Microsoft

CVPR

微软亚洲研究院创研论坛

CVPR 2020 论文分享会

01

Reusing Discriminators for Encoding: Towards Unsupervised Image-to-Image Translation

陈润发, **黄文炳**, 黄柄惠, 孙富春, 方斌

清华大学计算机系

01 • Content

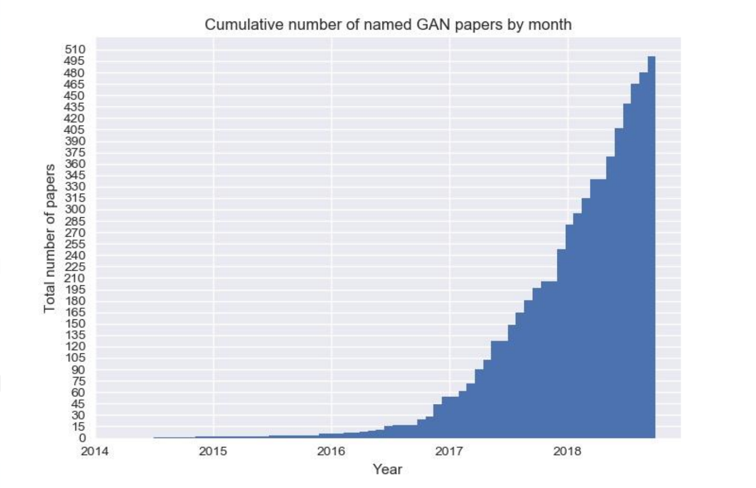
- Background
- Motivation
- NICE-GAN
- Experiments
- Conclusion

02 • Background

源域

目标域

GAN
ACGAN
BGAN
CGAN
DCGAN
EBGAN
fGAN
GoGAN
⋮

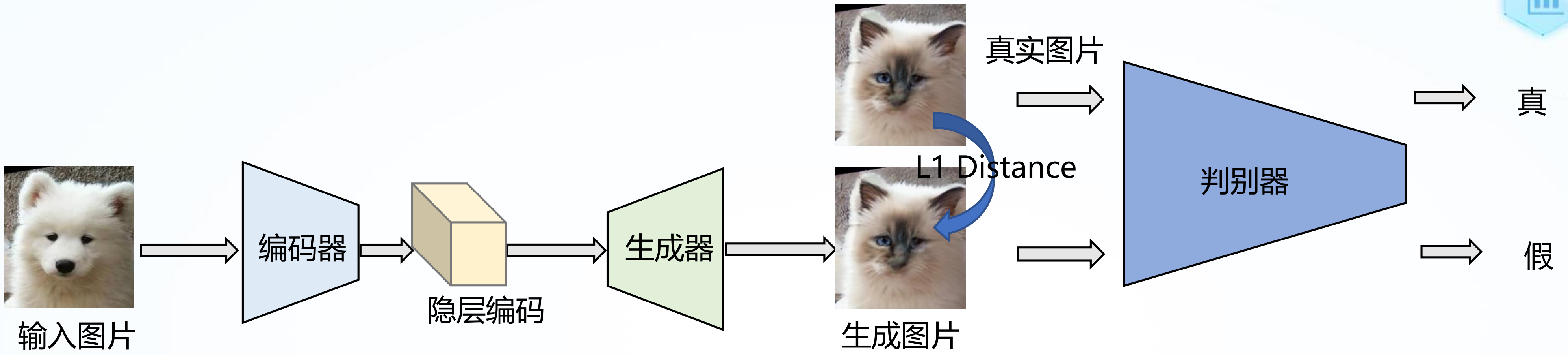


https://github.com/hindupuravinash/the-gan-zoo

跨域图片转换是图片编辑(换脸)、迁移学习、多模态融合等任务的关键步骤；

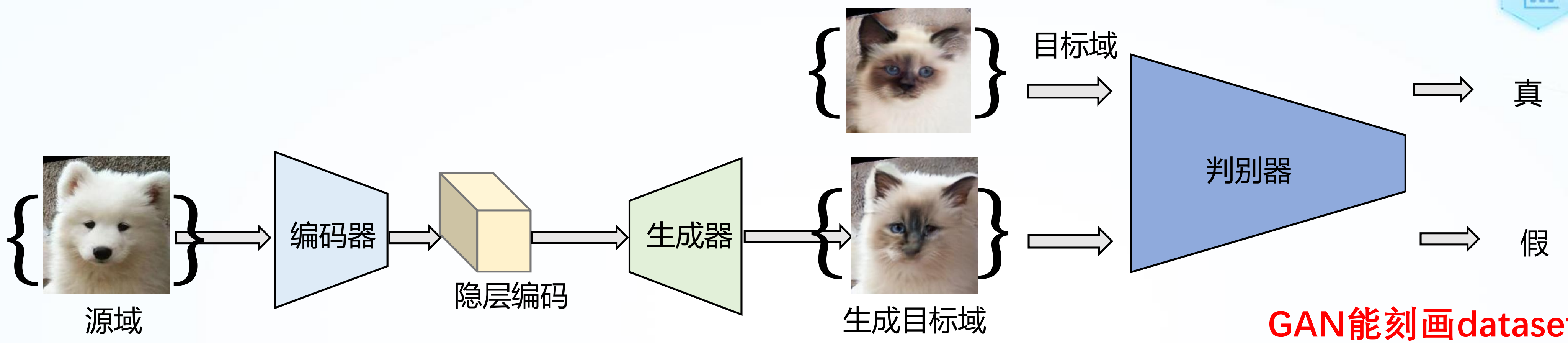
GAN已经成为图片转换技术是最有效的方法之一

03 • Background： 从监督图片转换到无监督图片转换



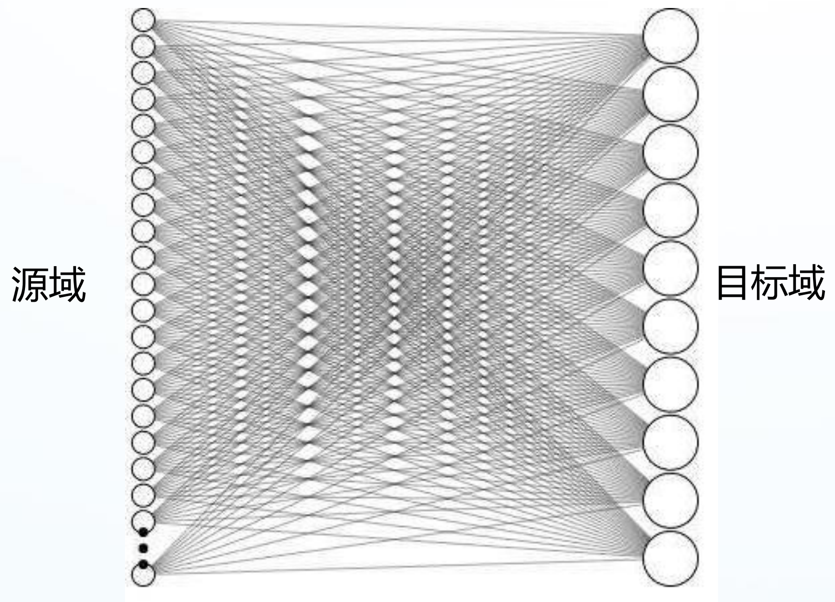
Isola et al. CVPR 2017

03 • Background： 从监督图片转换到无监督图片转换



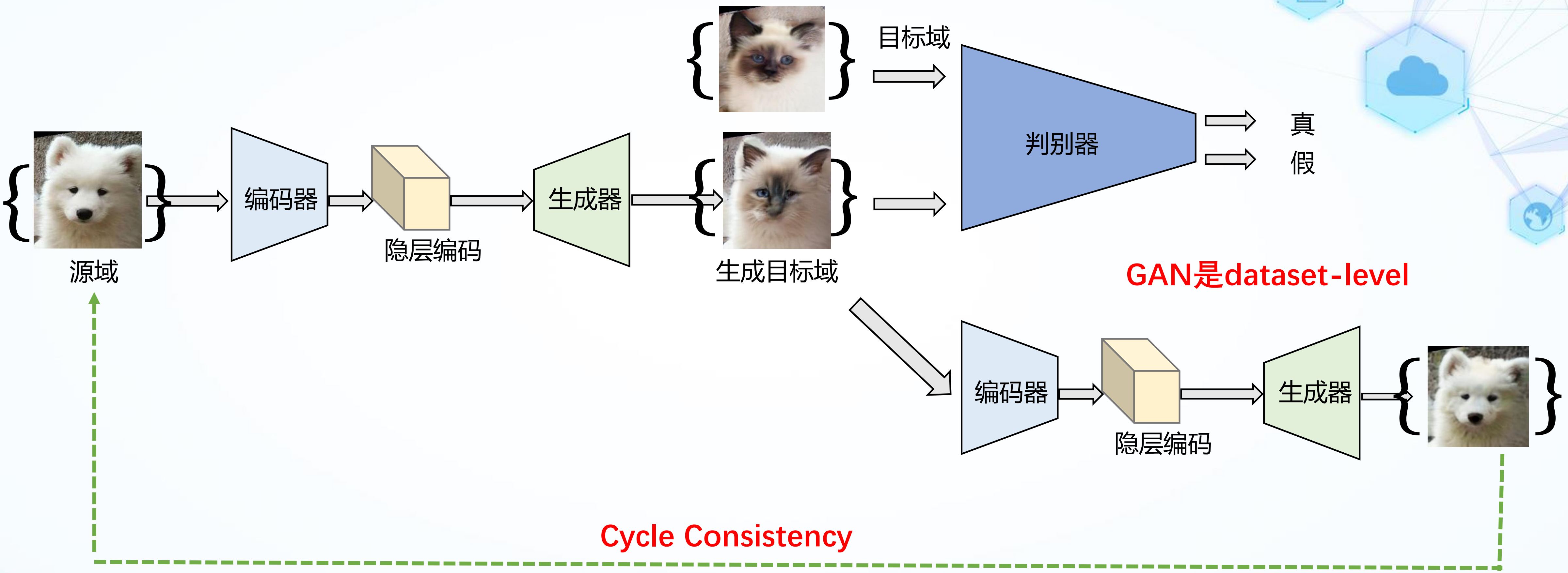
GAN能刻画dataset-level相似性

然而，存在无穷多的映射，无法捕捉不同域的instance-level的相似性关联



给定边缘分布P(X)和P(Y),
对应无穷多的联合分布P(X, Y)

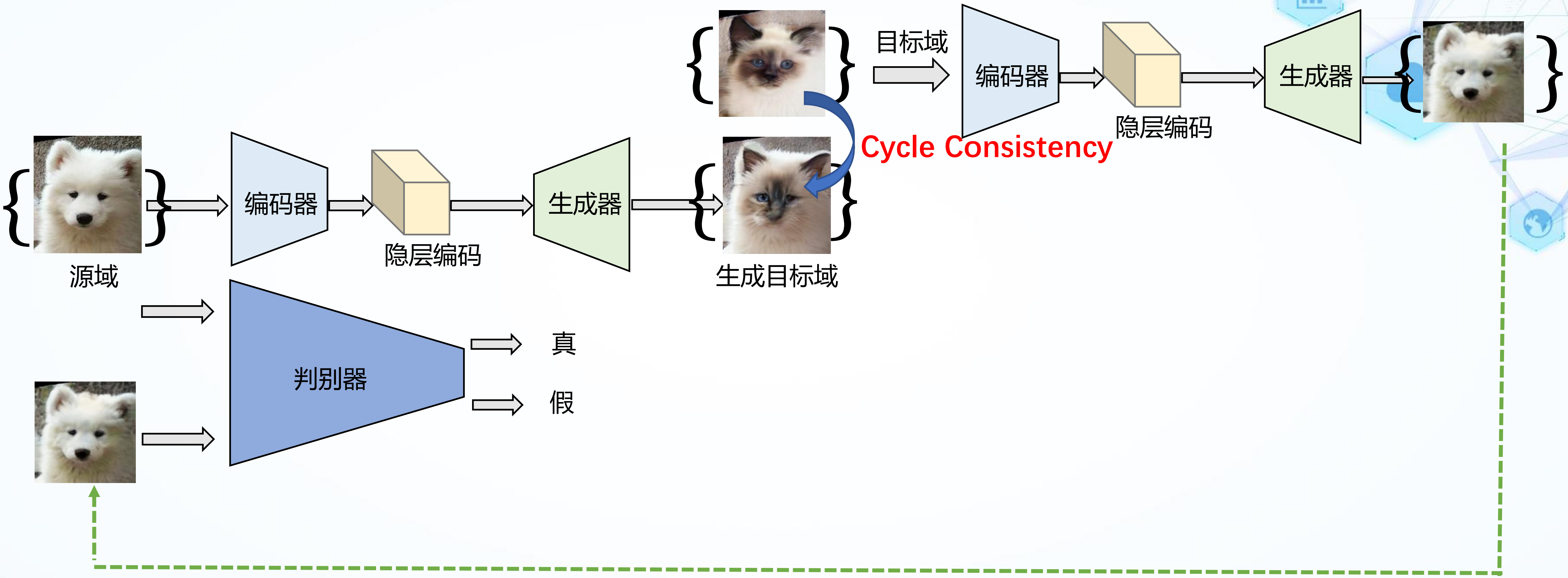
03 • Background： 从监督图片转换到无监督图片转换



Enforcing $\sum_{Y} P(X|Y)P(Y|X) = I$

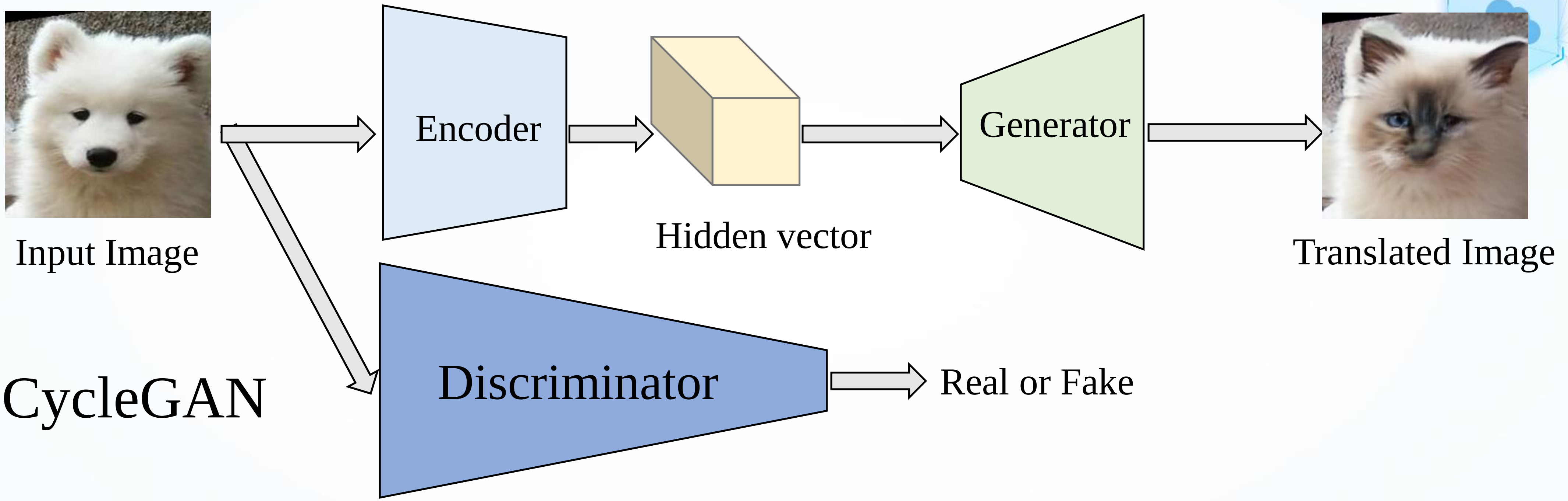
Zhu et al. ICCV 2017

03 • Background： 从监督图片转换到无监督图片转换



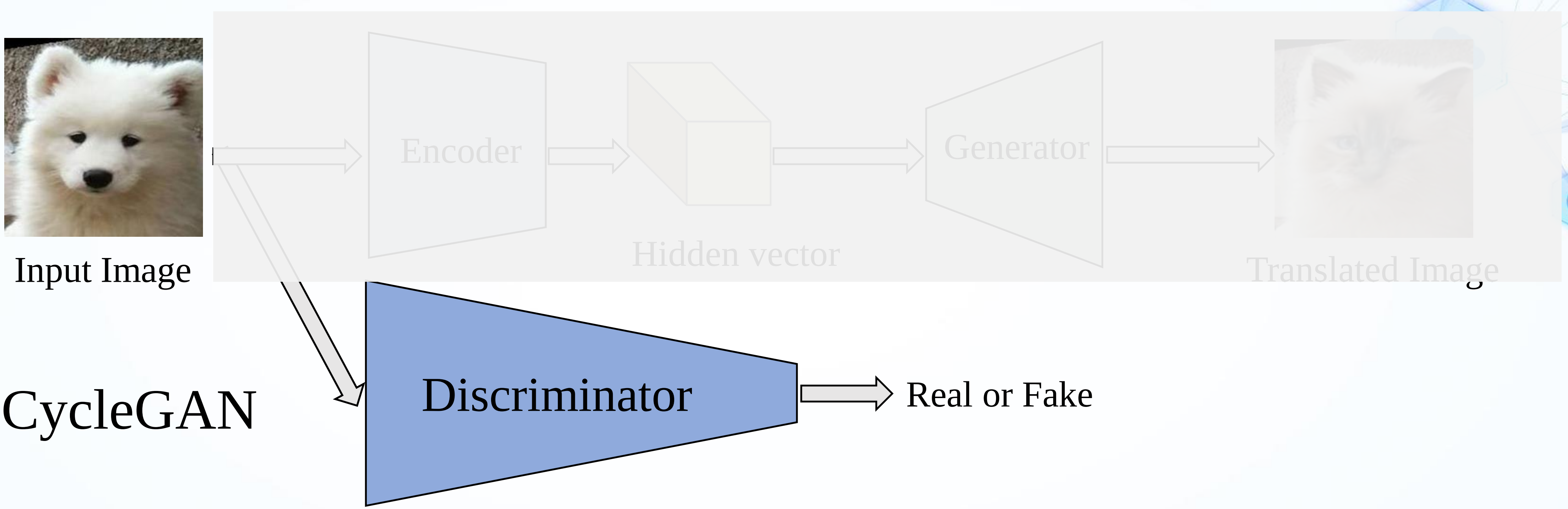
Zhu et al. ICCV 2017

04 • Background：无监督图片转换



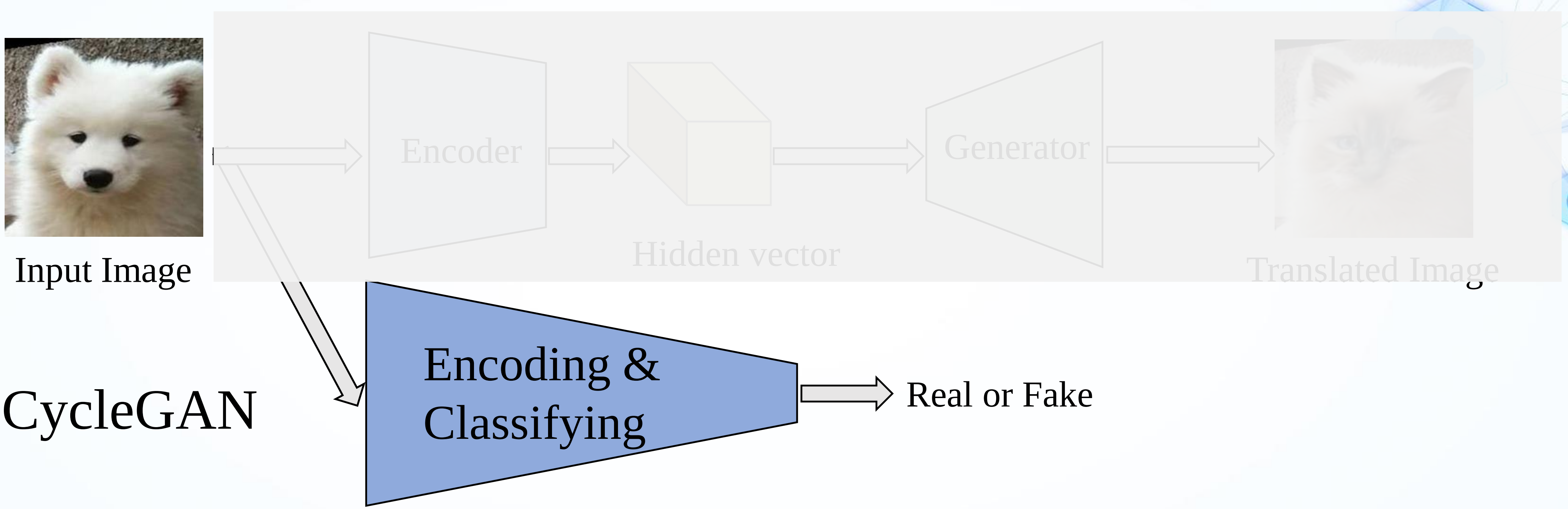
- ➢ **A Discriminator** for domain alignment by using GAN training.
- ➢ **An Encoder** to embed the input image to a low-dimension hidden space.
- ➢ **A Generator** to translate hidden vectors to images of the other domain

04 • Motivation： 判别器的角色是什么?



CycleGAN

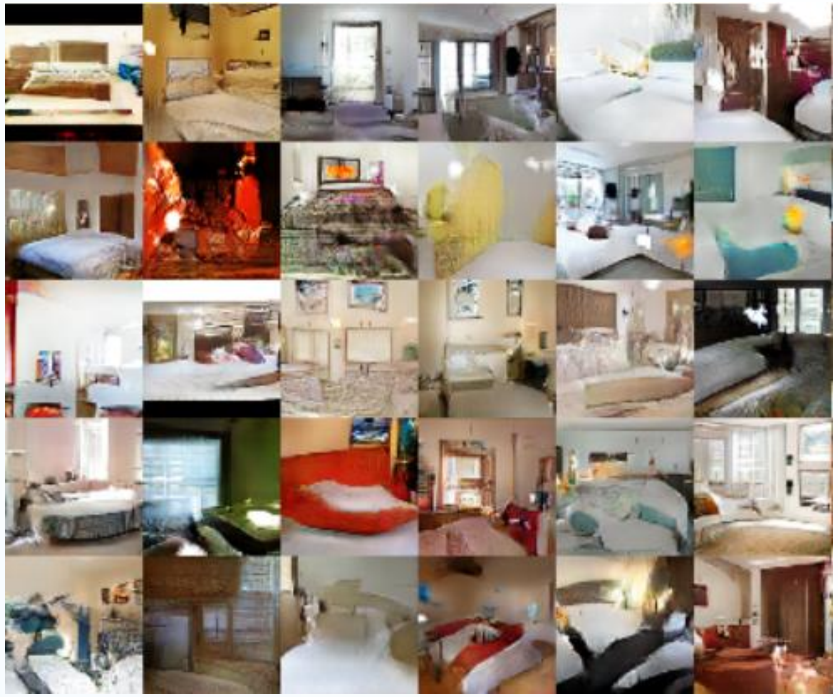
04 • Motivation：判别器的角色是什么？

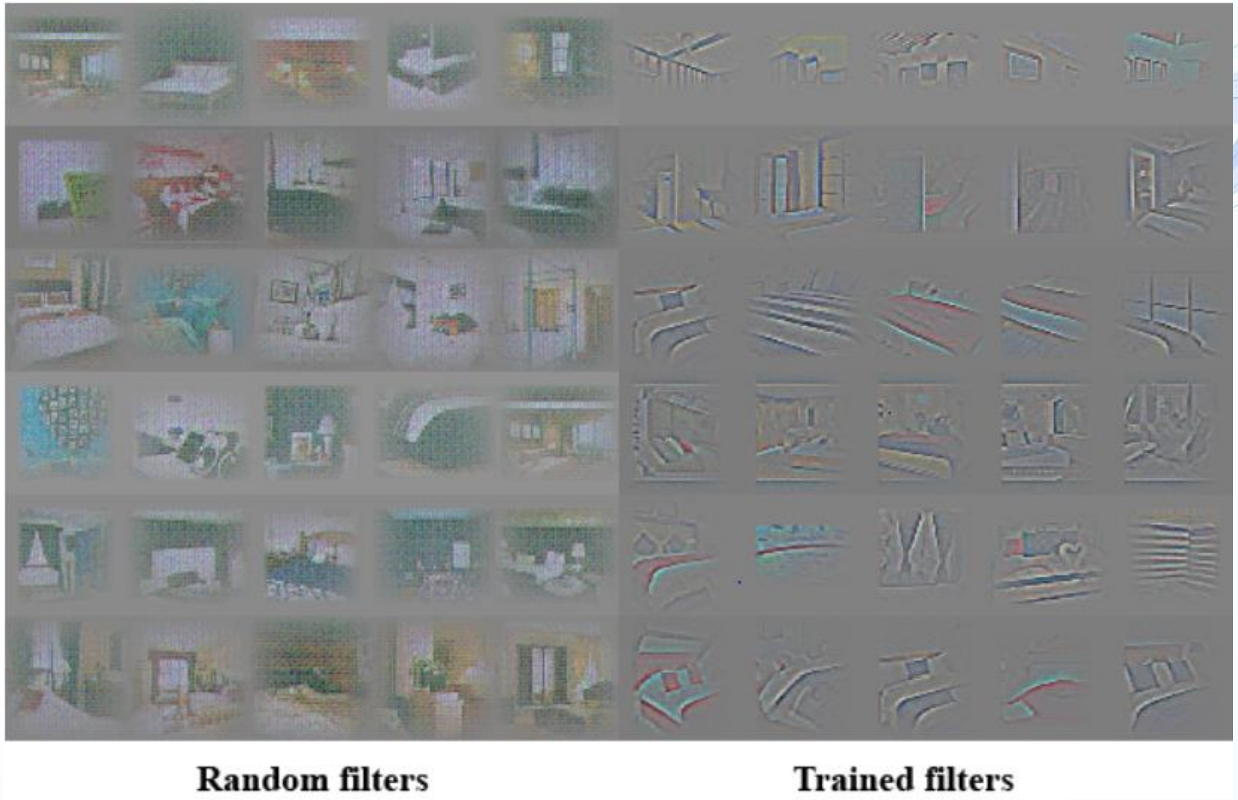


The discriminator should conduct semantics encoding of the input images before it can tell what images are true/false.

04 • Motivation：判别器的角色是什么?

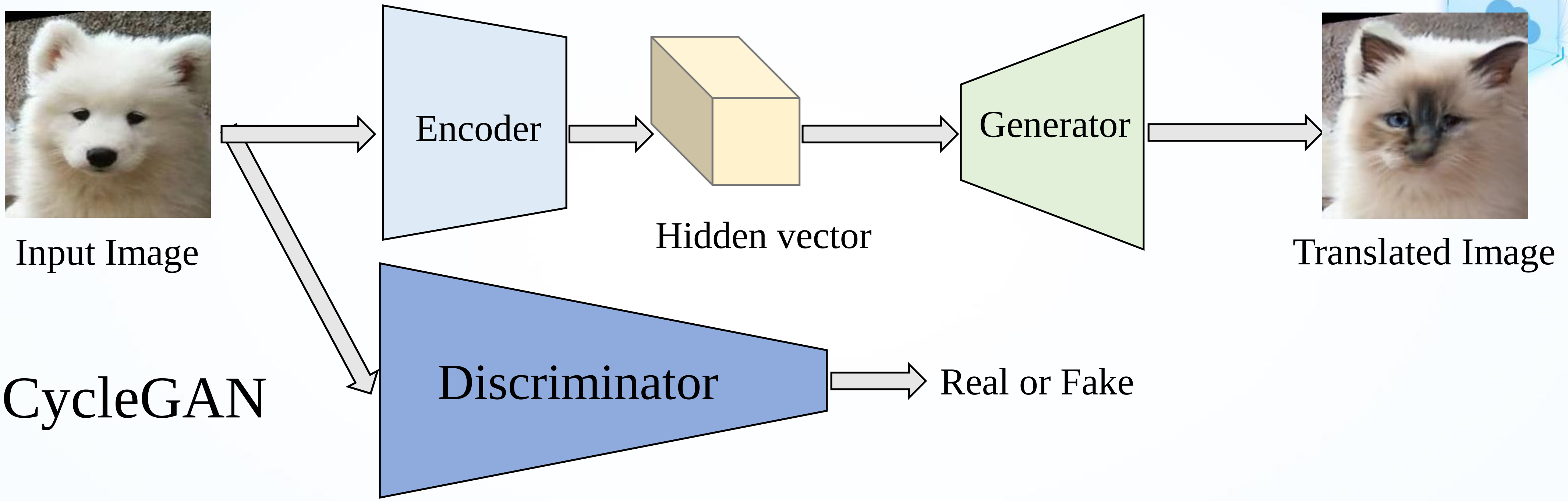
Radford ea al. DCGAN, ICLR 2016.





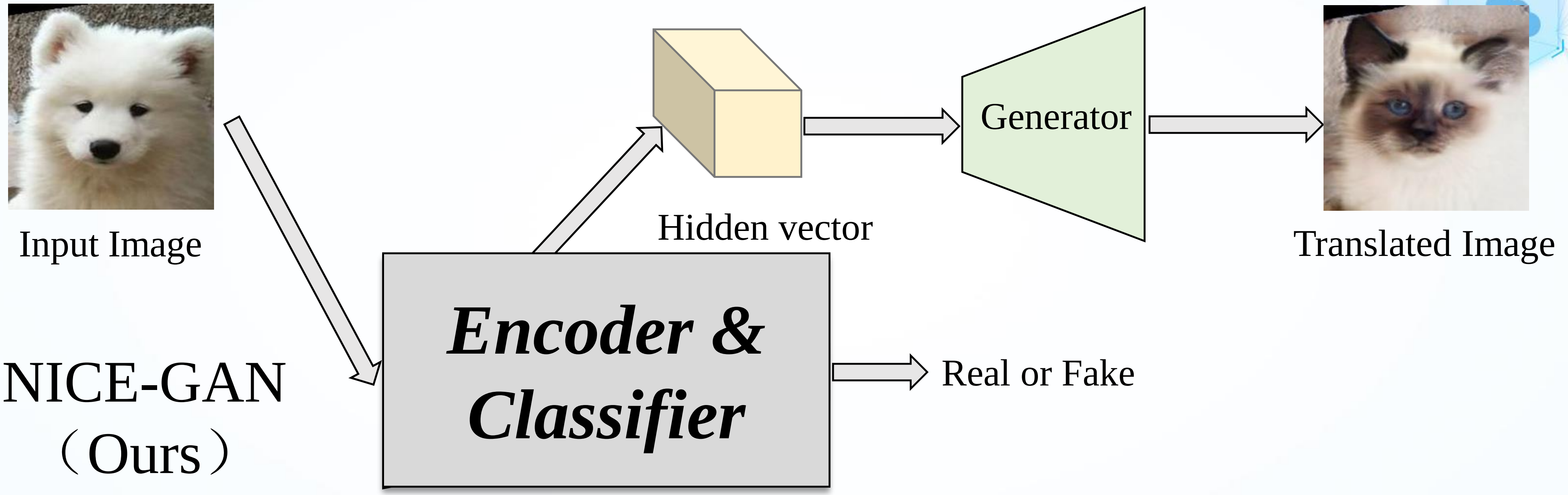
Strongly responses to the input image are observed in the first 6 layers of the discriminator after training.

04 • Motivation：判别器的角色是什么？



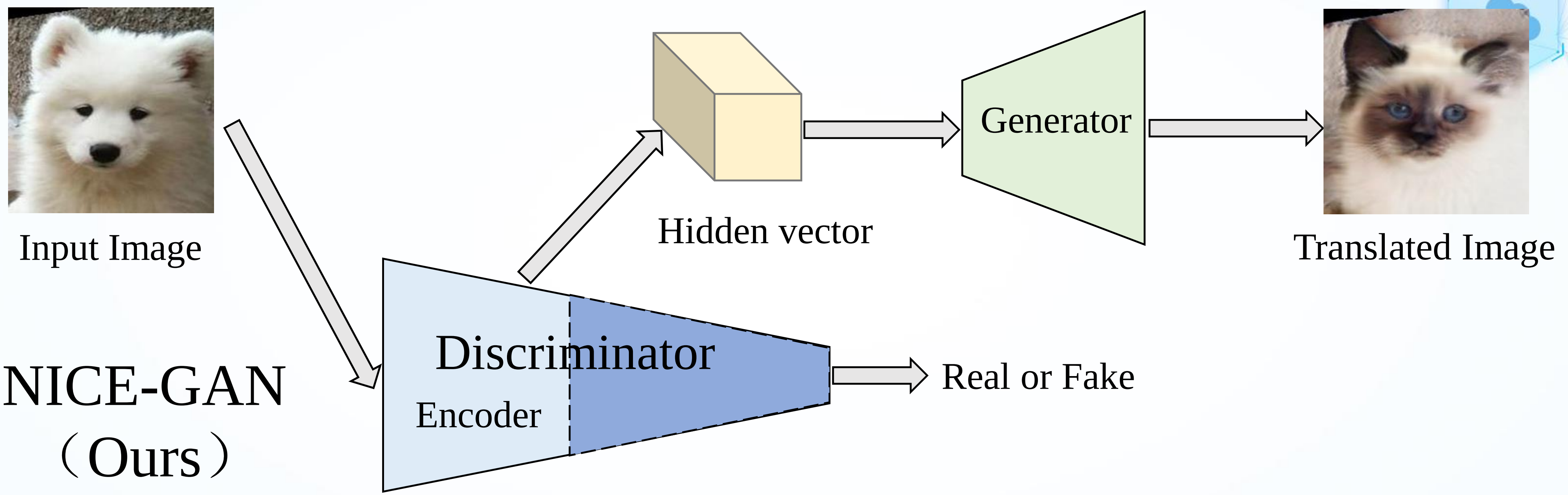
Can we integrate encoder into discriminator, since it has done the encoding thing

04 • Motivation：判别器的角色是什么？



- **A Discriminator** for domain alignment by using GAN training.
  Reusing early layers to embed the input image to a low-dimension hidden space.
- **A Generator** to translate hidden vectors to images of the other domain

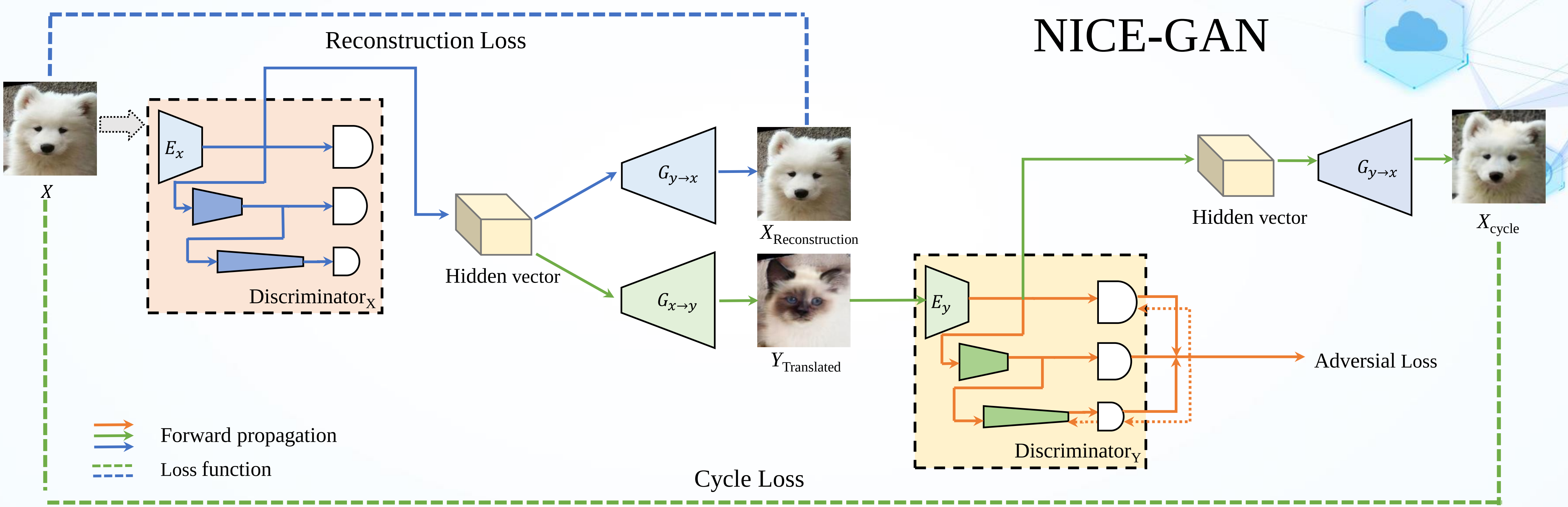
04 • Motivation：判别器的角色是什么?



NICE-GAN
（Ours）

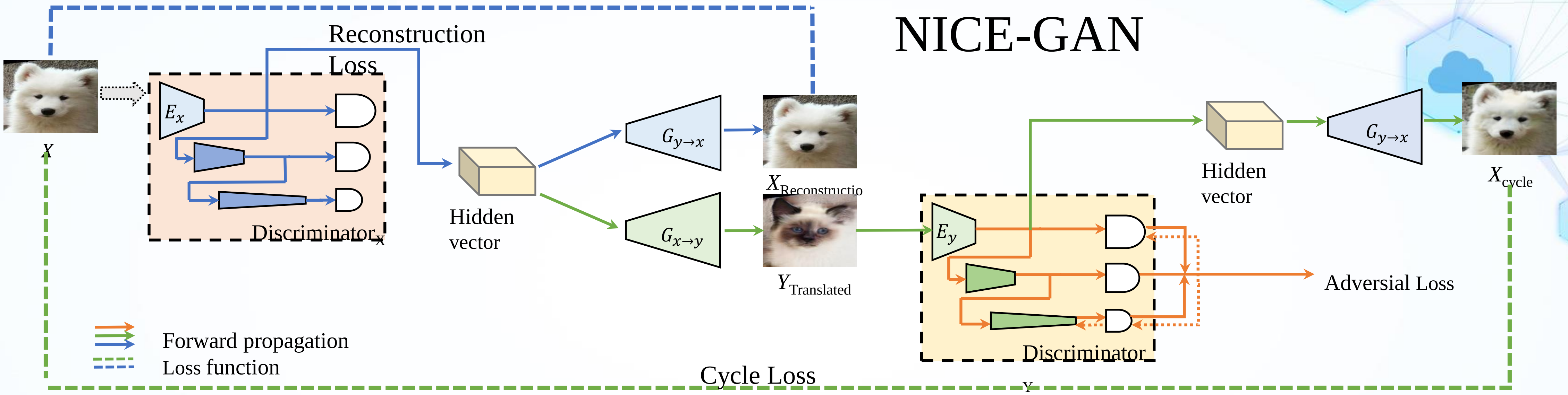
- ➢ **More compact architecture!**
- ➢ **More effective training of the encoder (directly using the discriminative loss)!**

05 • NICE-GAN (**N**o **I**ndependent **C**omponent for **E**ncoding)



- Reusing discriminator for encoding
- Multi-scale discriminator
- Self-reconstruction
- **Decoupled training**

05 • What and Why decoupled training?



$$\min_{E_x, G_{x\to y}} \max_{D_y = E_y \circ C_y} L_{gan}^{x\to y} + \min_{E_y, G_{y\to x}} \max_{D_x = E_x \circ C_x} L_{gan}^{y\to x}$$

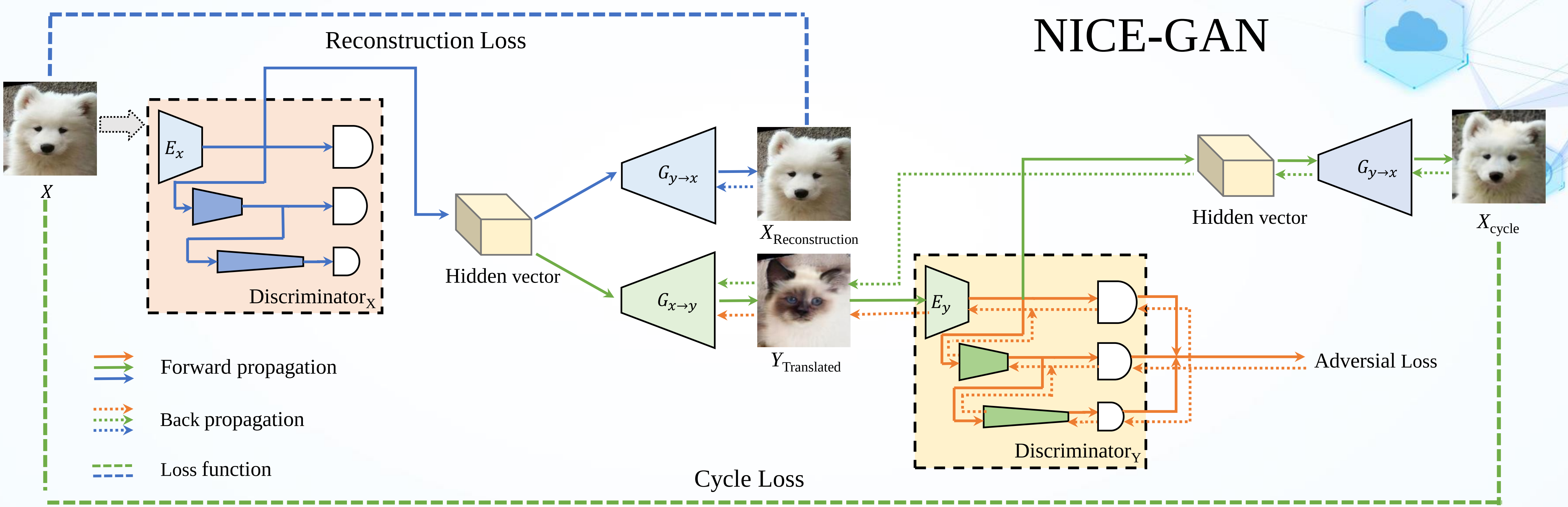
Decoupled Training

$$\min_{G_{x\to y}} \max_{D_y = E_y \circ C_y} L_{gan}^{x\to y} + \min_{G_{y\to x}} \max_{D_x = E_x \circ C_x} L_{gan}^{y\to x}$$

The encoders are minimized and maximized at the same time!!!

05

• The Overall Flowchart



Update Generator

Update iscriminator

06 • Experiments

- Comparisons with SOTAs (how does it perform quantitatively and qualitatively)
- Ablation Study (how does each component help)
- Analysis (why it works)

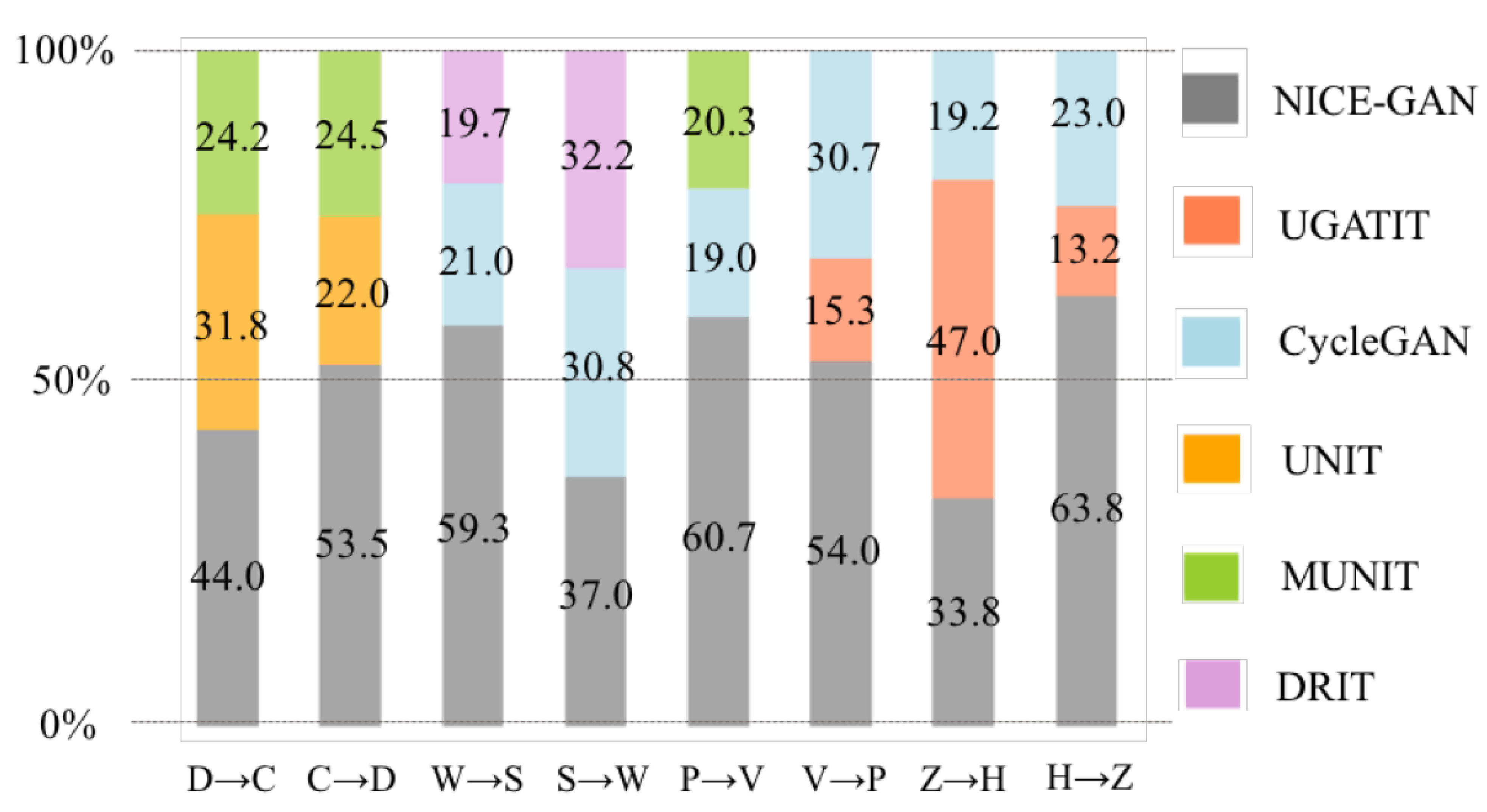
06

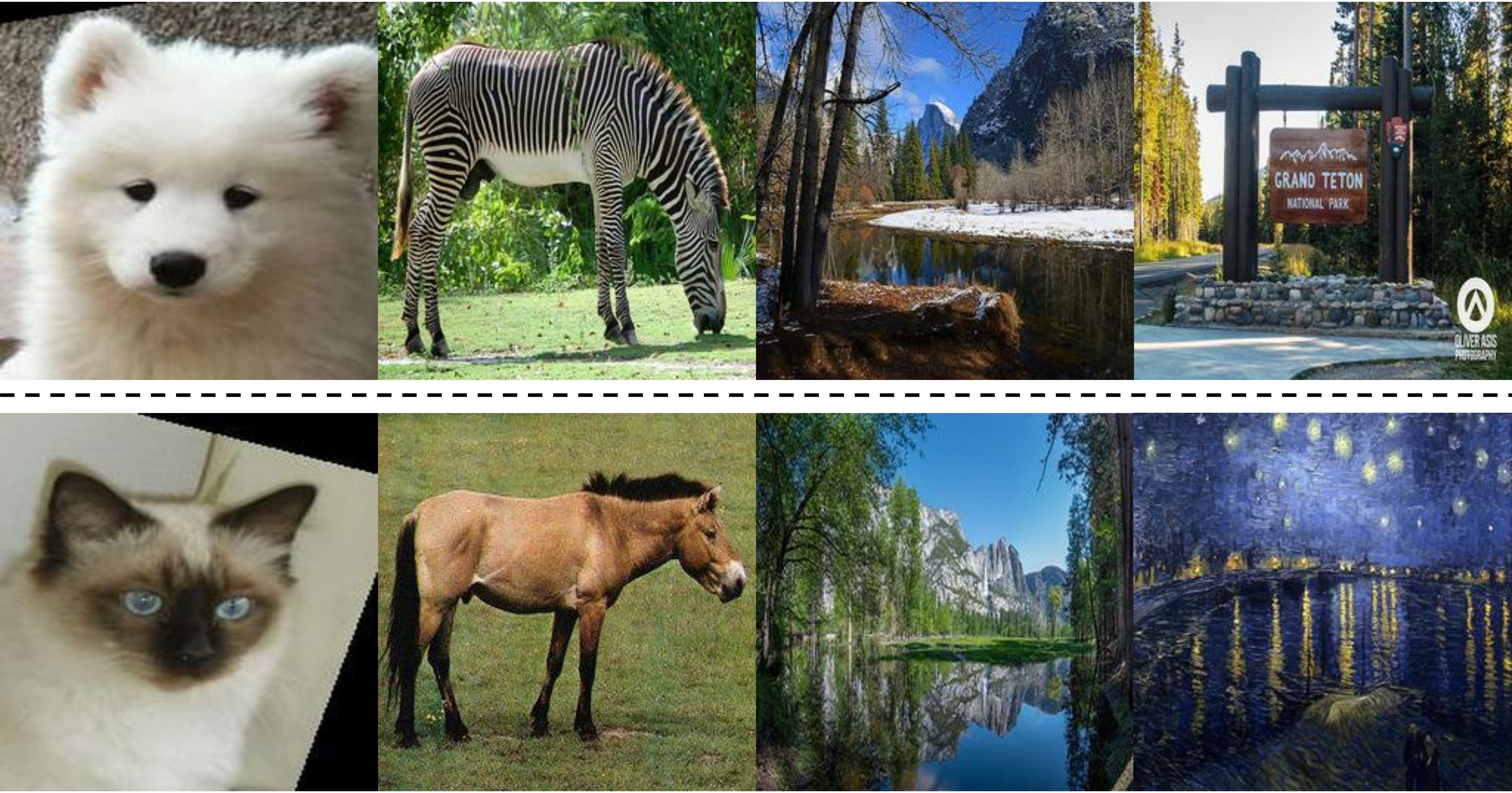
• Experiments: Comparison with SOTAs

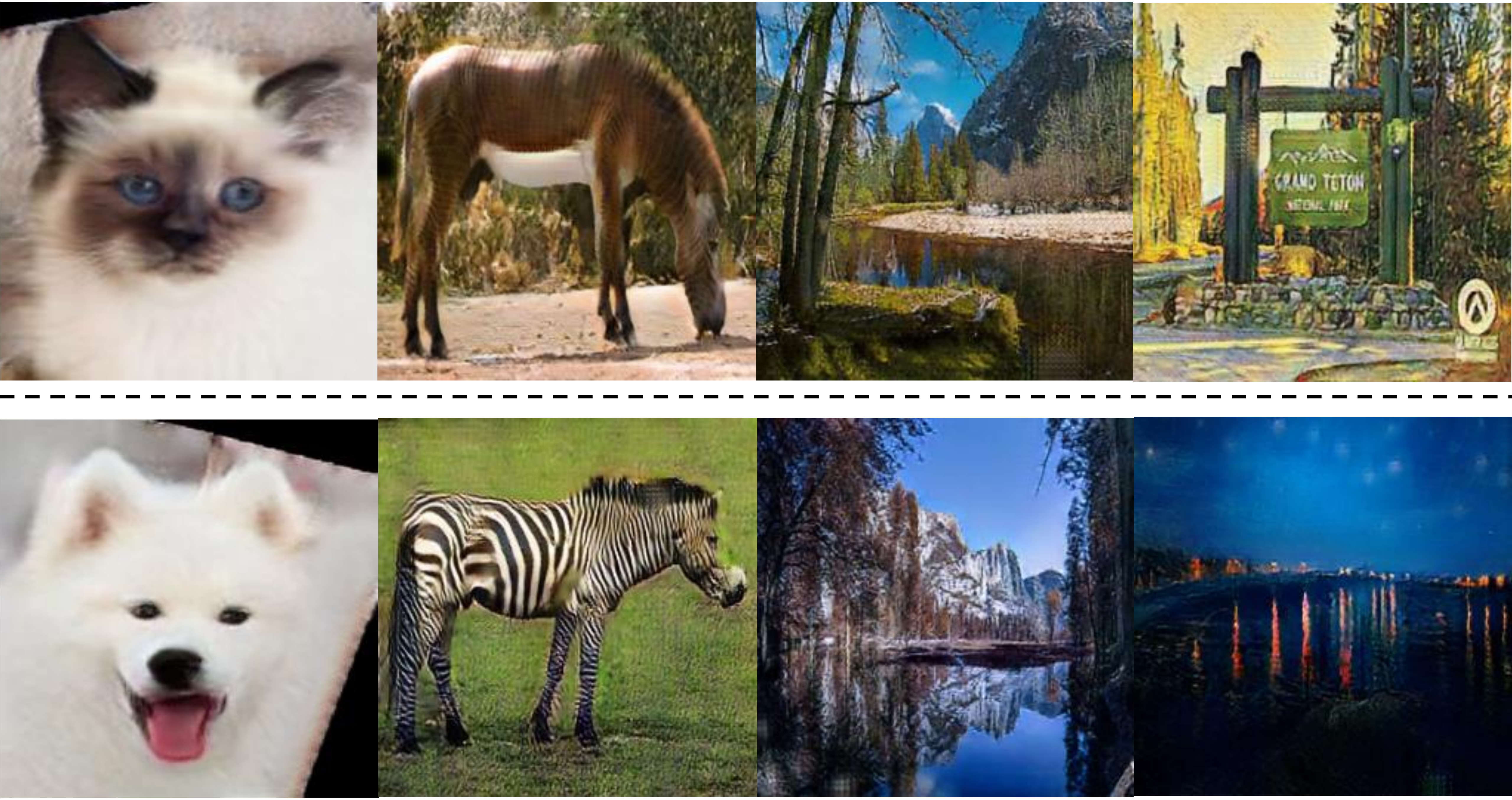
Method \ Dataset	dog → cat		winter → summer		photo → vangogh		zebra → horse	
	FID	KID × 100	FID	KID × 100	FID	KID × 100	FID	KID × 100
NICE-GAN	**48.79**	1.58	**76.44**	**1.22**	**122.27**	3.71	149.48	4.29
NICE-GAN*	51.98	1.50	79.02	1.35	122.59	**3.53**	150.57	4.43
U-GAT-IT-light	80.75	3.22	80.33	1.82	137.70	6.03	**145.47**	**3.39**
CycleGAN	119.32	4.93	79.58	1.36	136.97	4.75	156.19	5.54
UNIT	59.56	1.94	95.93	4.63	136.80	5.17	170.76	6.30
MUNIT	53.25	**1.26**	99.14	4.66	130.55	4.50	193.43	7.25
DRIT	94.50	5.20	78.61	1.69	136.24	5.43	200.41	10.12

Method \ Dataset	cat → dog		summer → winter		vangogh → photo		horse → zebra	
	FID	KID × 100	FID	KID × 100	FID	KID × 100	FID	KID × 100
NICE-GAN	**44.67**	**1.20**	**76.03**	**0.67**	**112.00**	**2.79**	**65.93**	**2.09**
NICE-GAN*	55.72	1.89	77.13	0.73	117.81	3.61	84.89	3.29
U-GAT-IT-light	64.36	2.49	88.41	1.43	123.57	4.91	113.44	5.13
CycleGAN	125.30	6.93	78.76	0.78	135.01	4.71	95.98	3.24
UNIT	63.78	1.94	112.07	5.36	143.96	7.44	131.04	7.19
MUNIT	60.84	2.42	114.08	5.27	138.86	6.19	128.70	6.92
DRIT	79.57	4.57	81.64	1.27	142.69	5.62	111.63	7.40

06 • Experiments: Human preference results

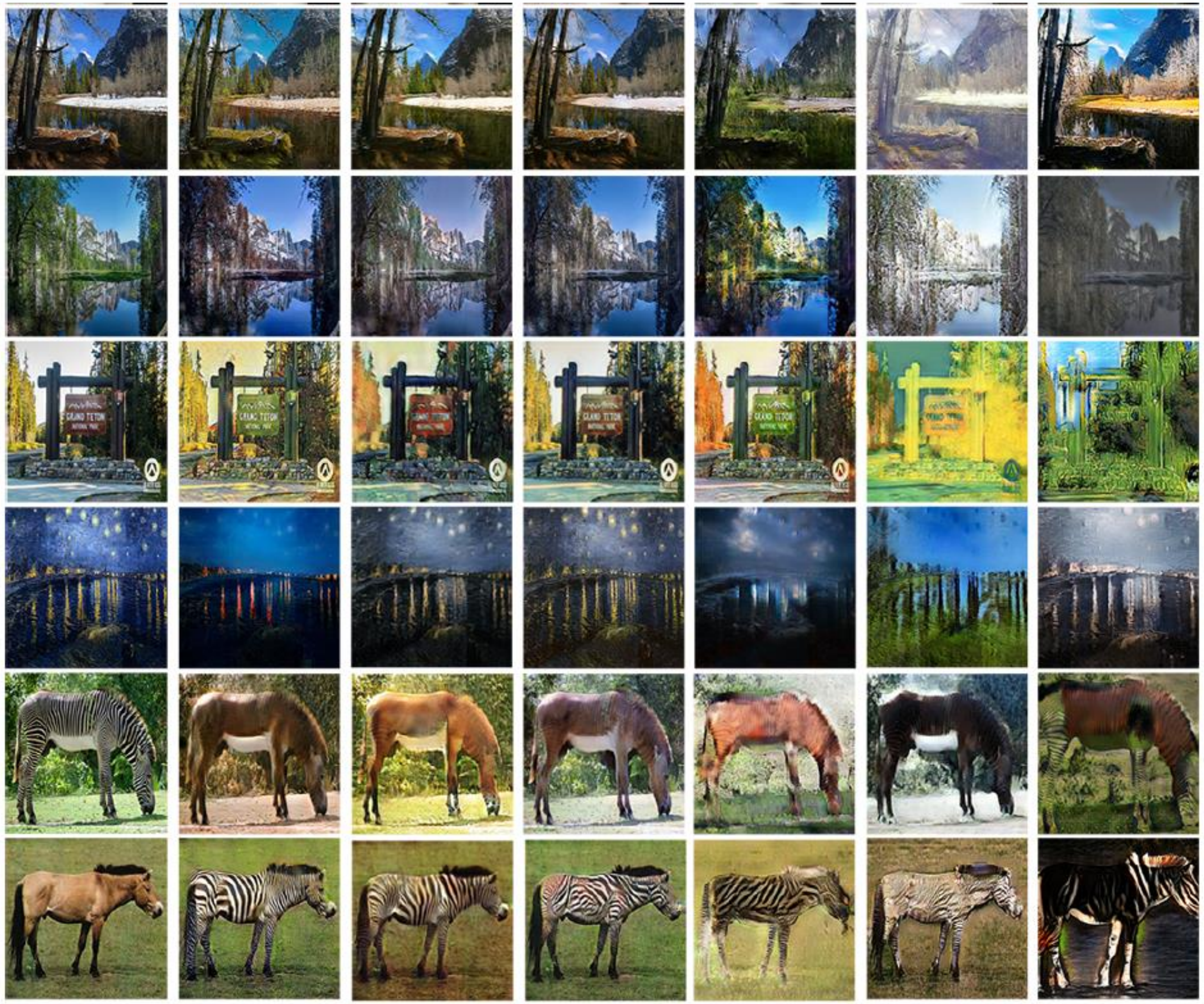



GRAND TETON
NATIONAL PARK
OLIVER ASIS
PHOTOGRAPHY


GRAND TETON

06

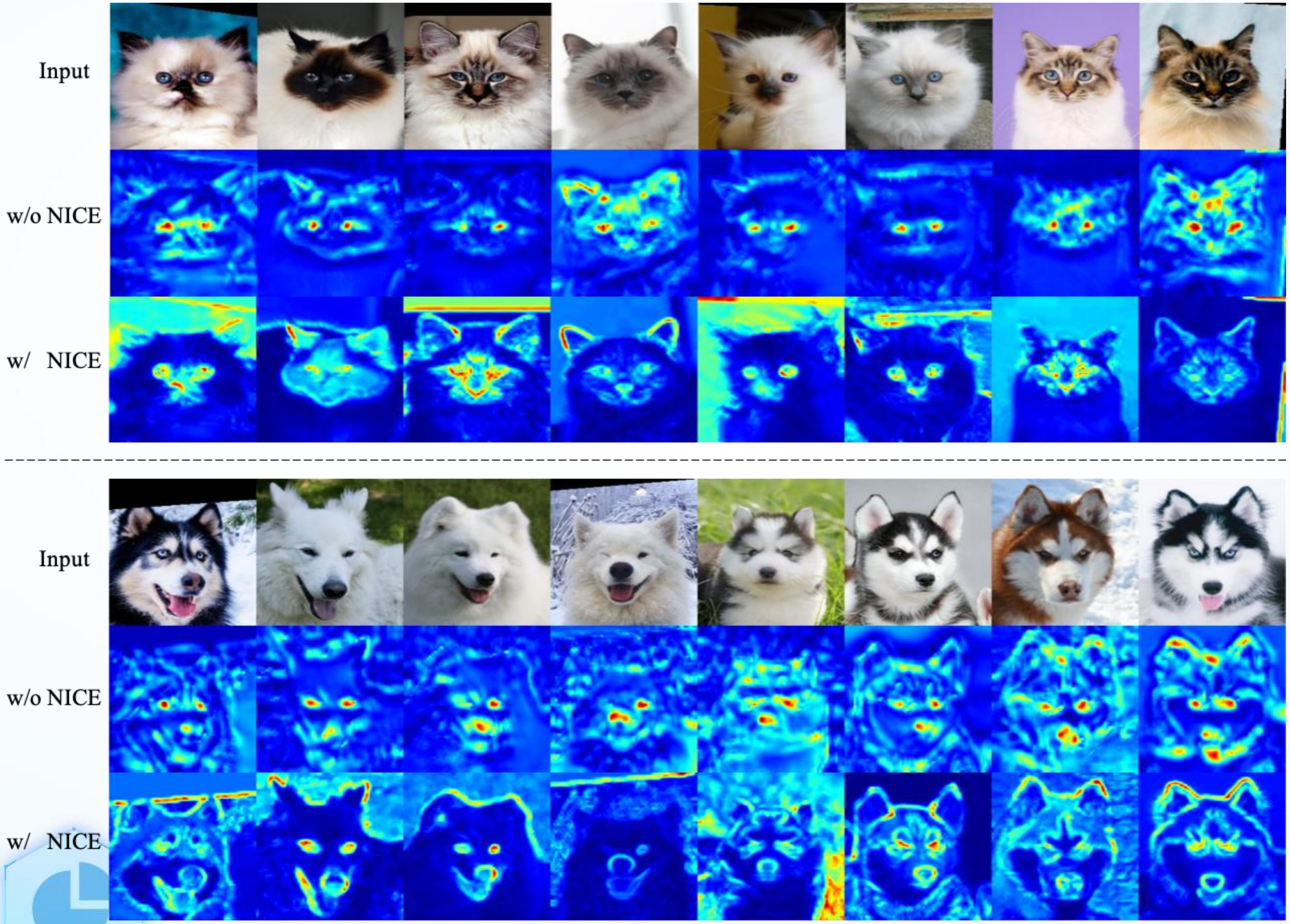
• Experiments: Visualization



a)Input b)NICE-GAN c)UGATIT d)CycleGAN e)UNIT f)MUNIT g)DRIT

06

• Experiments: Visualization



06

Experiments: More Compact Architecture

Table 2: Total number of parameters and FLOPs of network modules. NICE-GAN* are the version that the generator network is composed of only four residual blocks.

Method \ Module	Total number of params(FLOPs)	
	Generators	Discriminators
U-GAT-IT-light	21.2M(105.0G)	112.8M(15.8G)
NICE-GAN	16.2M(67.6G)	93.7M(12.0G)
NICE-GAN*	11.5M(48.2G)	93.7M(12.0G)

06 • Experiments: Ablation Study

Data Set	NICE	RA	C_x^0	C_x^1	C_x^2	FID	KID $\times$ 100
dog → cat	×	×	✓	✓	✓	80.75	3.22
	×	✓	✓	✓	✓	67.60	2.94
	✓	×	✓	✓	✓	63.80	3.27
	−	✓	✓	✓	✓	**48.55**	**1.23**
	✓	✓	✓	✓	✓	48.79	1.58
	+	✓	✓	✓	✓	53.52	1.84
	✓	✓	✓	✓	×	203.56	15.27
	✓	✓	✓	×	×	216.03	18.57
cat → dog	×	×	✓	✓	✓	64.36	2.49
	×	✓	✓	✓	✓	64.62	2.41
	✓	×	✓	✓	✓	51.49	1.68
	−	✓	✓	✓	✓	52.92	1.82
	✓	✓	✓	✓	✓	**44.67**	**1.20**
	+	✓	✓	✓	✓	54.90	2.17
	✓	✓	✓	✓	×	238.62	21.41
	✓	✓	✓	×	×	231.24	22.12

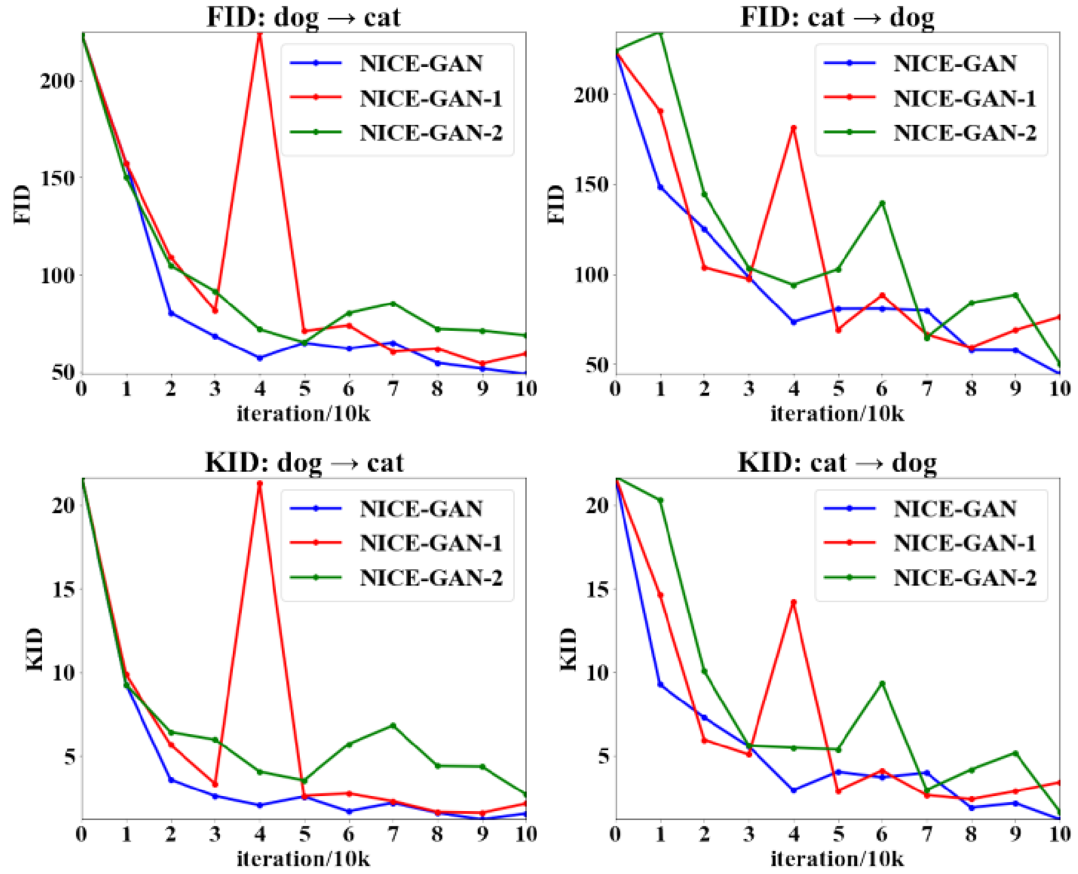
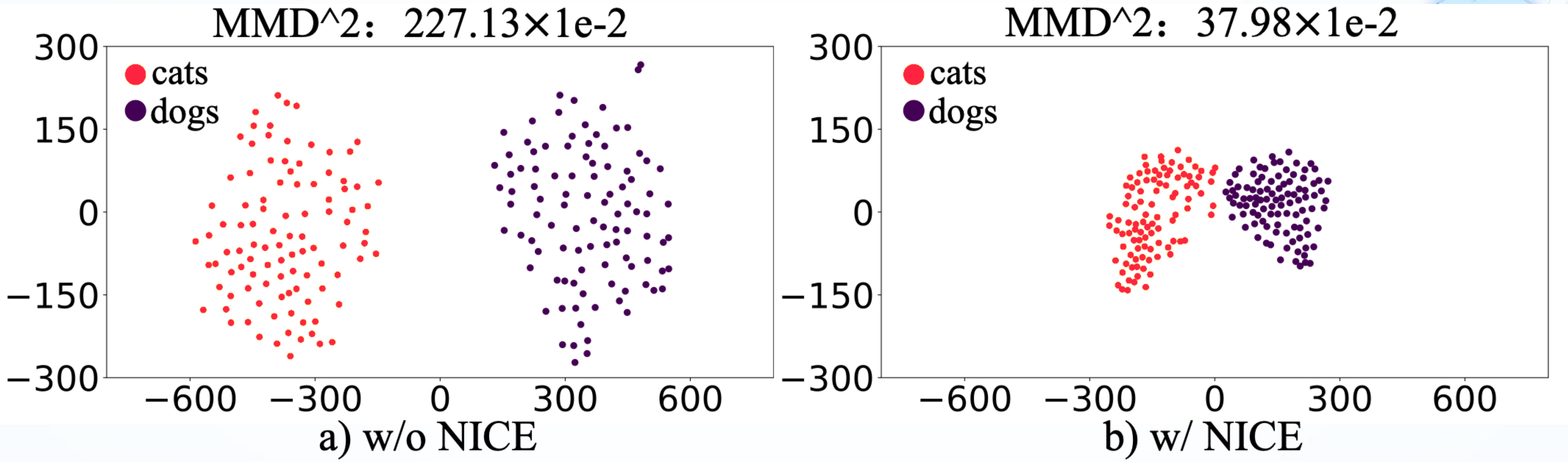


Figure 6: **Decoupled Training Analysis.** NICE-GAN:

06 • Experiments: Why it works?



- Shortening the transition path between domains in the latent space,
- NICE-GAN can probably facilitate domain translation in the image space.

07 • Conclusion

- This paper contends a novel role of the discriminator by reusing it for encoding the images of the target domain.

- We develop a decoupled training strategy by which the encoder is only trained when maximizing the adversary loss while keeping frozen otherwise.

- Extensive experiments on four popular benchmarks demonstrate the superior performance of NICE-GAN over state-of-the-art methods in terms of FID, KID, and also human preference.

谢 谢 观 看

THANK YOU