Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Jun-Yan Zhu* Taesung Park* Phillip Isola Alexei A. Efros

Berkeley AI Research (BAIR) laboratory, UC Berkeley

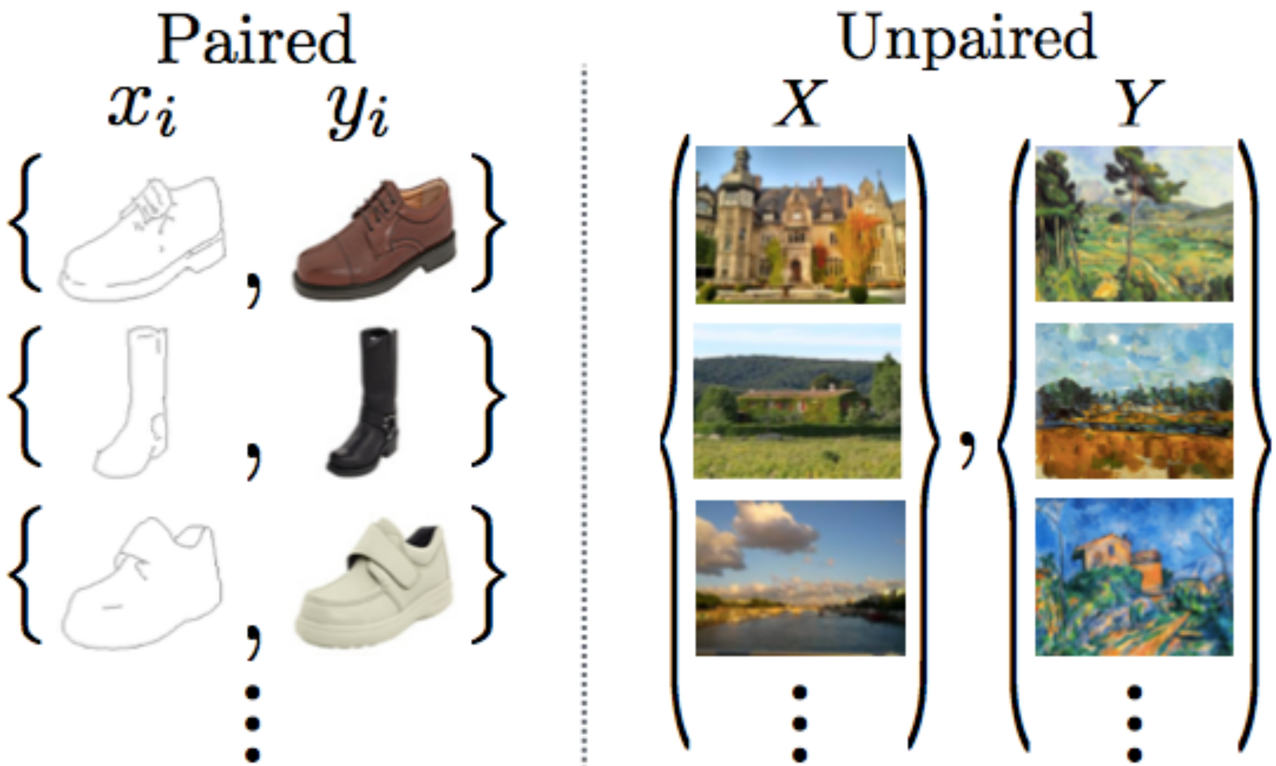
In CVPR 2017

Yufeng Jiang

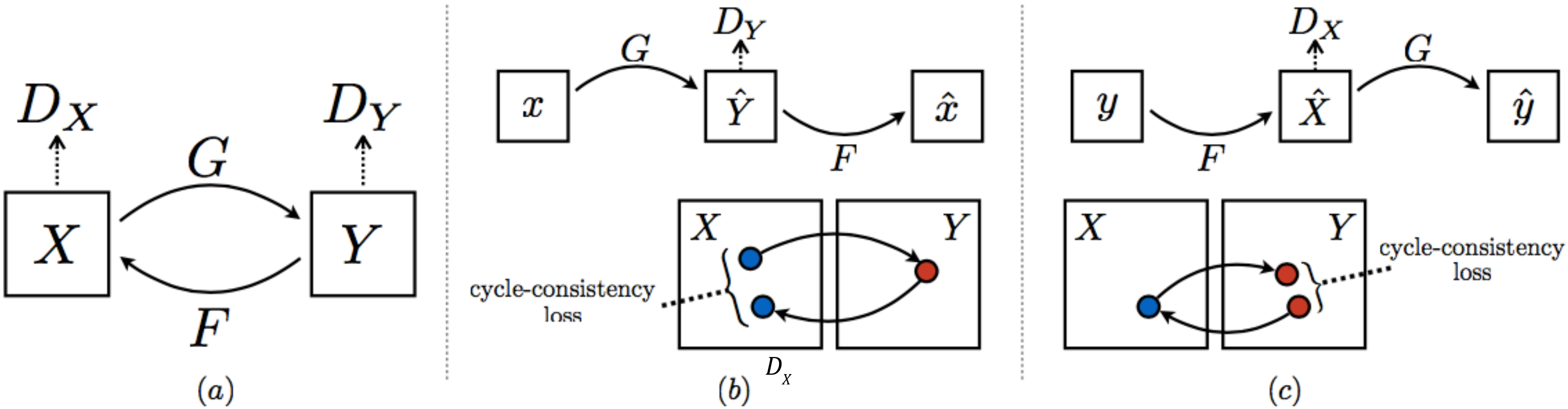
2018.09.10

Motivation

- Paired datasets are rare.
- Images from two different fields are hard to match.



Approach



Formula variable definition：

Domain: X, Y

Image sets of domain: {x}, {y}

Map: G: X -> Y

Map: F： Y -> X

D_X: discriminate {x} and {F(y)}

D_Y: discriminate {y} and {G(x)}

Loss function

- Adversarial loss

G: X -> Y Discriminator： D_Y

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]$$

$$\min_G \max_{D_Y} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y)$$

F: Y -> X Discriminator: D_X

$$\min_F \max_{D_X} \mathcal{L}_{\text{GAN}}(F, D_X, Y, X)$$

Loss function

- Cycle consistency loss

Forward cycle consistency： x -> G(x) –> F(G(x)) ≈ x

Backward cycle consistency: y -> F(y) -> G(F(y)) ≈ y

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

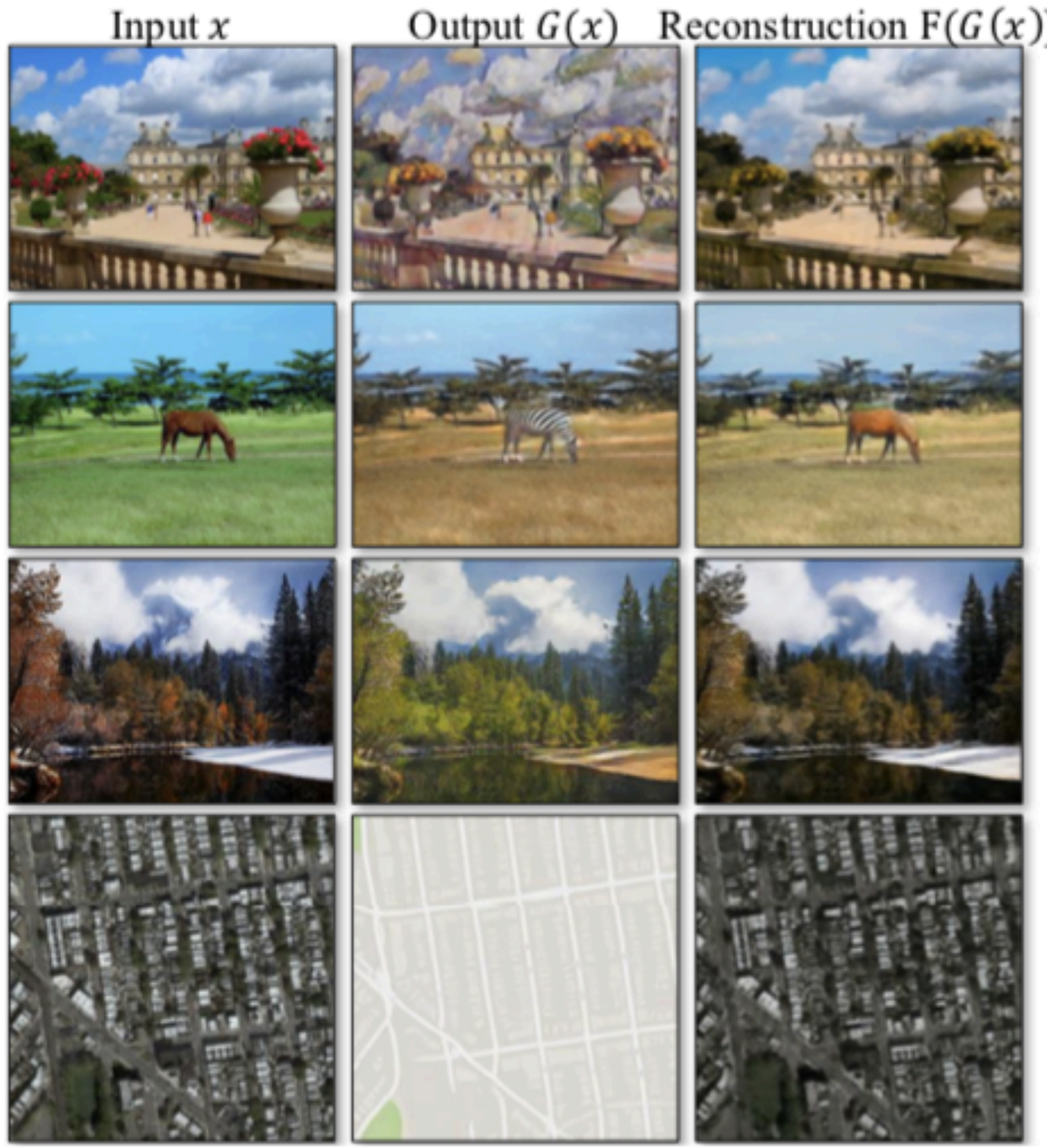
Use cycle consistency loss

Photo <-> Cezanne

Horses <-> Zebras

Winter <-> Summer Yosemite

Photos <-> Google maps



Full objective

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y)\\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X)\\ &+ \lambda\mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

$$G^*, F^* = \arg\min_{G,F}\max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y).$$

Implementation

- Network Architecture

  Generator： two stride-2 convolutions
  several residual blocks
  two stride-1/2 convolutions
  instance normalization

  Discriminator: 70 × 70 PatchGAN

Evaluation

- Evaluation Metrics
  - AMT perceptual studies
  - FCN score
  - Semantic segmentation metrics
- Baselines
  - CoGAN
  - SimGAN
  - Feature loss + GAN
  - BiGAN/ALI
  - pix2pix

Comparison against baselines

Loss	Map → Photo % Turkers labeled *real*	Photo → Map % Turkers labeled *real*
CoGAN [32]	$0.6\% \pm 0.5\%$	$0.9\% \pm 0.5\%$
BiGAN/ALI [9, 7]	$2.1\% \pm 1.0\%$	$1.9\% \pm 0.9\%$
SimGAN [46]	$0.7\% \pm 0.5\%$	$2.6\% \pm 1.1\%$
Feature loss + GAN	$1.2\% \pm 0.6\%$	$0.3\% \pm 0.2\%$
CycleGAN (ours)	$\mathbf{26.8\%} \pm \mathbf{2.8\%}$	$\mathbf{23.2\%} \pm \mathbf{3.4\%}$

Table 1: AMT "real vs fake" test on maps↔aerial photos at 256 × 256 resolution.

Comparison against baselines

Loss	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN [32]	0.40	0.10	0.06
BiGAN/ALI [9, 7]	0.19	0.06	0.02
SimGAN [46]	0.20	0.10	0.04
Feature loss + GAN	0.06	0.04	0.01
CycleGAN (ours)	**0.52**	**0.17**	**0.11**
pix2pix [22]	0.71	0.25	0.18

Table 2: FCN-scores for different methods, evaluated on Cityscapes labels→photo.

Comparison against baselines

Loss	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN [32]	0.45	0.11	0.08
BiGAN/ALI [9, 7]	0.41	0.13	0.07
SimGAN [46]	0.47	0.11	0.07
Feature loss + GAN	0.50	0.10	0.06
CycleGAN (ours)	**0.58**	**0.22**	**0.16**
pix2pix [22]	0.85	0.40	0.32

Table 3: Classification performance of photo→labels for different methods on Cityscapes.

Comparison against baselines

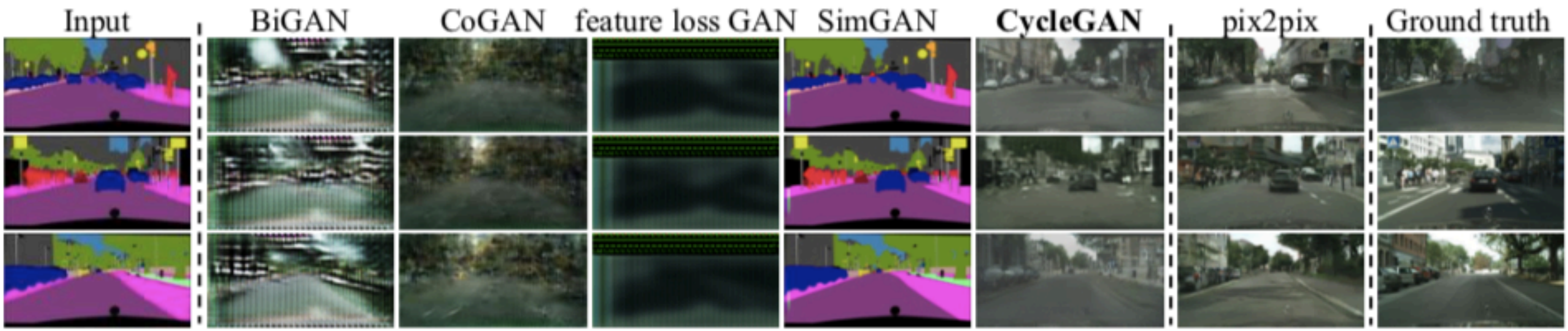


Figure 5:
Different methods for mapping labels <-> photos trained on Cityscapes images.

Comparison against baselines

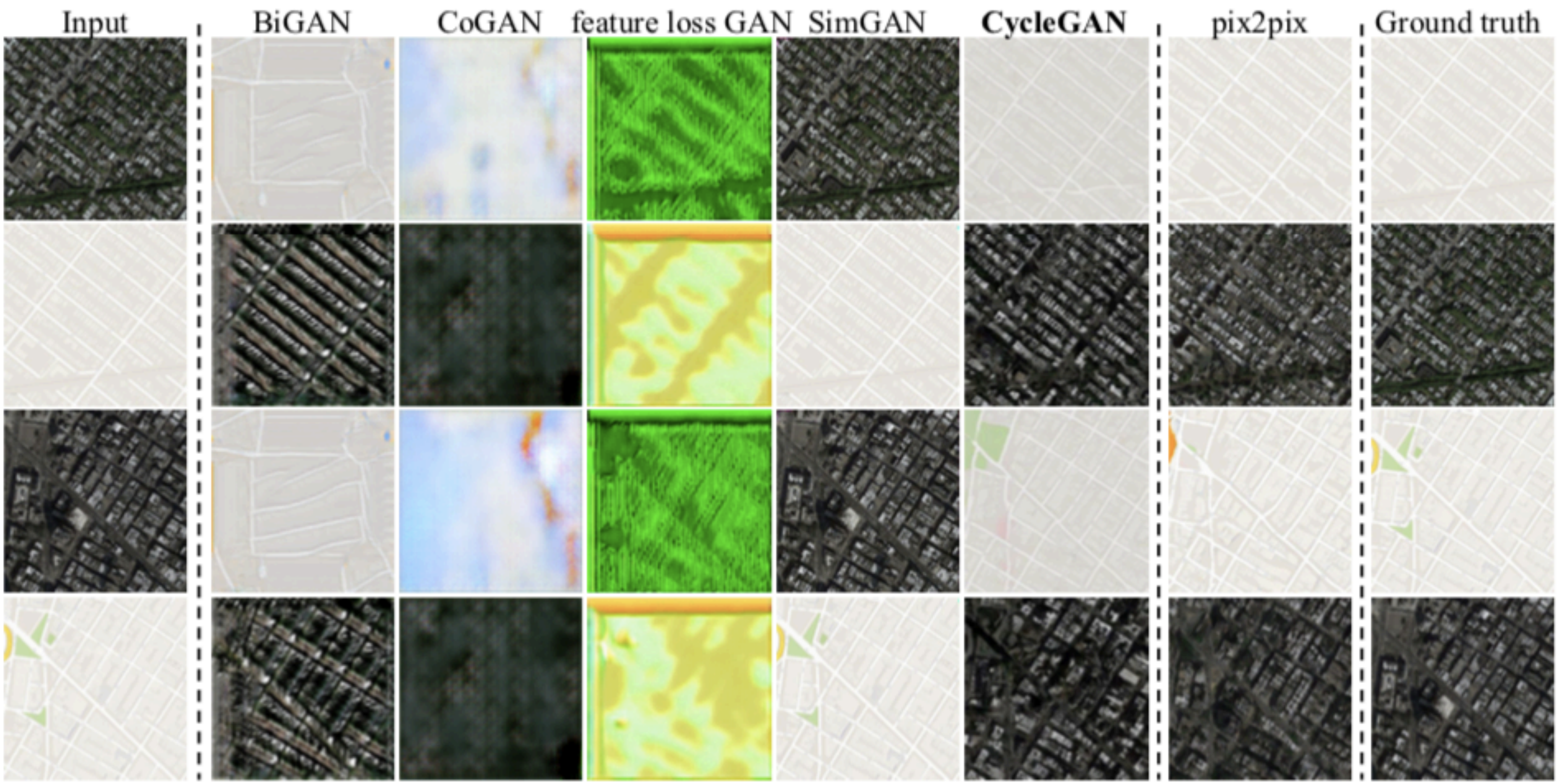


Figure 6:
Different methods for mapping aerial photos <-> maps on Google Maps.

Analysis of the loss function

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Cycle alone	0.22	0.07	0.02
GAN alone	0.51	0.11	0.08
GAN + forward cycle	**0.55**	**0.18**	**0.12**
GAN + backward cycle	0.39	0.14	0.06
CycleGAN (ours)	0.52	0.17	0.11

Table 4: Ablation study: FCN-scores for different variants of our method, evaluated on Cityscapes labels→photo.

Analysis of the loss function

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Cycle alone	0.10	0.05	0.02
GAN alone	0.53	0.11	0.07
GAN + forward cycle	0.49	0.11	0.07
GAN + backward cycle	0.01	0.06	0.01
CycleGAN (ours)	**0.58**	**0.22**	**0.16**

Table 5: Ablation study: classification performance of photo→labels for different losses, evaluated on Cityscapes.

Analysis of the loss function

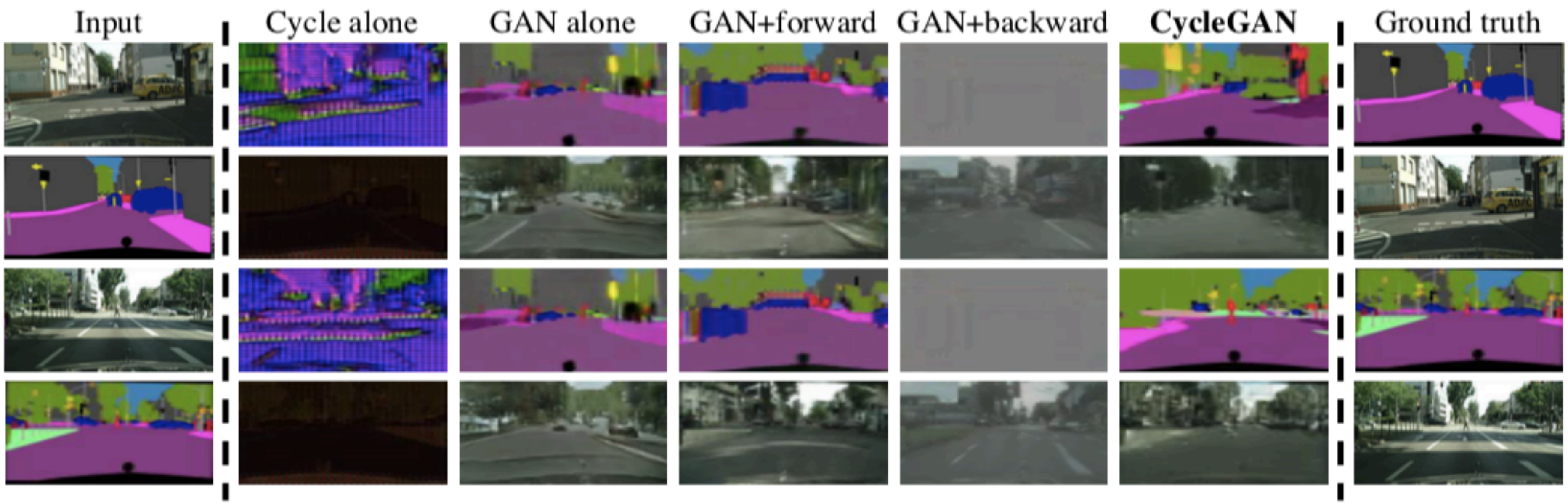
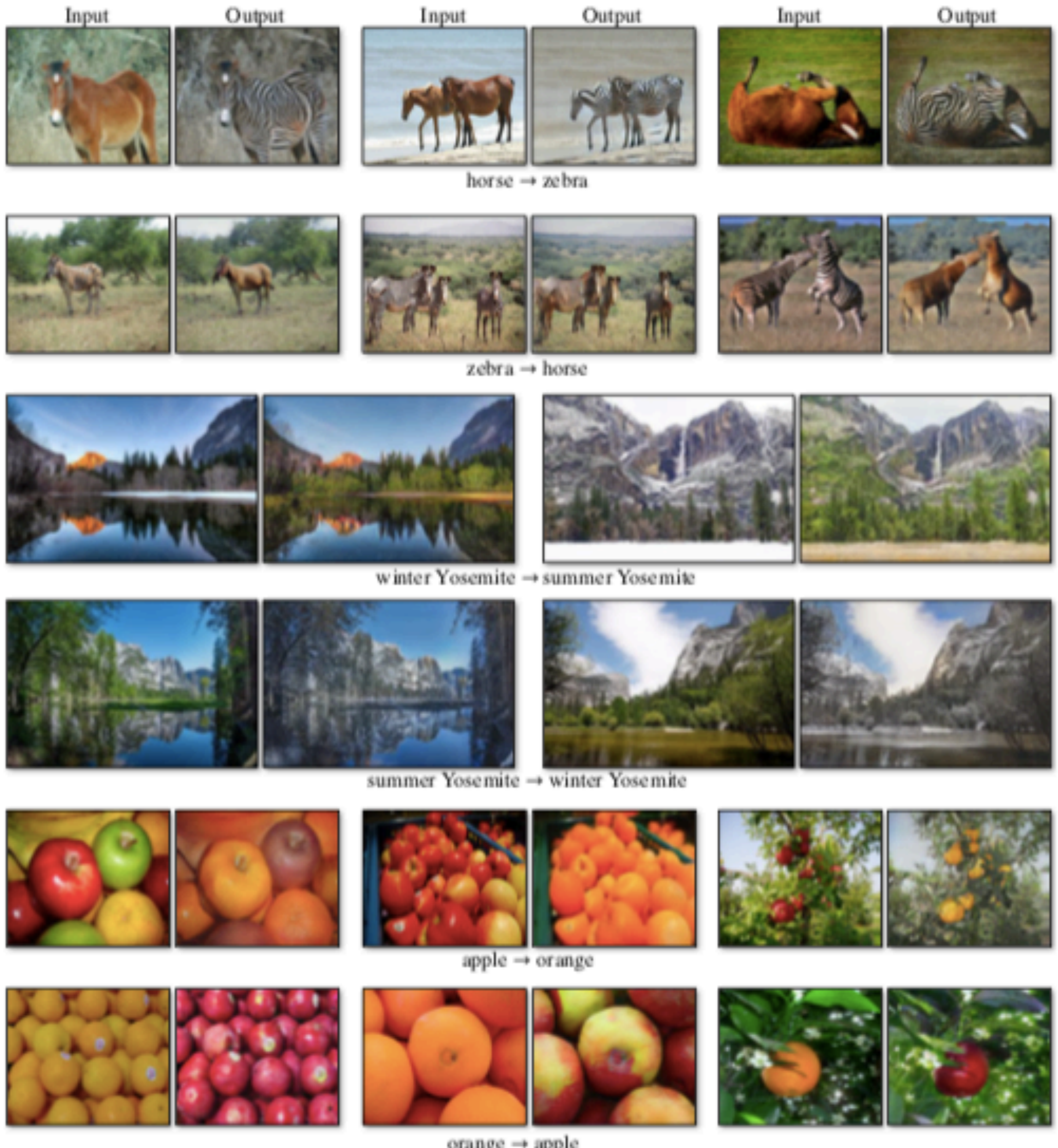


Figure 7:
Different variants of our method for mapping labels↔photos trained on cityscapes.

Results


Input
Output
Input
Output
Input
Output
horse → zebra
zebra → horse
winter Yosemite → summer Yosemite
summer Yosemite → winter Yosemite
apple → orange
orange → apple

Thanks