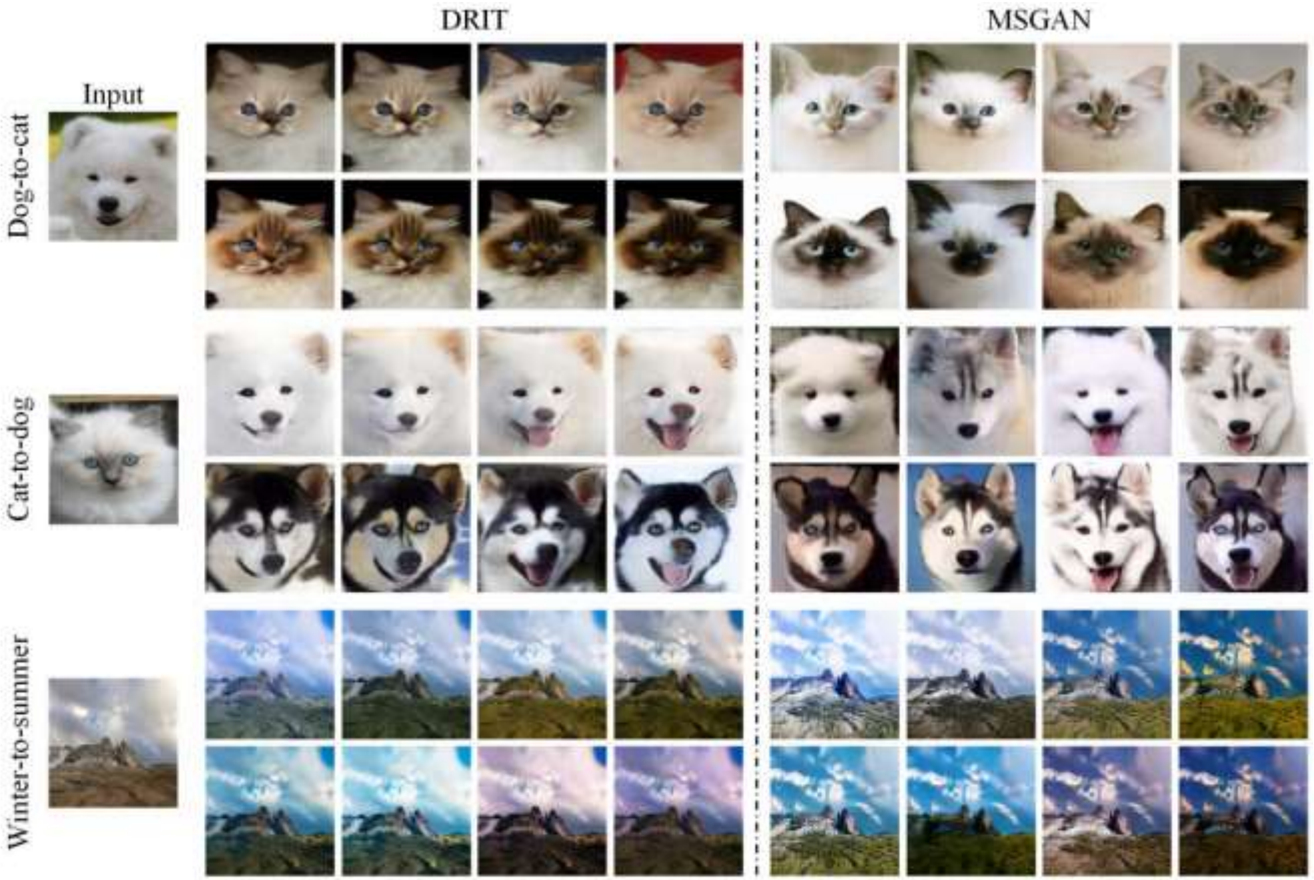


Experiment



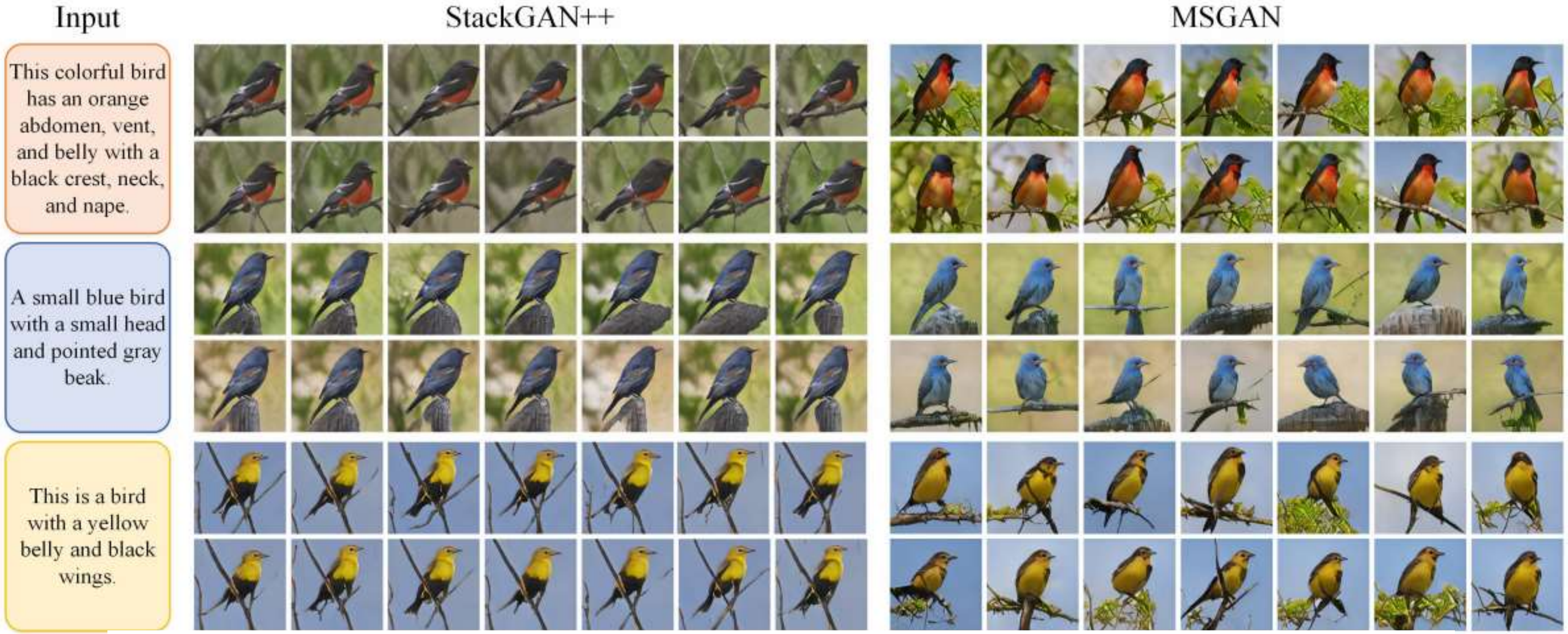
Datasets	Facades		
	Pix2Pix [11]	MSGAN	BicycleGAN [35]
FID ↓	139.19 ± 2.94	$\mathbf{92.84 \pm 1.00}$	98.85 ± 1.21
NDB↓	14.40 ± 1.82	$\mathbf{12.40 \pm 0.55}$	13.80 ± 0.45
JSD↓	0.074 ± 0.012	$\mathbf{0.038 \pm 0.004}$	0.058 ± 0.004
LPIPS↑	0.0003 ± 0.0000	$\mathbf{0.1894 \pm 0.0011}$	0.1413 ± 0.0005
Datasets	**Maps**		
	Pix2Pix [11]	MSGAN	BicycleGAN [35]
FID ↓	168.99 ± 2.58	152.43 ± 2.52	$\mathbf{145.78 \pm 3.90}$
NDB↓	49.00 ± 1.00	$\mathbf{41.60 \pm 0.55}$	46.60 ± 1.34
JSD↓	0.088 ± 0.018	0.031 ± 0.003	$\mathbf{0.023 \pm 0.002}$
LPIPS↑	0.0016 ± 0.0003	$\mathbf{0.2189 \pm 0.0004}$	0.1150 ± 0.0007

Experiment



Datasets	Summer2Winter		Winter2Summer	
	DRIT [15]	MSGAN	DRIT [15]	MSGAN
FID ↓	57.24 ± 2.03	$\mathbf{51.85 \pm 1.16}$	47.37 ± 3.25	$\mathbf{46.23 \pm 2.45}$
NDB↓	25.60 ± 1.14	$\mathbf{22.80 \pm 2.96}$	30.60 ± 2.97	$\mathbf{27.80 \pm 3.03}$
JSD↓	0.066 ± 0.005	$\mathbf{0.046 \pm 0.006}$	0.049 ± 0.009	$\mathbf{0.038 \pm 0.004}$
LPIPS↑	0.1150 ± 0.0003	$\mathbf{0.1468 \pm 0.0005}$	0.0965 ± 0.0004	$\mathbf{0.1183 \pm 0.0007}$
Datasets	Cat2Dog		Dog2Cat	
	DRIT [15]	MSGAN	DRIT [15]	MSGAN
FID↓	22.74 ± 0.28	$\mathbf{16.02 \pm 0.30}$	62.85 ± 0.21	$\mathbf{29.57 \pm 0.23}$
NDB↓	42.00 ± 2.12	$\mathbf{27.20 \pm 0.84}$	41.00 ± 0.71	$\mathbf{31.00 \pm 0.71}$
JSD↓	0.127 ± 0.003	$\mathbf{0.084 \pm 0.002}$	0.272 ± 0.002	$\mathbf{0.068 \pm 0.001}$
LPIPS↑	0.245 ± 0.002	$\mathbf{0.280 \pm 0.002}$	0.102 ± 0.001	$\mathbf{0.214 \pm 0.001}$

Experiment



	Conditioned on text descriptions		Conditioned on text codes	
	StackGAN++ [32]	MSGAN	StackGAN++ [32]	MSGAN
FID ↓	25.99 ± 4.26	$\mathbf{25.53 \pm 1.83}$	$\mathbf{27.12 \pm 1.15}$	27.94 ± 3.10
NDB↓	38.20 ± 2.39	$\mathbf{30.60 \pm 2.51}$	39.00 ± 0.71	$\mathbf{30.60 \pm 2.41}$
JSD↓	0.092 ± 0.005	$\mathbf{0.073 \pm 0.003}$	0.102 ± 0.016	$\mathbf{0.095 \pm 0.016}$
LPIPS↑	0.362 ± 0.004	$\mathbf{0.373 \pm 0.007}$	0.156 ± 0.004	$\mathbf{0.207 \pm 0.005}$

Experiment

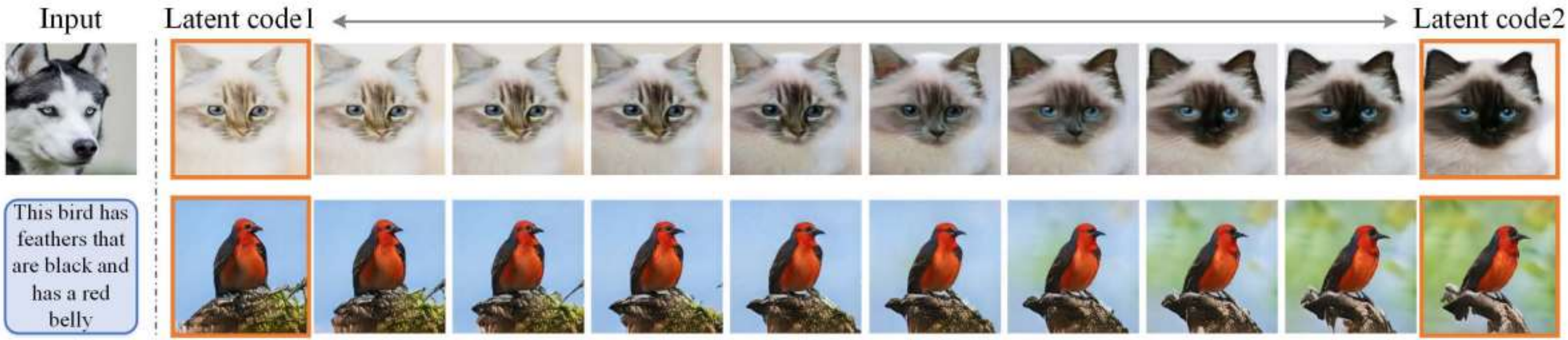


Table 1: **NDB and JSD results on the CIFAR-10 dataset.**

Metrics	Models	airplane	automobile	bird	cat	deer
NDB $\downarrow$	DCGAN	49.60 ± 3.43	53.00 ± 7.28	**34.40 ± 6.11**	46.00 ± 1.41	**44.80 ± 3.90**
	MSGAN	**46.60 ± 7.40**	**51.80 ± 2.28**	39.40 ± 1.95	**41.80 ± 3.70**	46.80 ± 4.92
JS $\downarrow$	DCGAN	0.034 ± 0.001	0.035 ± 0.002	**0.025 ± 0.002**	0.030 ± 0.002	**0.033 ± 0.001**
	MSGAN	**0.031 ± 0.001**	**0.033 ± 0.001**	0.027 ± 0.001	**0.027 ± 0.001**	0.035 ± 0.003
		dog	frog	horse	ship	truck
NDB $\downarrow$	DCGAN	50.40 ± 4.62	52.00 ± 3.81	54.40 ± 4.04	42.80 ± 5.45	47.80 ± 4.55
	MSGAN	**33.80 ± 3.27**	**42.00 ± 2.92**	**47.60 ± 5.03**	**41.00 ± 2.92**	**43.80 ± 6.61**
JS $\downarrow$	DCGAN	0.033 ± 0.001	0.034 ± 0.002	0.035 ± 0.001	0.029 ± 0.003	0.032 ± 0.001
	MSGAN	**0.024 ± 0.001**	**0.030 ± 0.002**	**0.033 ± 0.003**	**0.027 ± 0.001**	**0.029 ± 0.003**