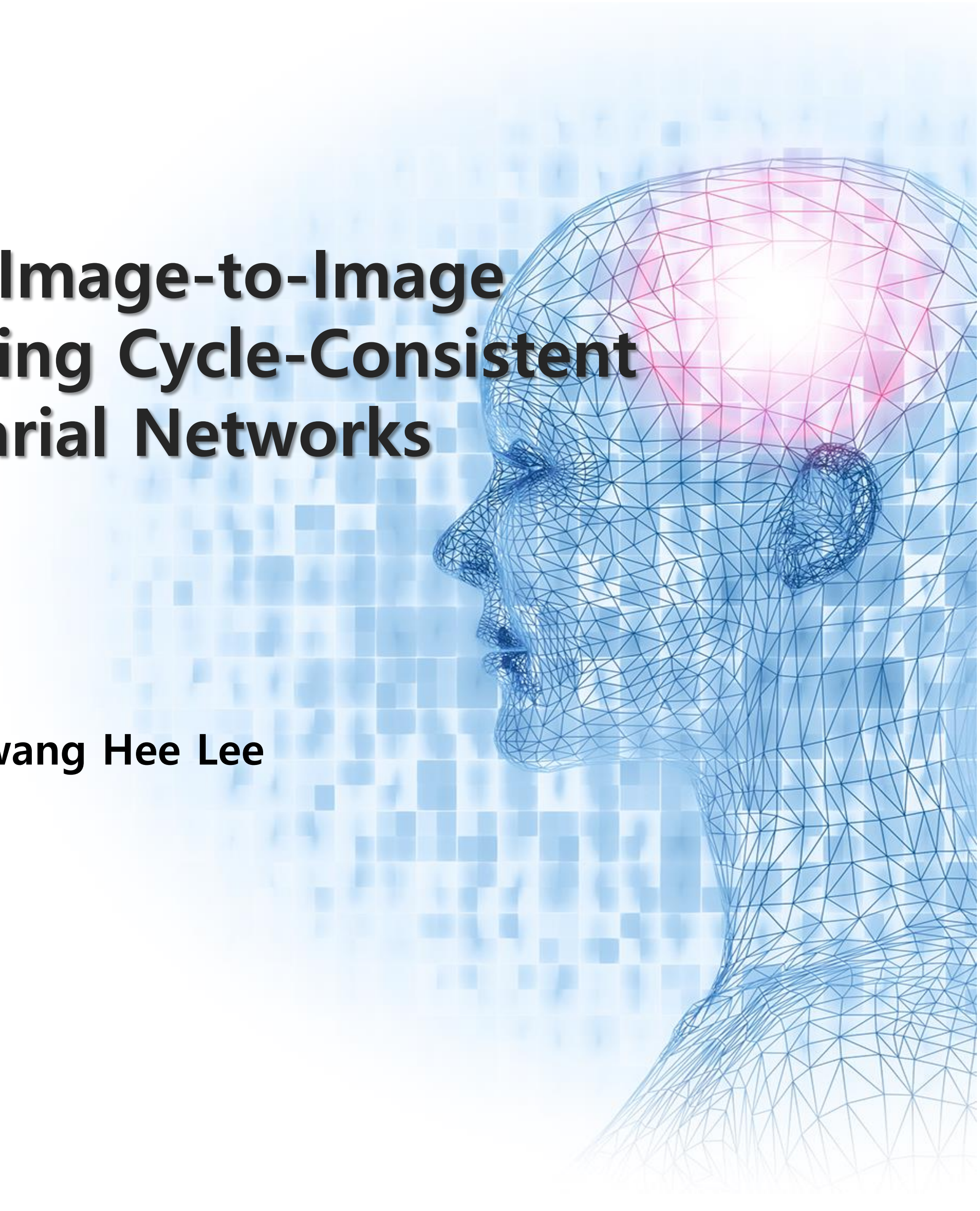
ARTIFICIAL INTELLIGENCE
지능정보기술연구원

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

Kwang Hee Lee



ARTIFICIAL
INTELLIGENCE
RESEARCH
INSTITUTE

Goal

- Given unpaired two image collections, automatically "translate" an image from one into the other and vice versa
- Style transfer, object transfiguration, attribute transfer and photo enhancement

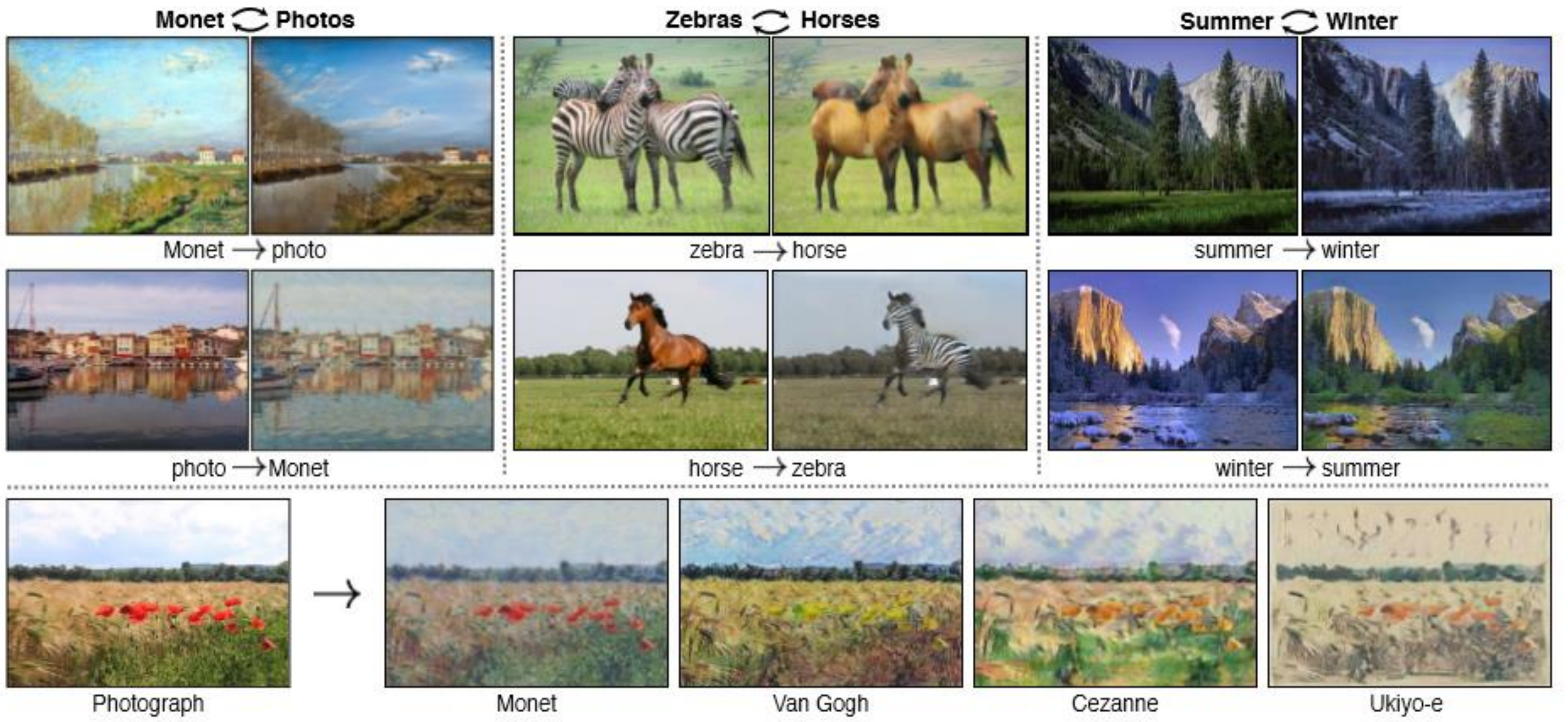


Figure 1: Given any two unordered image collections X and Y, our algorithm learns to automatically "translate" an image from one into the other and vice versa: *(left)* 1074 Monet paintings and 6753 landscape photos from Flickr; *(center)* 1177 zebras and 939 horses from ImageNet; *(right)* 1273 summer and 854 winter Yosemite photos from Flickr. Example application *(bottom)*: using a collection of paintings of a famous artist, learn to render a user's photograph into their style.

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

ARTIFICIAL
INTELLIGENCE
RESEARCH
INSTITUTE

Goal

Obtaining **paired training data** can be **difficult** and **expensive.**

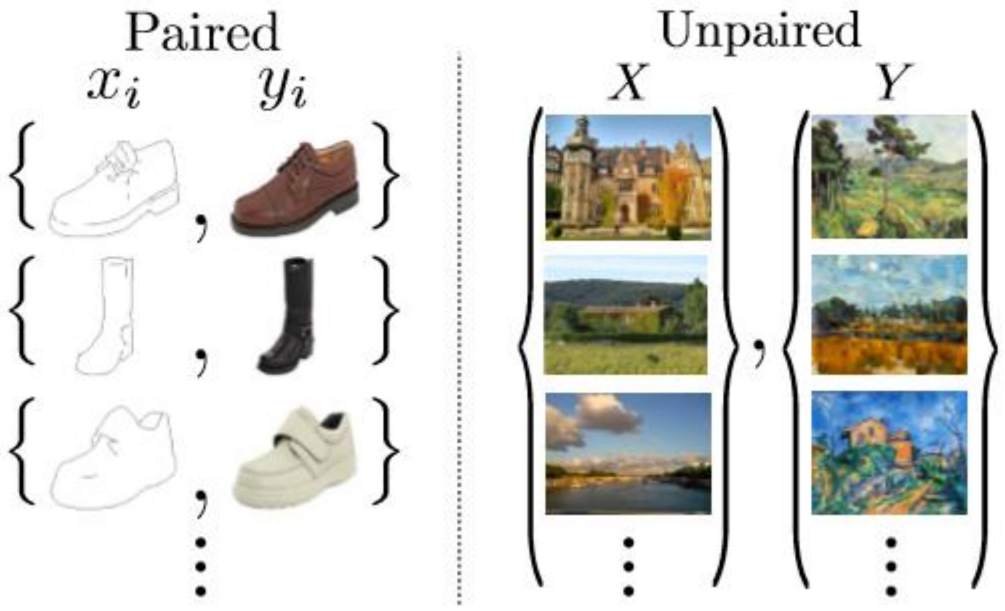
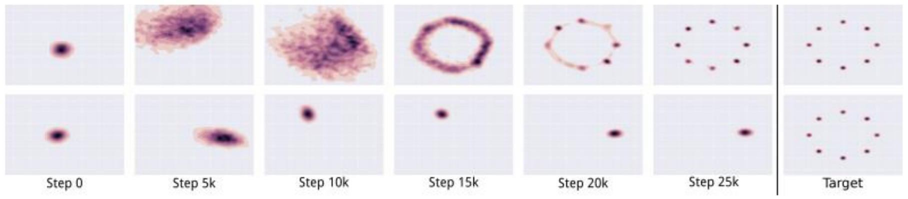


Figure 2: *Paired* training data (left) consists of training examples $\{x_i, y_i\}_{i=1}^{N}$, where the y_i that corresponds to each x_i is given [18]. We instead consider *unpaired* training data (right), consisting of a source set $\{x_i\}_{i=1}^{N} \in X$ and a target set $\{y_j\}_{j=1}^{M} \in Y$, with no information provided as to which x_i matches which y_j.

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Idea

- Assumption : Underlying relationship between the domains
- G : X->Y
- F : Y->X
- **Problem :** infinitely many mappings(under-constrained), mode collapse (need additional condition)



- G and F should be inverses of each other, bijections(one-to-one correspondence).
- Cycle consistency loss : F(G(x)) ≒ x and G(F(y)) ≒ y

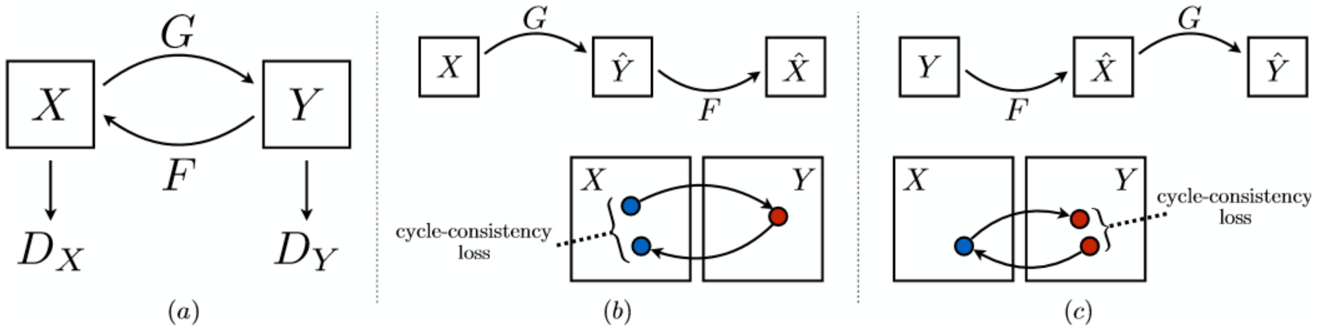


Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X, F, and X. To further regularize the mappings, we introduce two "cycle consistency losses" that capture the intuition that if we translate from one domain to the other and back again we should arrive where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

ARTIFICIAL
INTELLIGENCE
RESEARCH
INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Related Works

- **Unpaired Image-to-Image Translation**

  - not rely on any task-specific, pre-defined similarity function between the input and output

- **Neural Style Transfer**

  - learning the mapping between two domains, rather than between two specific images
  - can be applied to other tasks where single transfer methods don't perform well

- **Cycle Consistency**

  - 3D shape matching, co-segmentation, dense semantic alignment and depth estimation
  - introducing a similar loss

Artificial Intelligence Research Institute

Proposed Method

- **Formulation**
  - Domain : X and Y
  - Training samples : $\{x_i\}_{i=1}^{N} \in X \quad \{y_j\}_{j=1}^{M} \in Y$
  - Mappings : $G : X \rightarrow Y \quad F : Y \rightarrow X$
  - Adversarial discriminator : D_X and D_Y

$$\begin{tikzcd}
\boxed{X} \arrow[r, bend left, "G"] \arrow[d] & \boxed{Y} \arrow[l, bend left, "F"] \arrow[d] \\
D_X & D_Y
\end{tikzcd}$$

aims to distinguish $\boldsymbol{x}$ and $\{F(y)\}$

aims to distinguish $\{y\}$ and $\{G(x)\}$

Proposed Method

- **Loss = Adversarial Loss + Cycle Consistency Loss**

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

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

- **Adversarial Loss**

$$G : X \rightarrow Y$$

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

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

Proposed Method

- Loss = Adversarial Loss + Cycle Consistency Loss

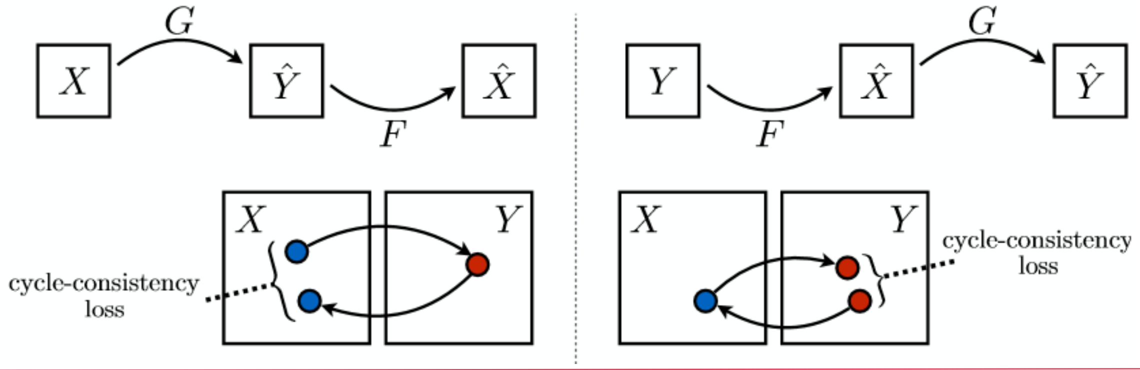
$$\mathcal{L}(G, F, D_X, D_Y) = \boxed{\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X)} + \boxed{\lambda \mathcal{L}_{\text{cyc}}(G, F)}$$

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

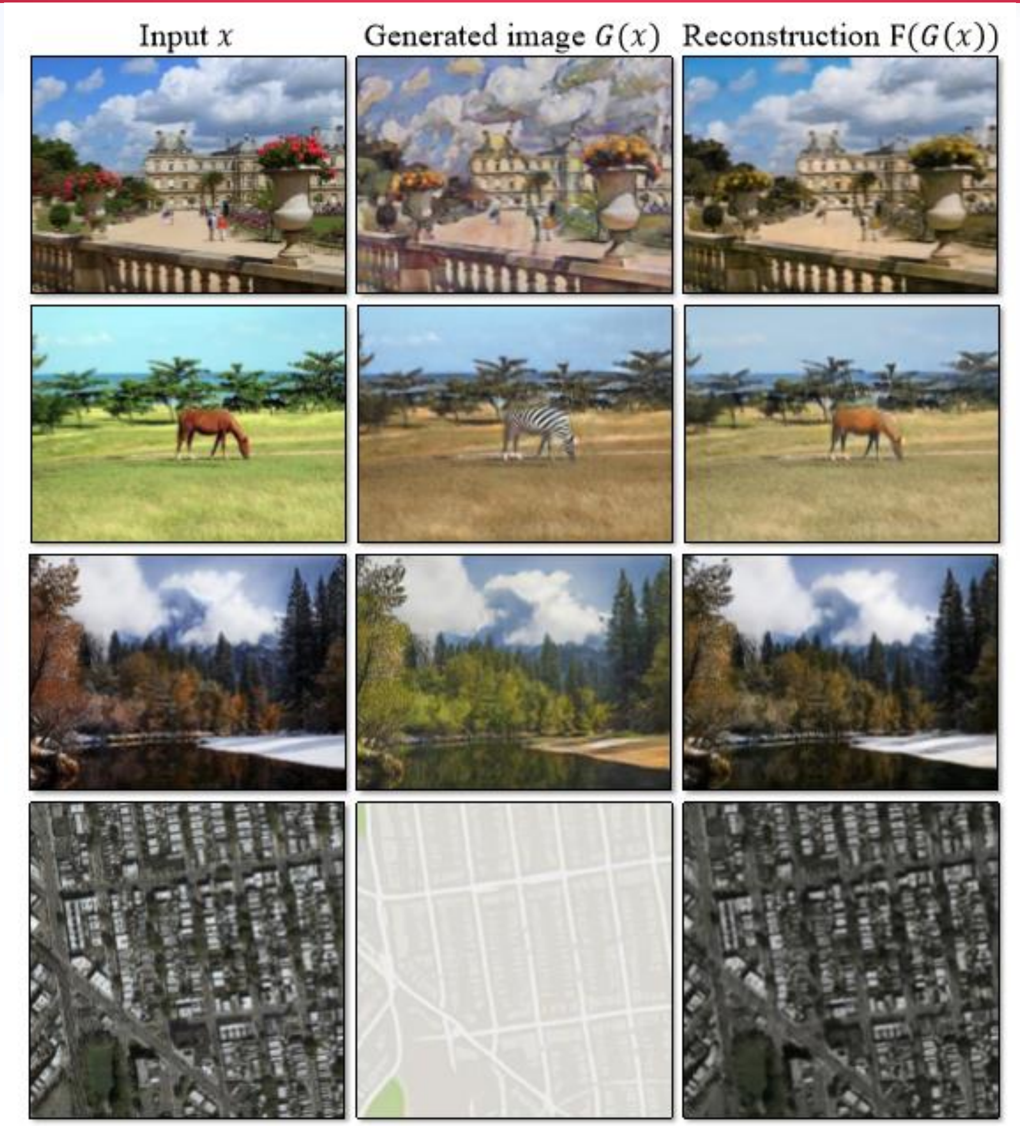
To further reduce the space of possible mapping functions,

- Cycle Consistency Loss

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



Proposed Method

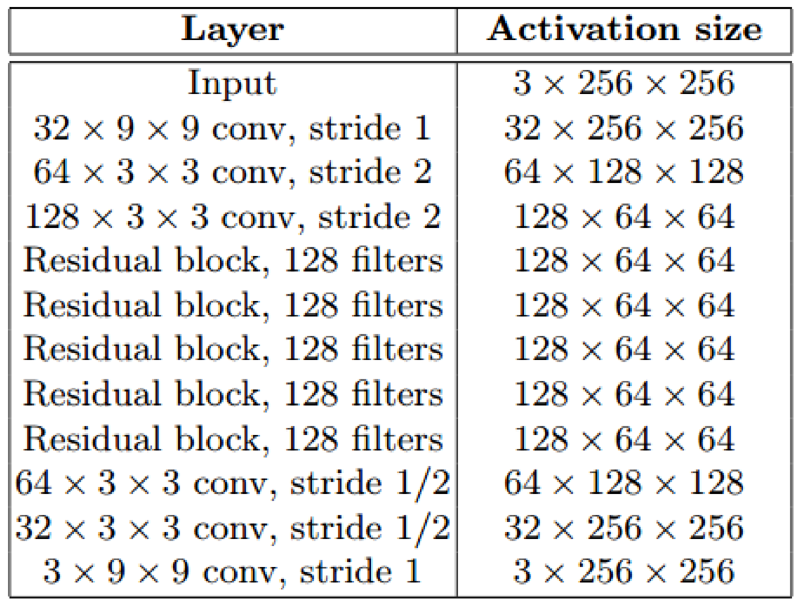


ARTIFICIAL INTELLIGENCE
지능정보기술연구원

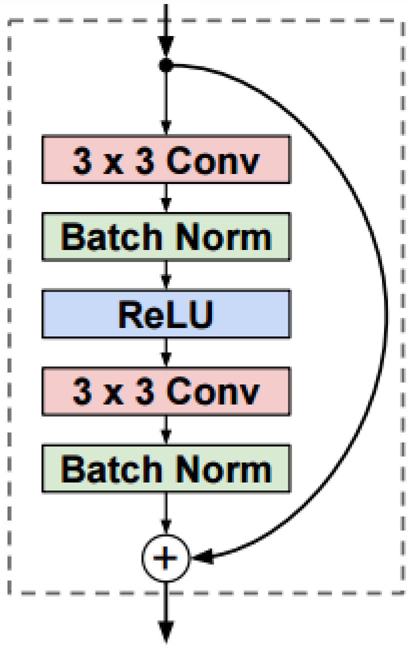
Implemention

- **Network Architecture**

-Generator :



Layer	Activation size
Input	$3 \times 256 \times 256$
$32 \times 9 \times 9$ conv, stride 1	$32 \times 256 \times 256$
$64 \times 3 \times 3$ conv, stride 2	$64 \times 128 \times 128$
$128 \times 3 \times 3$ conv, stride 2	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
$64 \times 3 \times 3$ conv, stride 1/2	$64 \times 128 \times 128$
$32 \times 3 \times 3$ conv, stride 1/2	$32 \times 256 \times 256$
$3 \times 9 \times 9$ conv, stride 1	$3 \times 256 \times 256$



- Discriminator : PatchGAN(70x70)

Implemention

- **Training details**
  - **Stable training : LSGAN**

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

$$\Downarrow$$

$$\mathcal{L}_{\mathrm{LSGAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\mathrm{data}}(y)}[(D_Y(y) - 1)^2] + \mathbb{E}_{x \sim p_{\mathrm{data}}(x)}[D_Y(G(x))^2]$$

Experiments

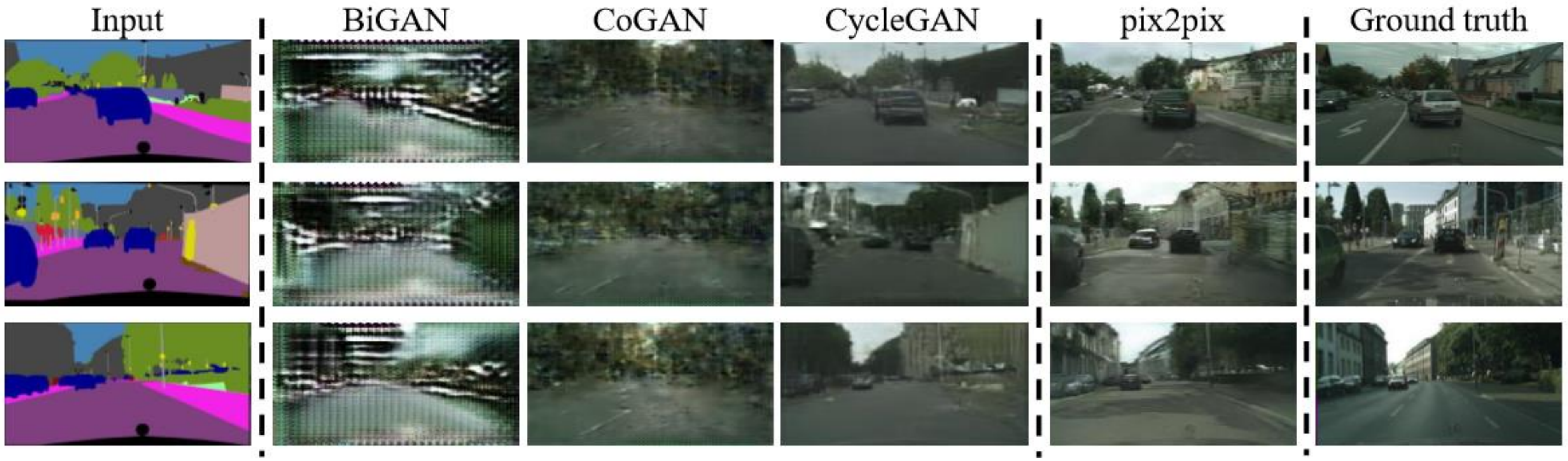
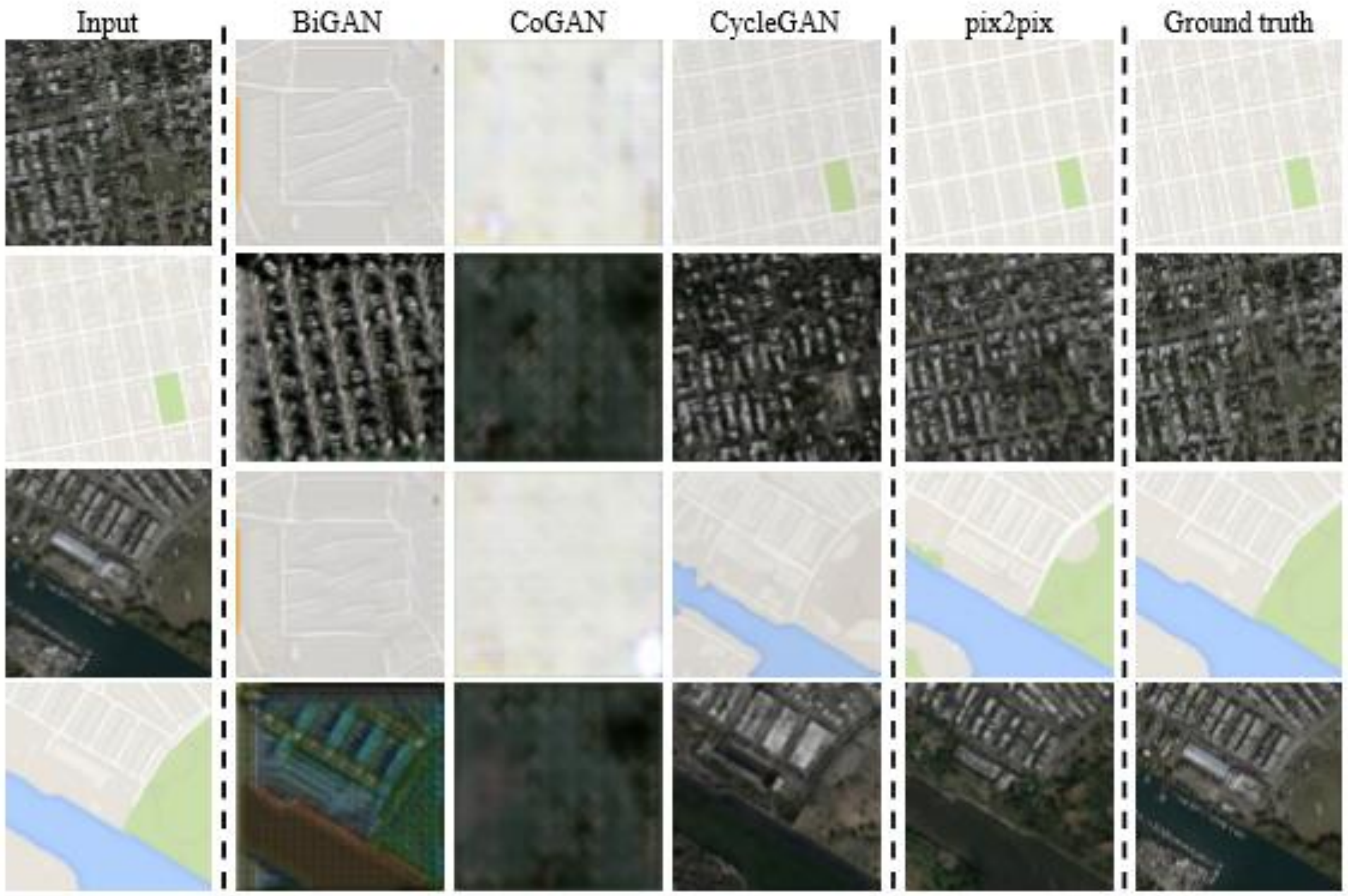


Figure 5: Different methods for mapping labels↔photos trained on cityscapes. From left to right: input, BiGAN [5, 6], CoupledGAN [27], CycleGAN (ours), pix2pix [18] trained on paired data, and ground truth.

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Experiments



Artificial Intelligence Research Institute

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Experiments

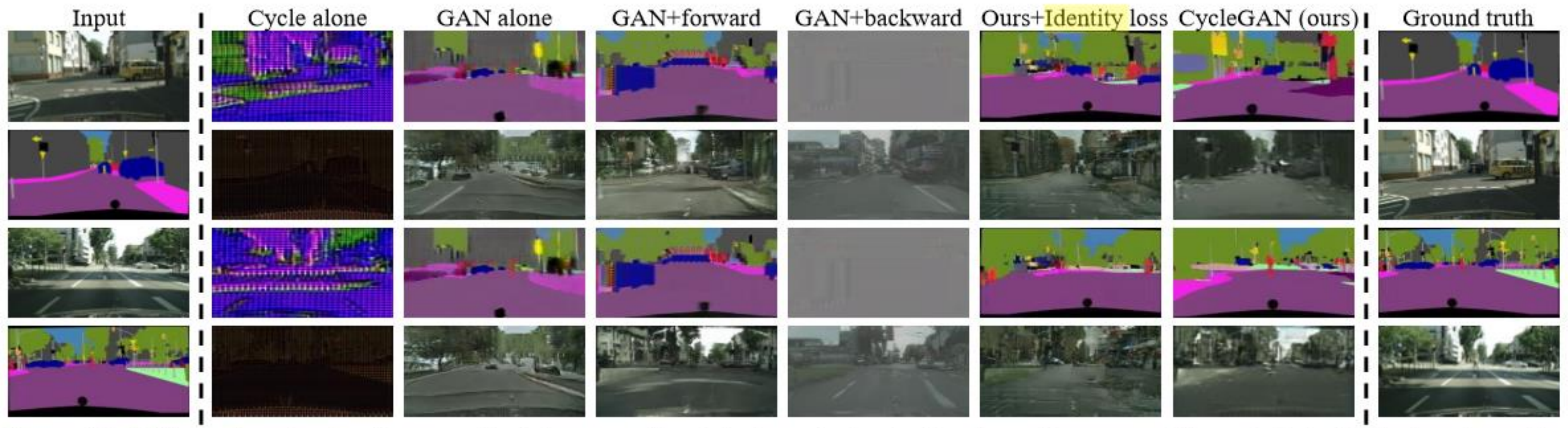


Figure 7: Different variants of our method for mapping labels↔photos trained on cityscapes. From left to right: input, cycle-consistency loss alone, adversarial loss alone, GAN + forward cycle-consistency loss ($F(G(x)) \approx x$), GAN + backward cycle-consistency loss ($G(F(y)) \approx y$), CycleGAN (our full method), and ground truth. Both *Cycle alone* and *GAN + backward* fail to produce images similar to the target domain. *GAN alone* and *GAN + forward* suffer from mode collapse, producing identical label maps regardless of the input photo.

ARTIFICIAL
INTELLIGENCE
RESEARCH
INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Experiments

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

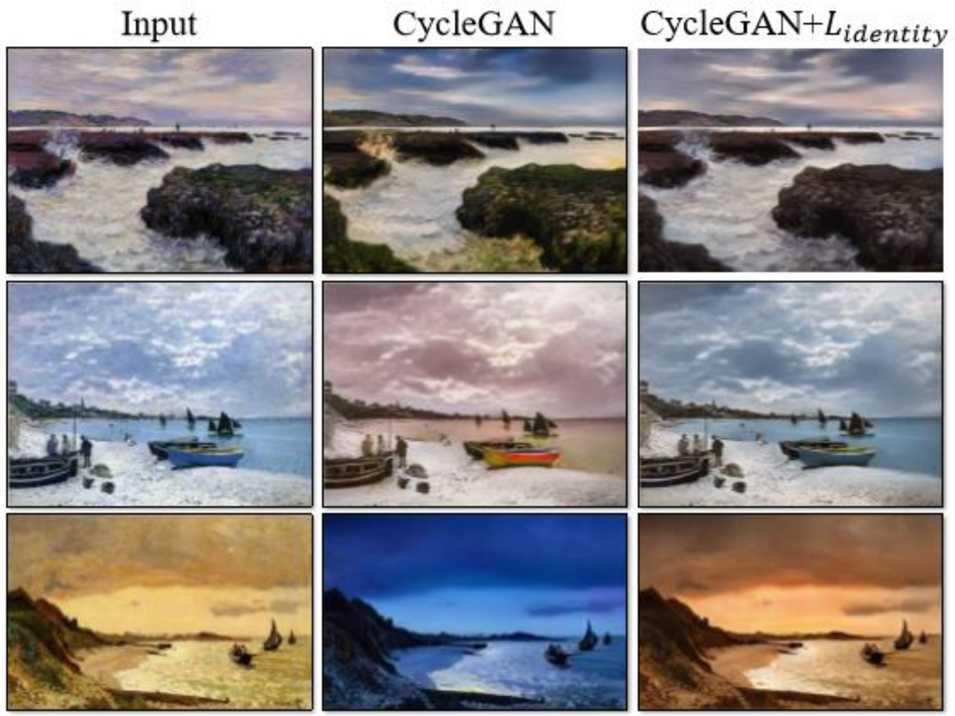


Figure 9: The effect of the *identity mapping loss* on Monet→ Photo. From left to right: input paintings, CycleGAN without identity mapping loss, CycleGAN with identity mapping loss. The identity mapping loss helps preserve the color of the input paintings.

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

AIRI
ARTIFICIAL INTELLIGENCE
지능정보기술연구원

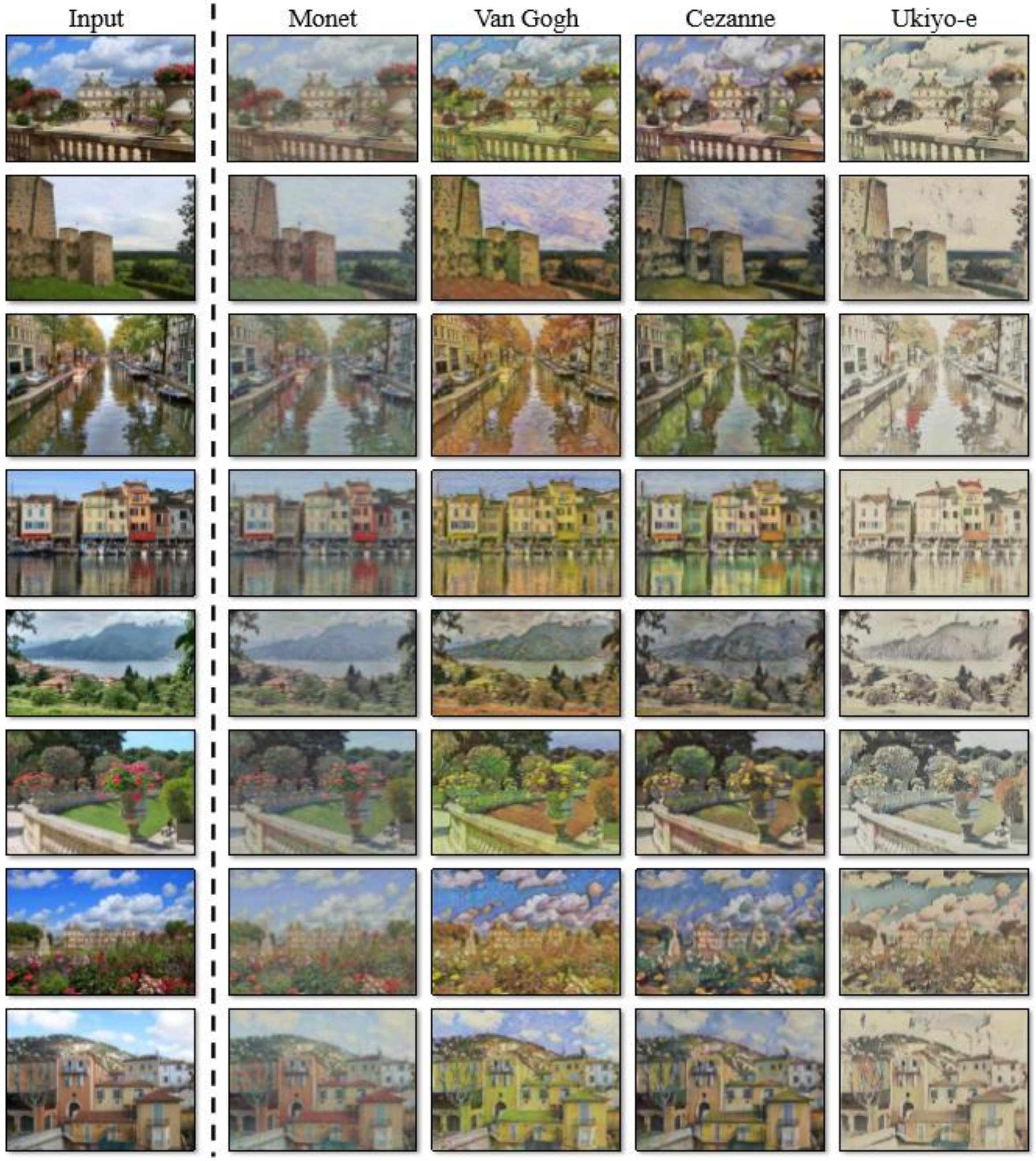


Figure 10: Collection style transfer: we transfer input images into the artistic styles of Monet, Van Gogh, Cezanne, and Ukiyo-e. Please see our website for additional examples.

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

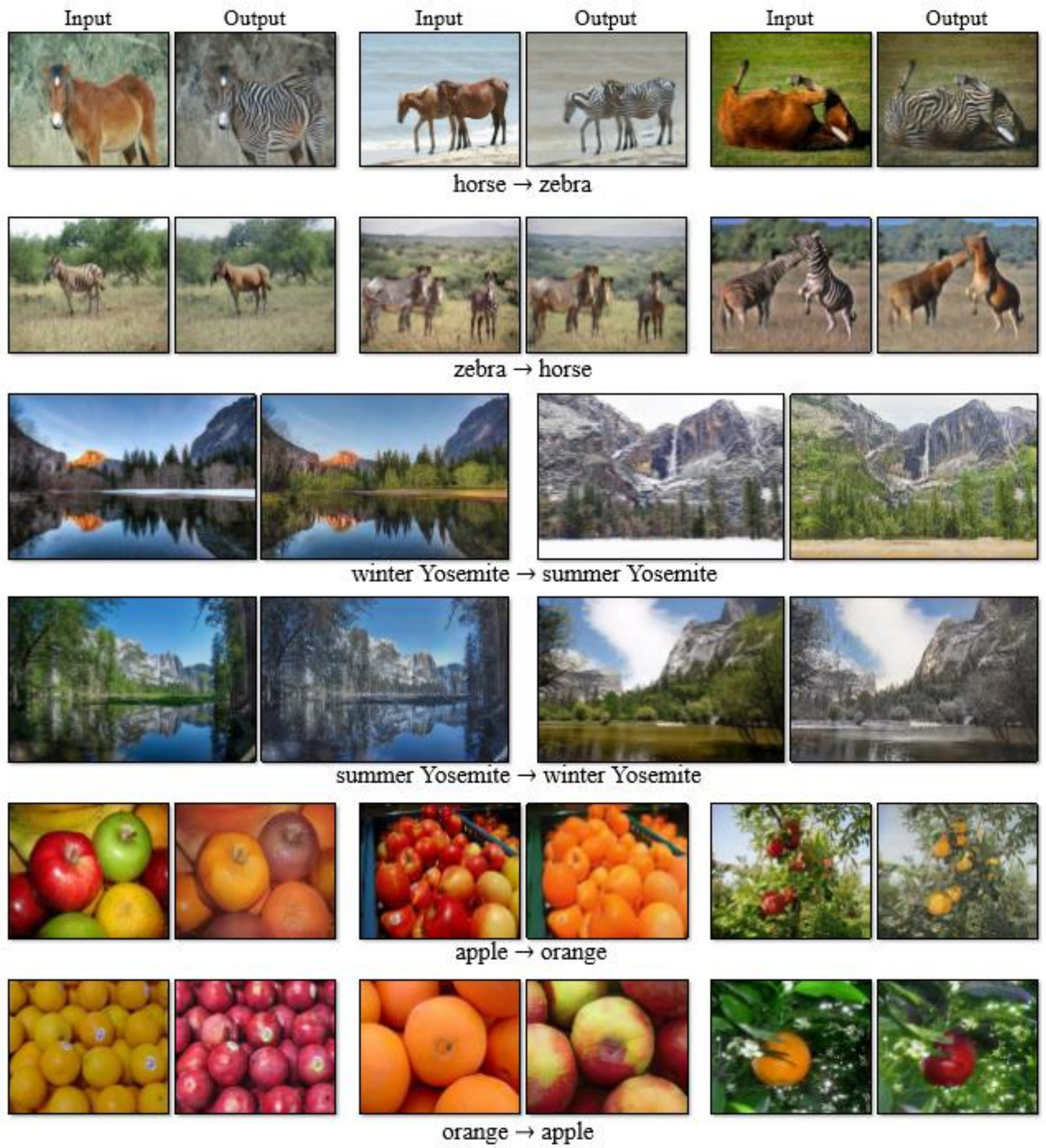


Figure 13: Our method applied to several translation problems. These images are selected as relatively successful results – please see our website for more comprehensive and random results. In the top two rows, we show results on object transfiguration between horses and zebras, trained on 939 images from the *wild horse* class and 1177 images from the *zebra* class in Imagenet [37]. The middle two rows show results on season transfer, trained on winter and summer photos of Yosemite from Flickr. In the bottom two rows, we train our method on 996 *apple* images and 1020 *navel orange* images from ImageNet.

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

CycleGAN vs Style Transfer

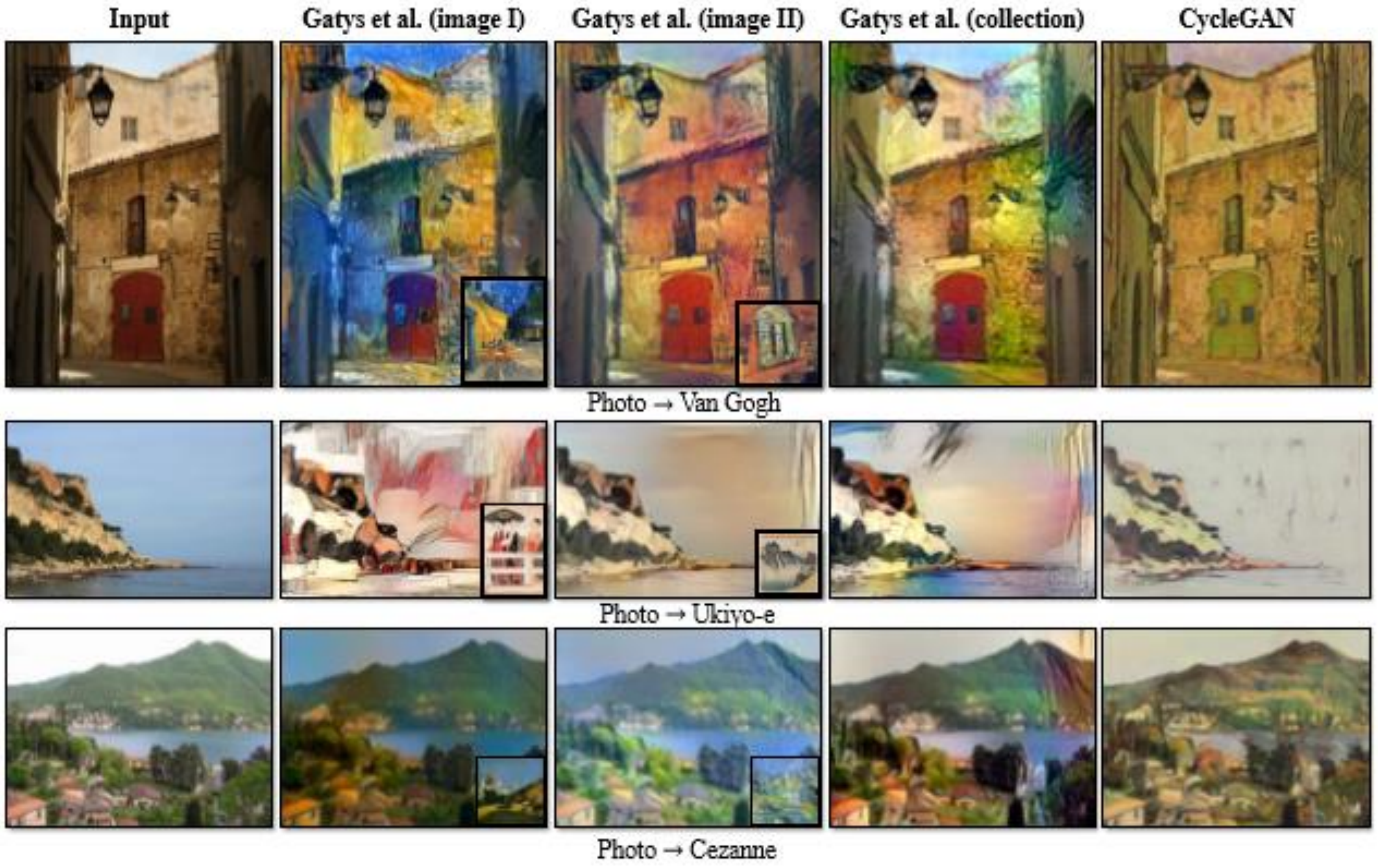


Figure 15: We compare our method with neural style transfer [10] on photo stylization. Left to right: input image, results from [10] using two different representative artworks as style images, results from [10] using the entire collection of the artist, and CycleGAN (ours).

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

CycleGAN vs Style Transfer

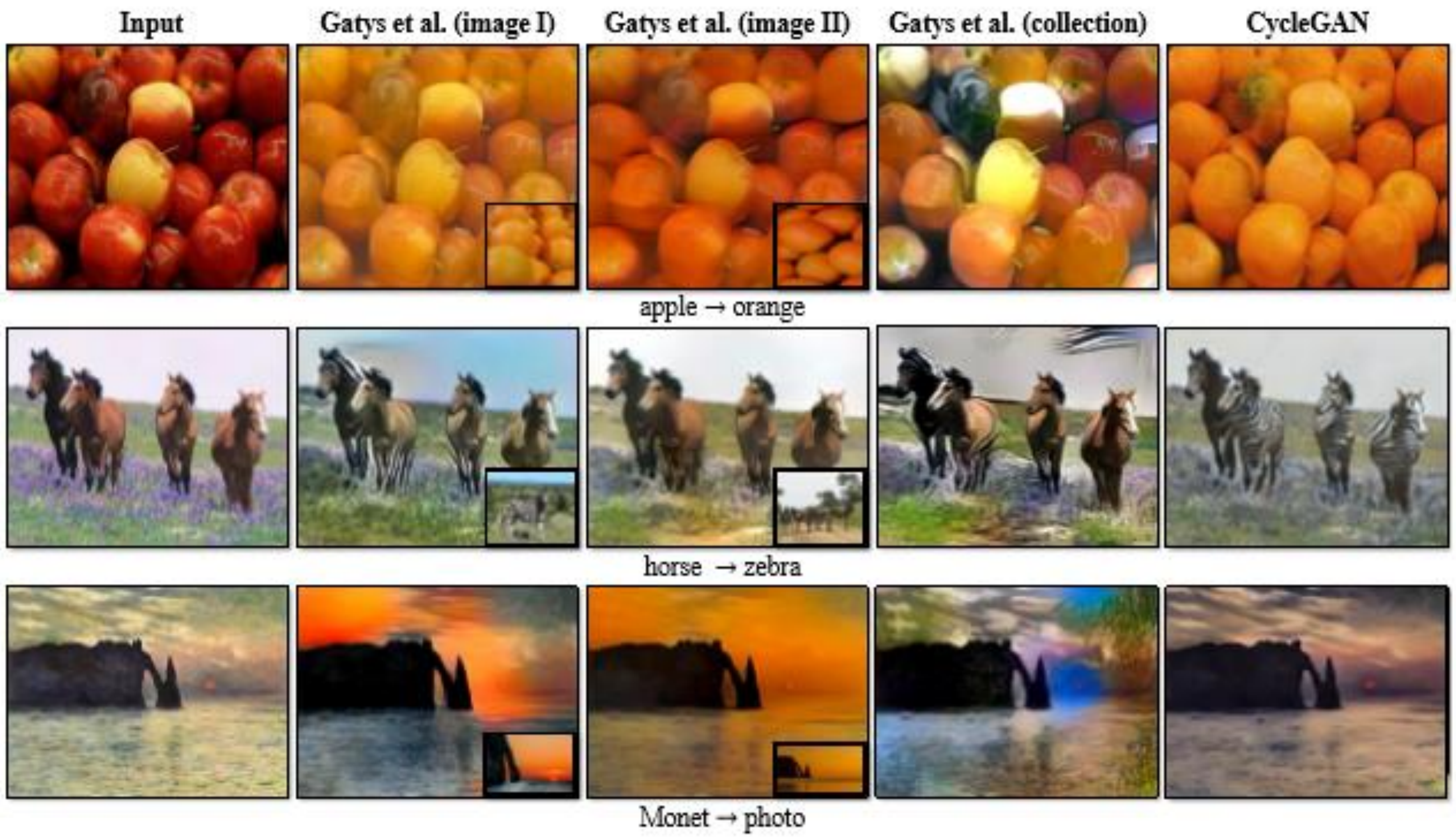


Figure 16: We compare our method with neural style transfer [10] on various applications. From top to bottom: apple→orange, horse→zebra, and Monet→photo. Left to right: input image, results from [10] using two different images as style images, results from [10] using all the images from the target domain, and CycleGAN (ours).

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Failure Cases

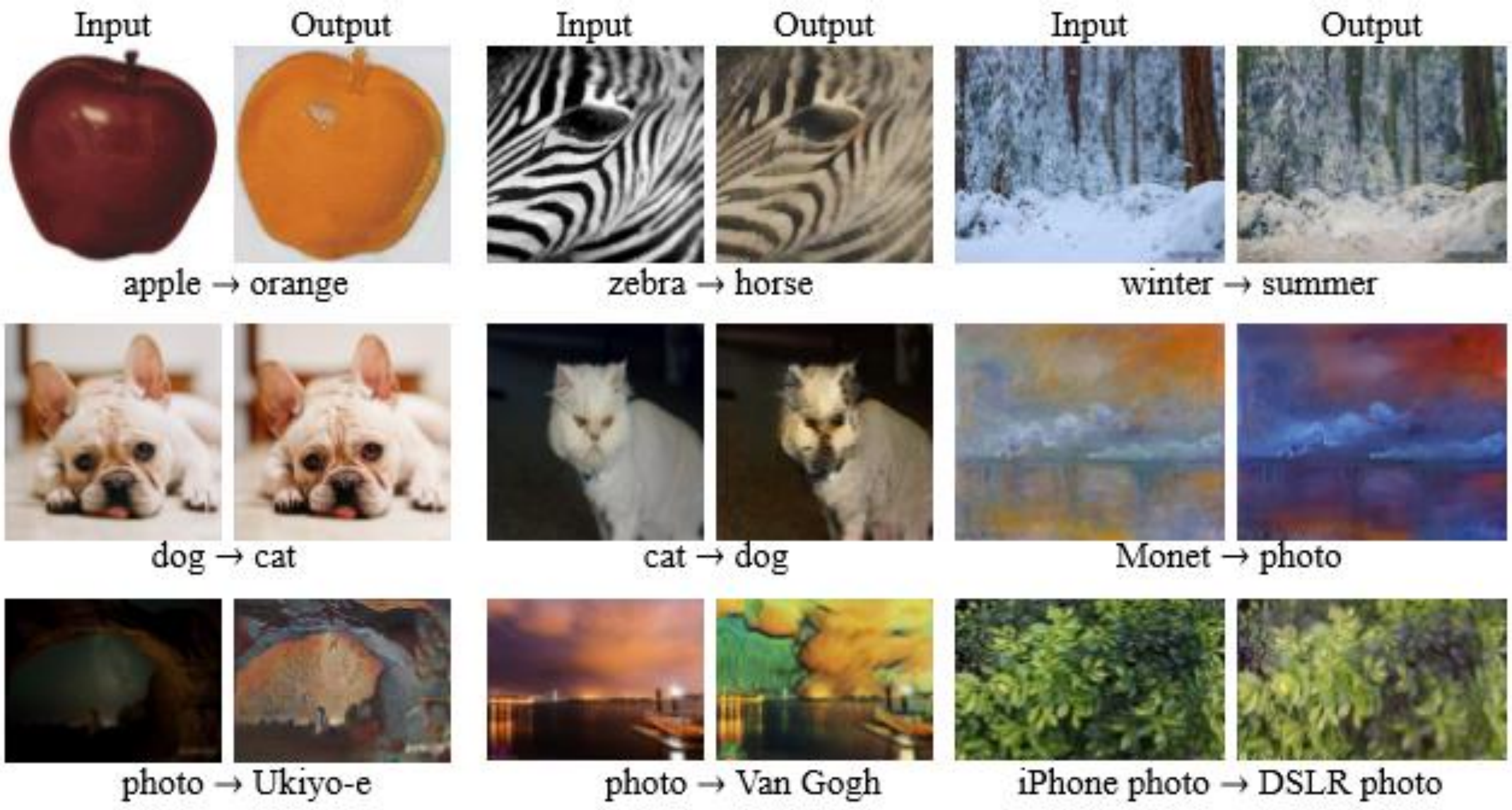


Figure 17: Typical failure cases of our method. Please see our website for more comprehensive results.

ARTIFICIAL INTELLIGENCE RESEARCH INSTITUTE

AIRI
ARTIFICIAL INTELLIGENCE
지능정보기술연구원

Limitation and Discussion

- **Future work : CycleGAN** fails on tasks that require geometric changes.
- Integrating weak or semi-supervised data may lead to substantially more powerful translators

ARTIFICIAL
INTELLIGENCE
RESEARCH
INSTITUTE