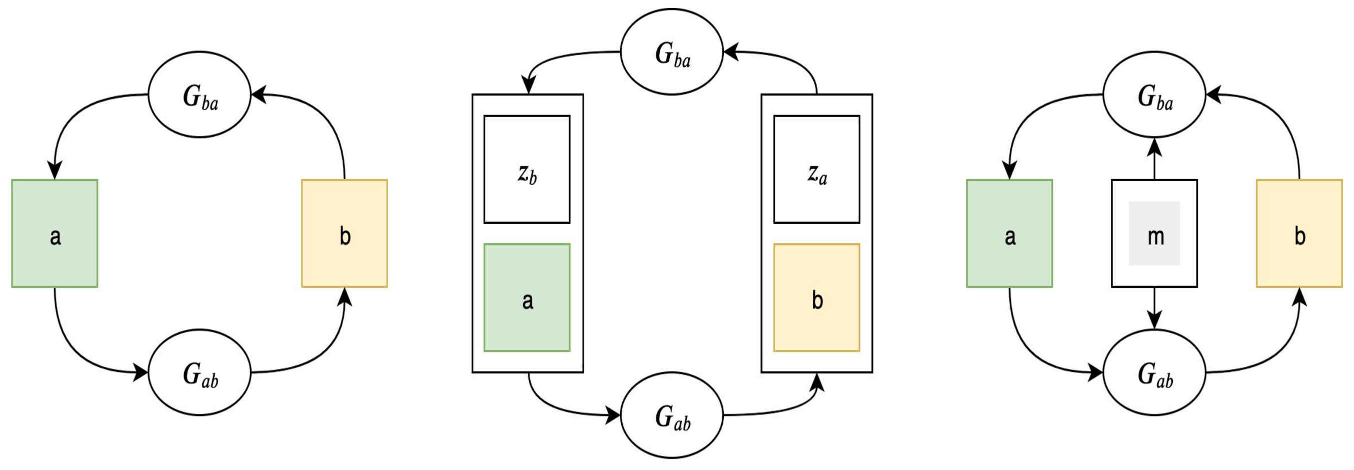
Mask CycleGAN: **Unpaired** Multi-modal Domain Translation with **Interpretable** Latent Variable

Minfa Wang (minfa@stanford.edu)

Demo: **bit.ly/mask_cgan**

Stanford Computer Science

Introduction



(a) CycleGAN (b) Augmented CycleGAN (c) Mask CycleGAN

CycleGAN is a popular approach for unpaired image-to-image translation between two domains. It has the limitation that the generator is **deterministic** w.r.t. input image.

People attempted to address this limitation by introducing latent variables typically modeled by multivariate Gaussian. However, it is **lack of interpretability**.

Mask CycleGAN aims to address both issues above by using **pixel mask** as latent variables. Its formulation is a full generalization of CycleGAN, and hence is at least equally expressive.

Notation

a: image from domain A
b: image from domain B
m: pixel mask
G_{AB} : generator mapping a to b
$\tilde{b} = G_{AB}(a, m)$: fake b
$a' = G_{BA}(\tilde{b})$: recovered a
$G_{BA}(a, m)$: same a

Technical Methods

The architecture is based on CycleGAN. **Our contributions** come from design of 3 components: *loss*, *mask* and *generator*.

GAN Loss

$$\mathcal{L}_{GAN}^{AF} = -\mathbb{E}_{a\sim A}[\log D_{AF}(a)] - \mathbb{E}_{\tilde{a}\sim\tilde{A}}[\log(1 - D_{AF}(\tilde{a}))]$$
$$\mathcal{L}_{GAN}^{AM} = -\mathbb{E}_{a\sim A}[\log D_{AM}(a \odot m)] - \mathbb{E}_{\tilde{a}\sim\tilde{A}}[\log(1 - D_{AM}(\tilde{a} \odot m))]$$
$$\mathcal{L}_{GAN}^{A} = \lambda_{GAN}^{M}\mathcal{L}_{GAN}^{AM} + (1 - \lambda_{GAN}^{M})\mathcal{L}_{GAN}^{AF}$$

Cycle Loss

$$\mathcal{L}_{CYC}^{A} = \lambda_{CYC}^{M}||(a - a') \odot m||_1 + (1 - \lambda_{CYC}^{M})||(a - a') \odot (1 - m)||_1$$

Identity Loss

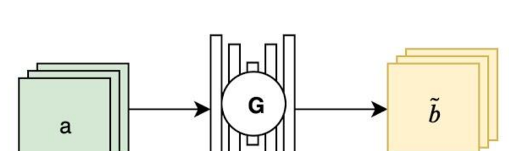
$$\mathcal{L}_{IDT}^{A} = ||a - G_{BA}(a, m)||_1$$

Mask is a pixel map of same shape as the image. We tried 2 masking schemes with results shown on the "Qualitative Results" section:

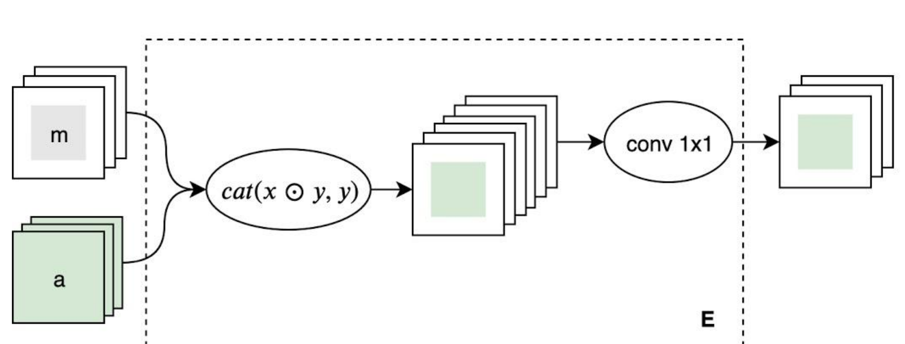
1. square, centered, size = 0.5, 0.8 and 1.0 of image size
2. heuristic: multiple squares, random position and size

In future work, we will experiment with *soft attention mask*.

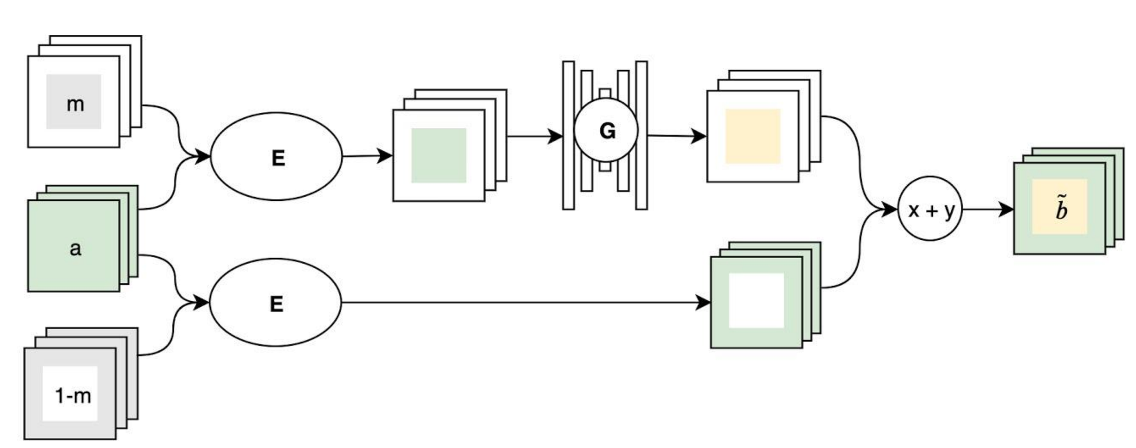
Generator has **encoders** that enforces **linear** interactions between the masked image and the mask, ensuring that the main generator network to only depend on masked region.



(a) Typical generator in CycleGAN



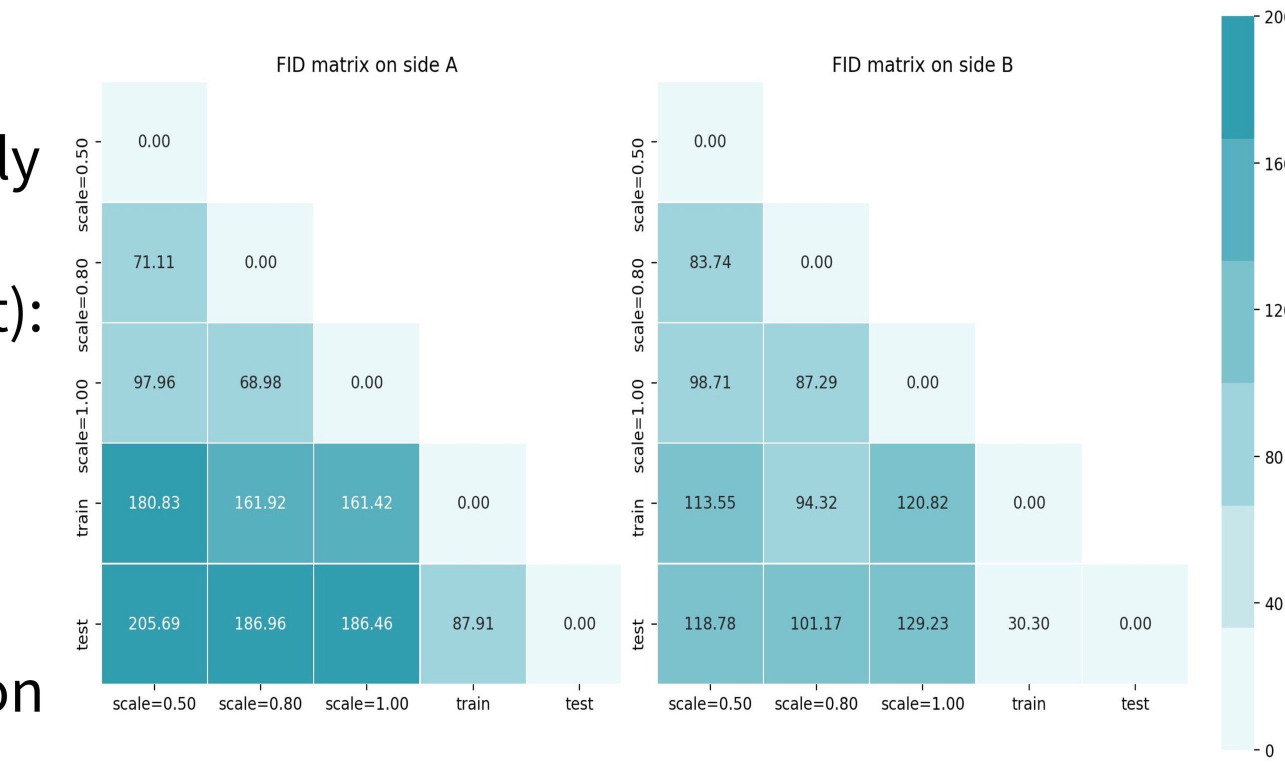
(b) Expanded view of Mask Encoder



(c) Generator in Mask CycleGAN

Quantitative Results

- Evaluated on horse2zebra dataset.
- scale=1.0 is **baseline**, approximately original CycleGAN
- FID_A(train, test) > FID_B(train, test): more variations in horses
- FID_A(scale, train) > FID_B(scale, train): horses are harder to fit
- FID_B(scale=0.8, test) < FID_B(scale=1.0, test): regularization



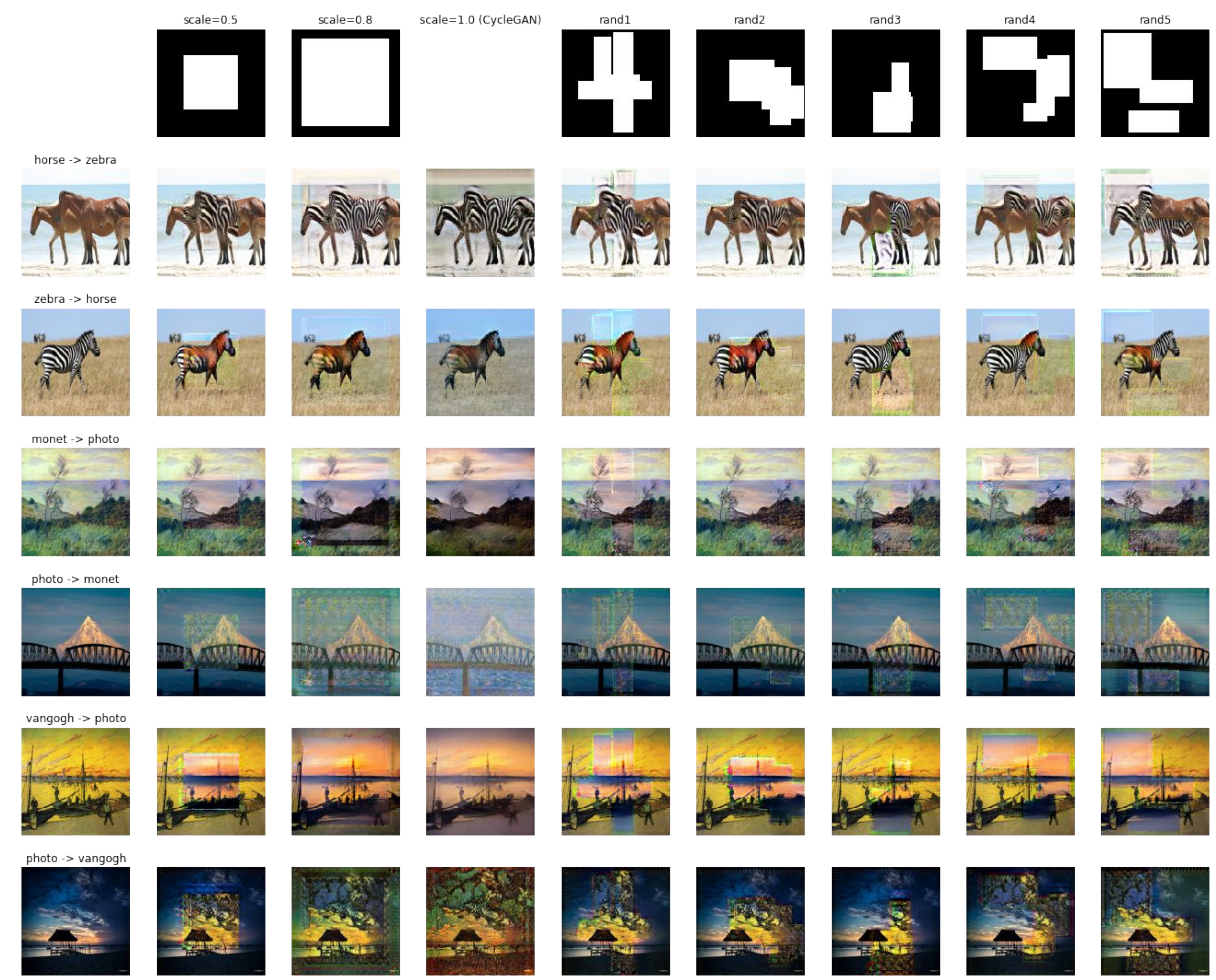
FID matrix on side A

	scale=0.50	scale=0.80	scale=1.00	train	test
scale=0.50	0.00				
scale=0.80	71.11	0.00			
scale=1.00	97.96	68.98	0.00		
train	180.83	161.92	161.42	0.00	
test	205.69	186.96	186.46	87.91	0.00

FID matrix on side B

	scale=0.50	scale=0.80	scale=1.00	train	test
scale=0.50	0.00				
scale=0.80	83.74	0.00			
scale=1.00	98.71	87.29	0.00		
train	113.55	94.32	120.82	0.00	
test	118.78	101.17	129.23	30.30	0.00

Qualitative Results



Key References

1. J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks.CoRR.
2. A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. C. Courville. Augmented cyclegan:Learning many-to-many mappings from unpaired data.CoRR,