GAN 補充說明

CycleGAN
ReCycleGAN

TA 鍾嘉峻

Note

- 11/11 : Lab6 release
  - Deadline : 12/2
- 11/18: you can ask any questions
  - Afternoon
- 11/25: Final exam
  - Open book
  - You can download the slides or bring your book
  - No internet!
- Final project assistence start from 12/9
- 12/31 : The last day you can submit your homework and demo

Some Useful Link

- Connected Paper
  - https://www.connectedpapers.com/
- Paper with Codes
  - https://paperswithcode.com/
- Arxiv Sanity
  - http://www.arxiv-sanity.com/

Outline

- Goal
- Related Work
- System / Method
- Result

If you want to access this slide , you can use the QR code above

Outline

- Goal
- Related Work
- System / Method
- Result

If you want to access this slide , you can use the QR code above

Goal

- Motivation
  - Here’s a youtube channel called “阿滴英文”
  - Two members: one is “阿滴”, and other is “滴妹”

If you want to watch their 2018 April’s Fool video

Goal

- Motivation
    - 阿滴英文 has a video for 2018 April’s Fool...

Goal

- 如果我可以有一個妹妹，跟我搭配一起教英文，應該會蠻有趣的....
- 滴妹..其實並不存在
- 她其實是用我的臉為模型，所做出來的一個電腦合成影像

醒醒吧!你沒有妹妹!!!!!QQQ

Goal

Don't Worry!! Let's create "sister" by ourselves

- 如果我可以有一個妹妹，跟我搭配一起教英文，應該會蠻有趣的....
- 滴妹..其實並不存在
- 她其實是用我的臉為模型，所做出來的一個電腦合成影像

~~阿不對!!把滴妹給我還來呀!!!~~

Goal

- Traditional way to generate the fake image

Green Screen

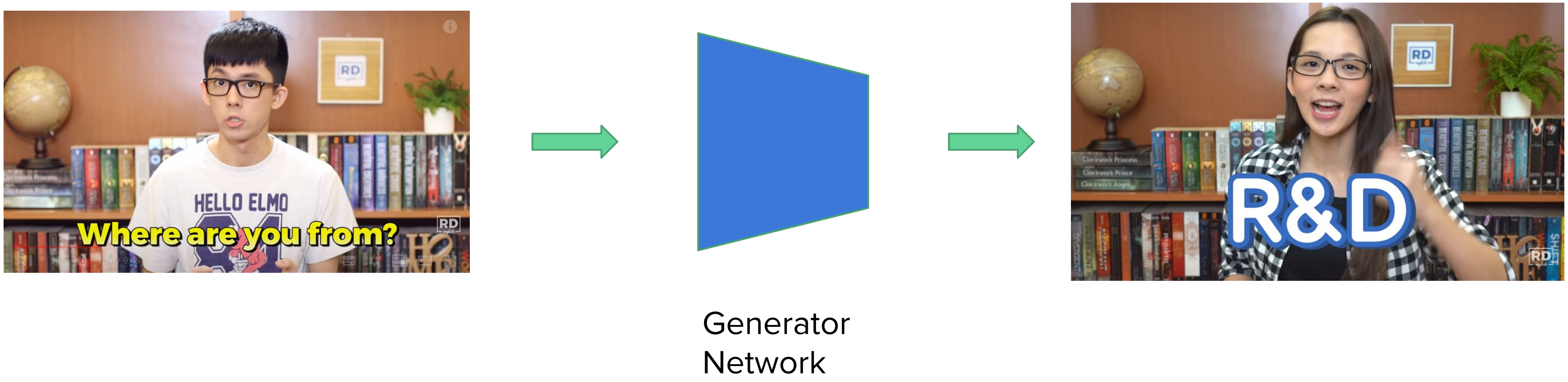
3D Model

Goal

- FaceAPP

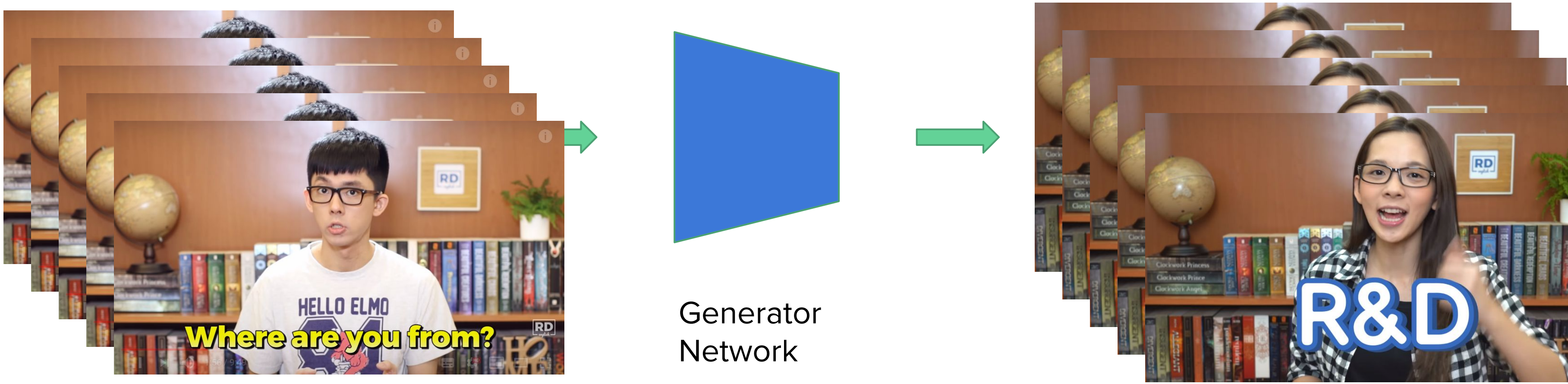
Goal

- Proposed goal
  - Transfer one image via neural network



Goal

- Proposed goal
  - Transfer lots of images via neural network, and merge them together



If you want to watch their 2018 April's Fool video

Goal

- Expected Result

Outline

- Goal
- Related Work
- Difficulties / Uniqueness
- Plan
- System / Method
- Result

If you want to access this slide , you can use the QR code above

Related Work

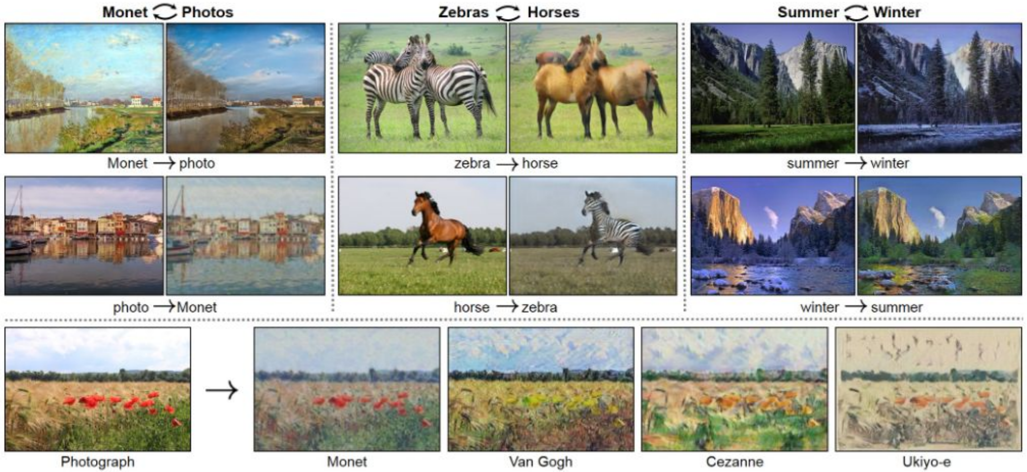
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

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

UC Berkeley

In ICCV 2017

Paper | PyTorch code | Torch code

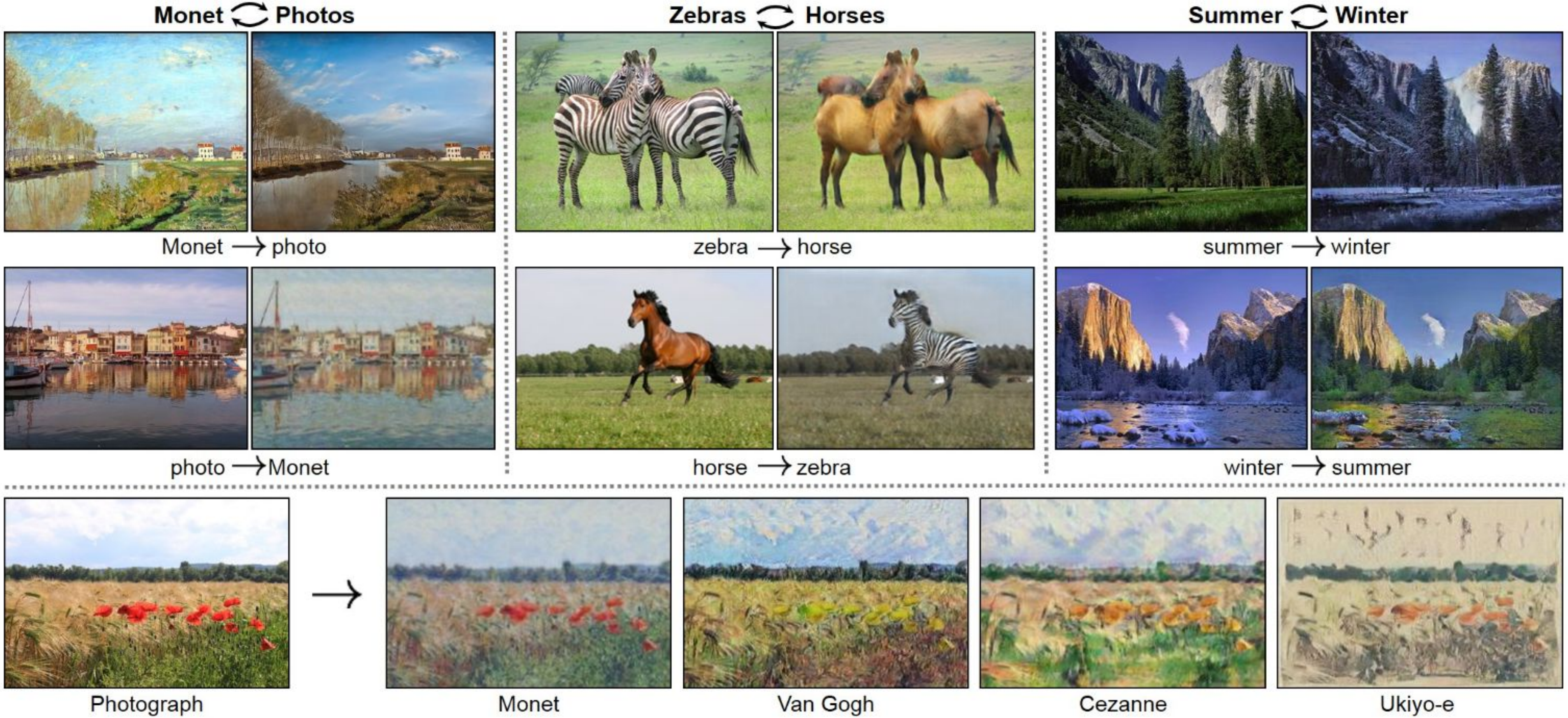


Abstract

Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available. We present an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping G: X → Y such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss. Because this mapping is highly under-constrained, we couple it with an inverse mapping F: Y → X and introduce a cycle consistency loss to push F(G(X)) ≈ X (and vice versa).

Source : https://junyanz.github.io/CycleGAN/

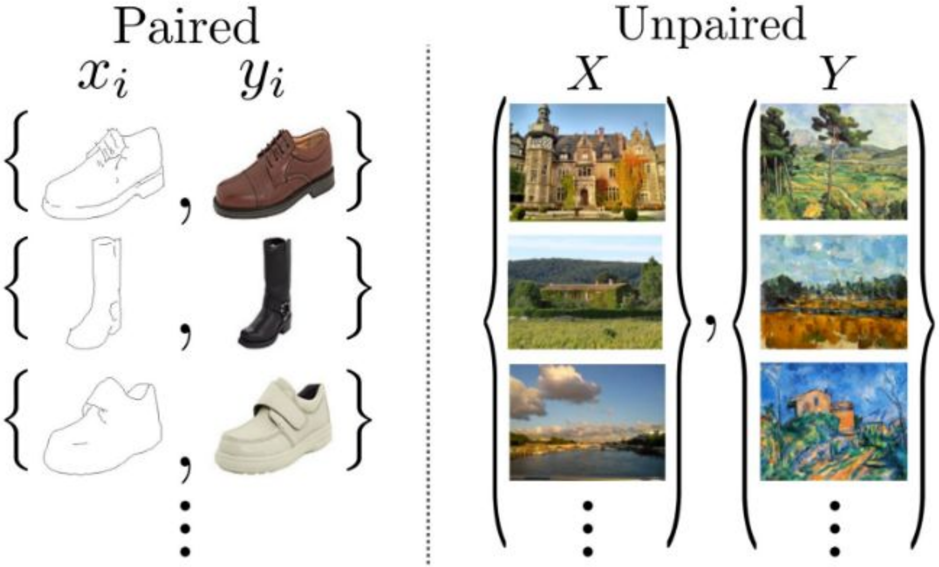
Cycle-GAN



Monet Photos
Zebras Horses
Summer Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer
Photograph
Monet
Van Gogh
Cezanne
Ukiyo-e

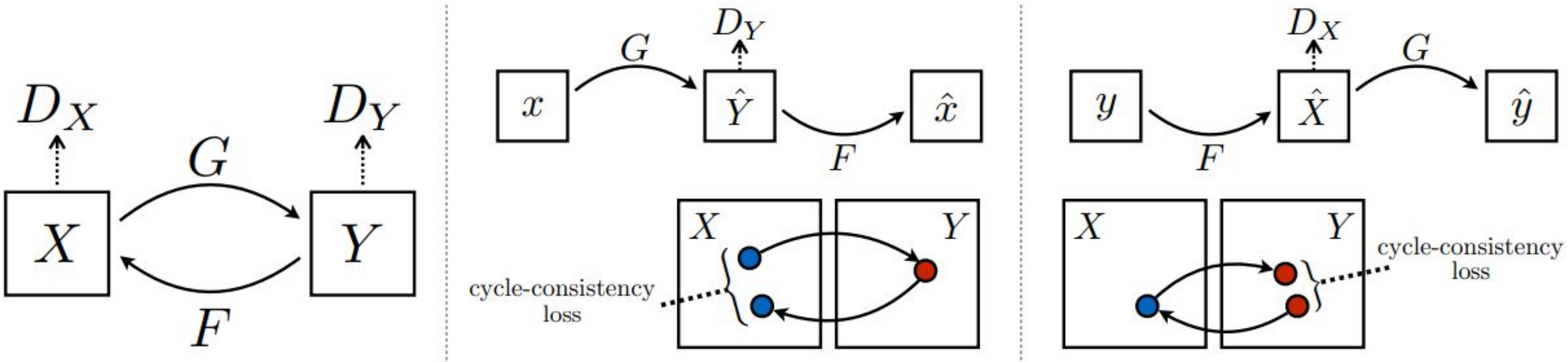
Cycle-GAN Dataset

- Can use unpair dataset
  - Pix2Pix depend on the availaibilty of training examples where the same data is available in both domains



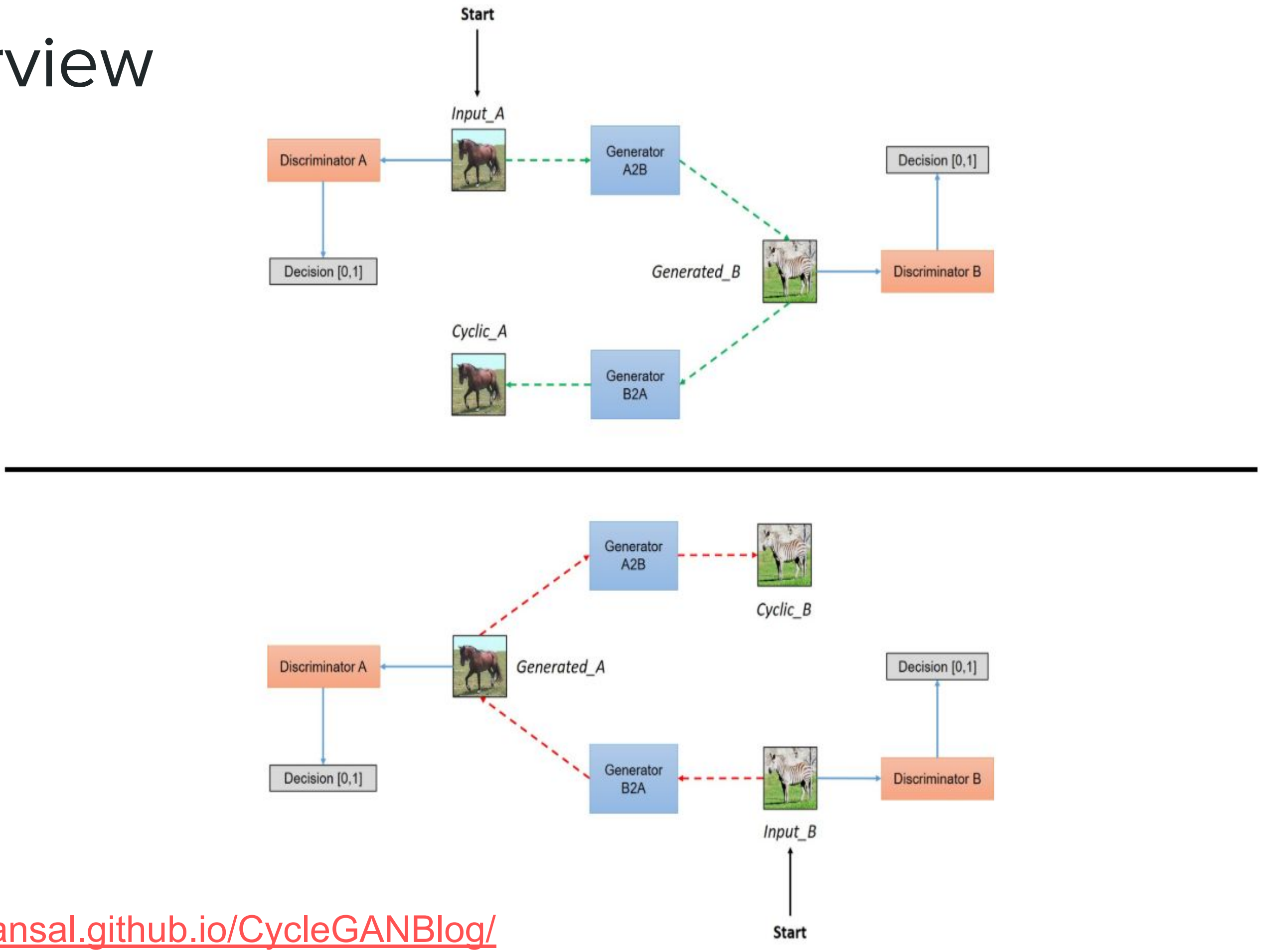
Cycle-GAN Architecture

- Two GAN Network
  - Two generators
  - Two discriminators



Cycle-GAN Overview

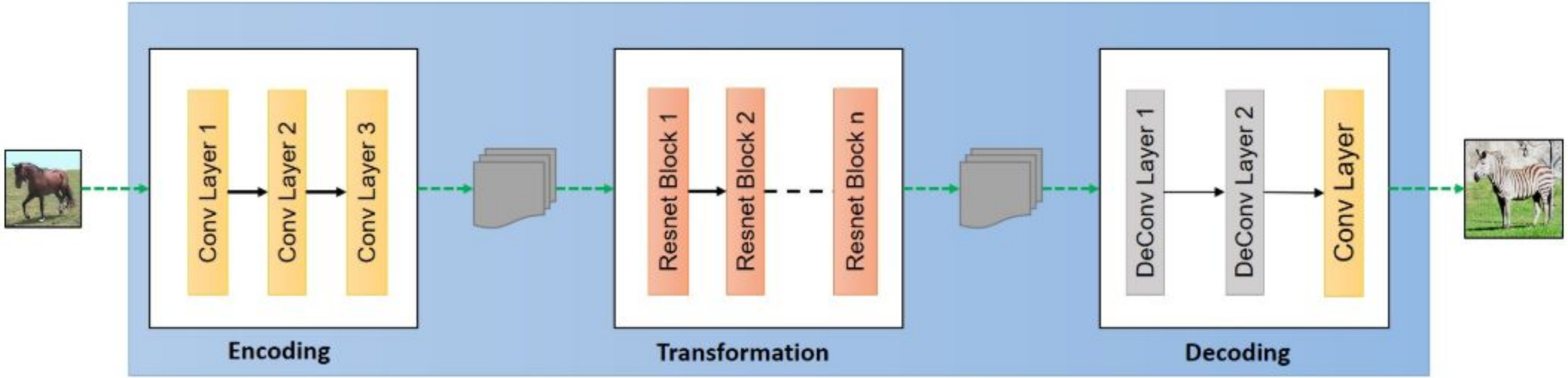
- Idea Overview



Source: https://hardikbansal.github.io/CycleGANBlog/

Cycle-GAN Architecture

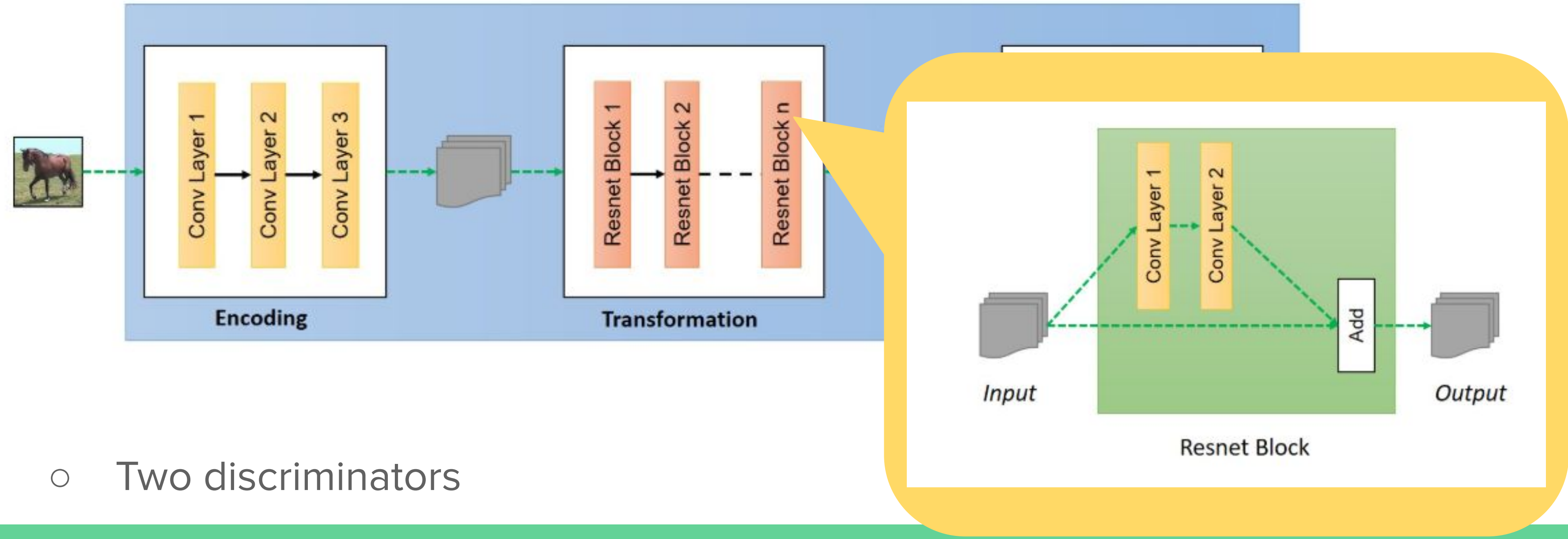
- Two GAN Network
  - **Two generators**



  - Two discriminators

Cycle-GAN Architecture

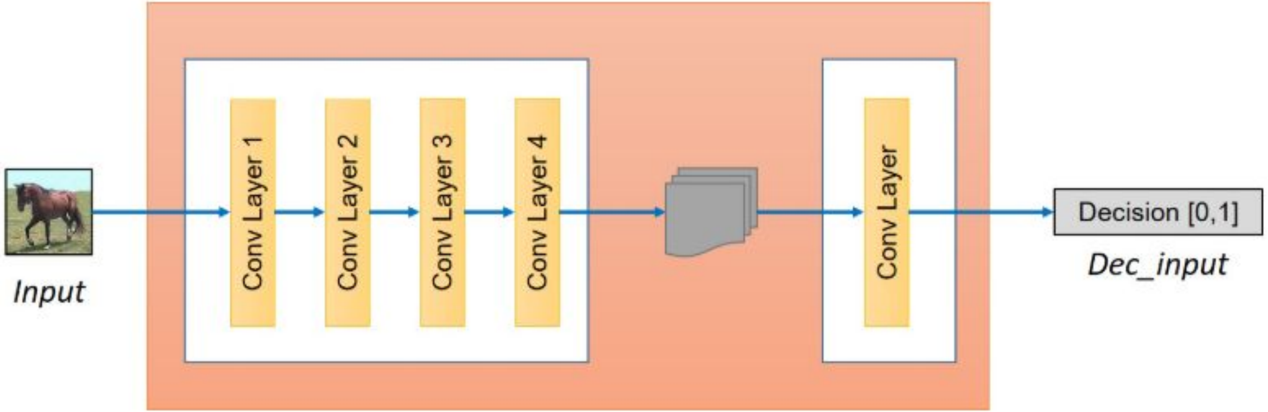
- Two GAN Network
    - **Two generators**



    - Two discriminators

Cycle-GAN Architecture

- Two GAN network
  - Two generators
  - **Two discriminators**



Cycle-GAN Loss function

- Two kinds of loss function
  - Adversarial Loss
  - Cycle-consistence loss

Cycle-GAN Loss function

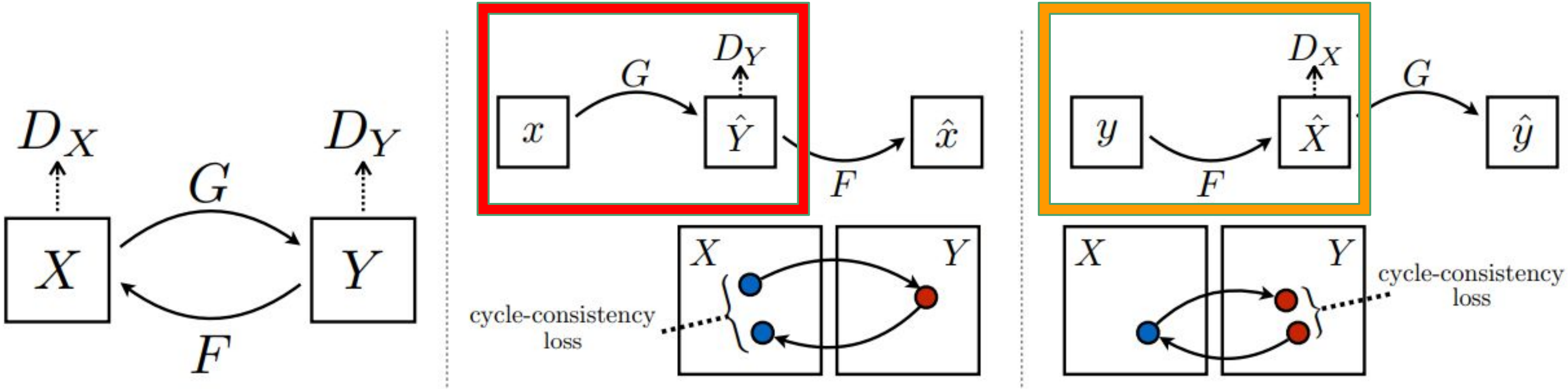
- Two kinds of loss function
  - Adversarial Loss
  - Cycle-consistence loss
- Full objective

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

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

Cycle-GAN Loss function

- Two kinds of loss function
  - **Adversarial Loss**



  - Cycle-consistence loss

Cycle-GAN Loss function

- Two kinds of loss function
  - **Adversarial Loss**
    - G : X -> Y

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

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

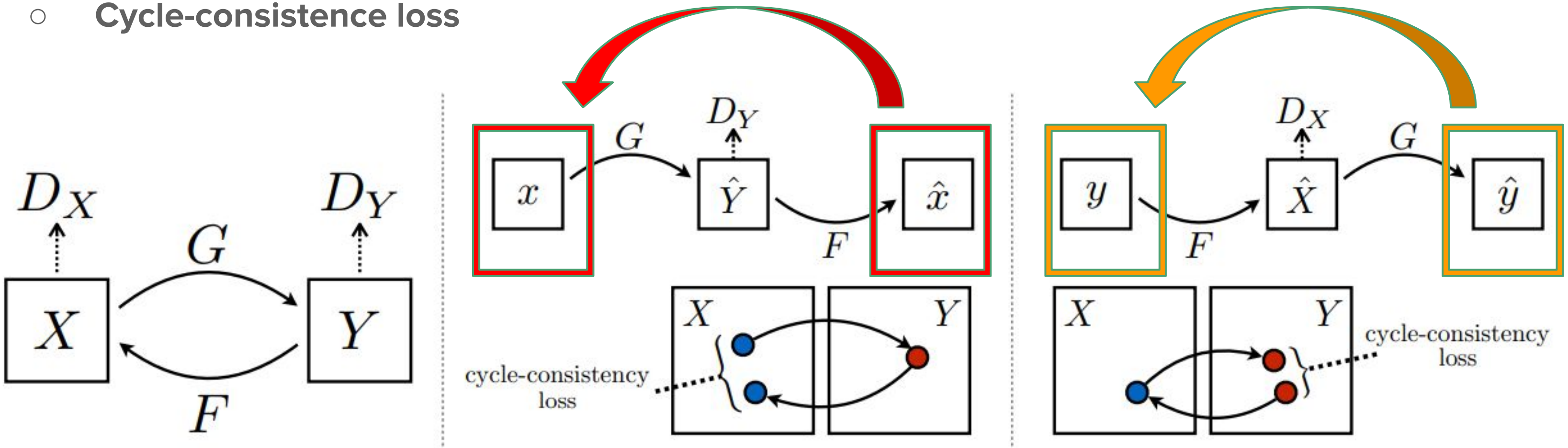
    - F : Y -> X, similar with “G : X -> Y”

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

  - Cycle-consistence loss

Cycle-GAN Loss function

- Two kinds of loss function
  - Adversarial Loss
  - **Cycle-consistence loss**



Cycle-GAN Loss function

- Two kinds of loss function
  - Adversarial Loss
  - **Cycle-consistence loss**

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

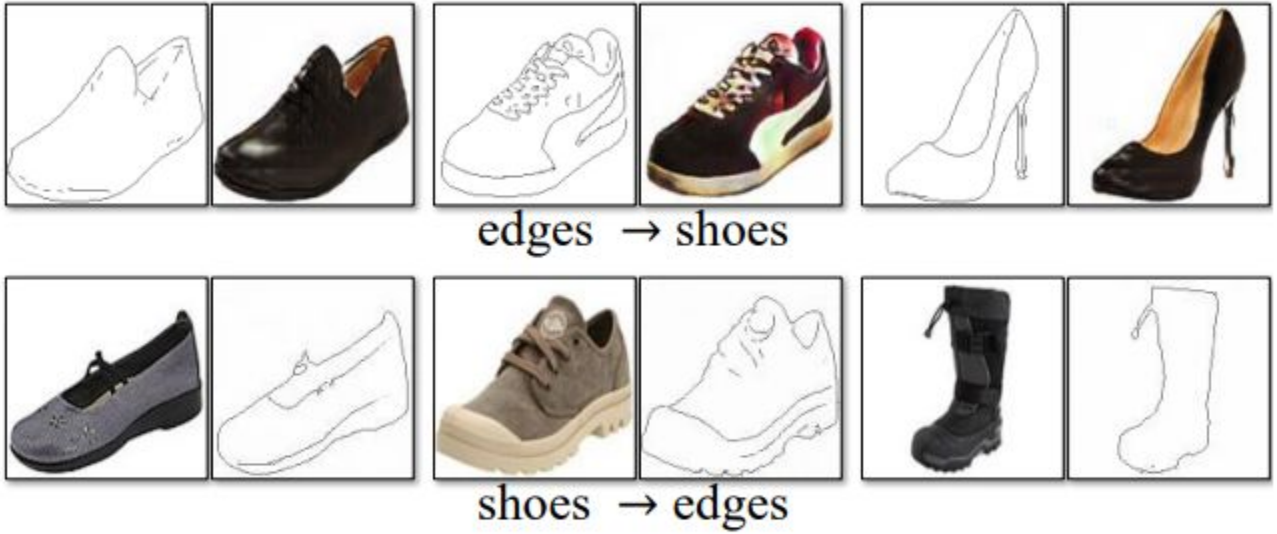
Cycle-GAN Loss function

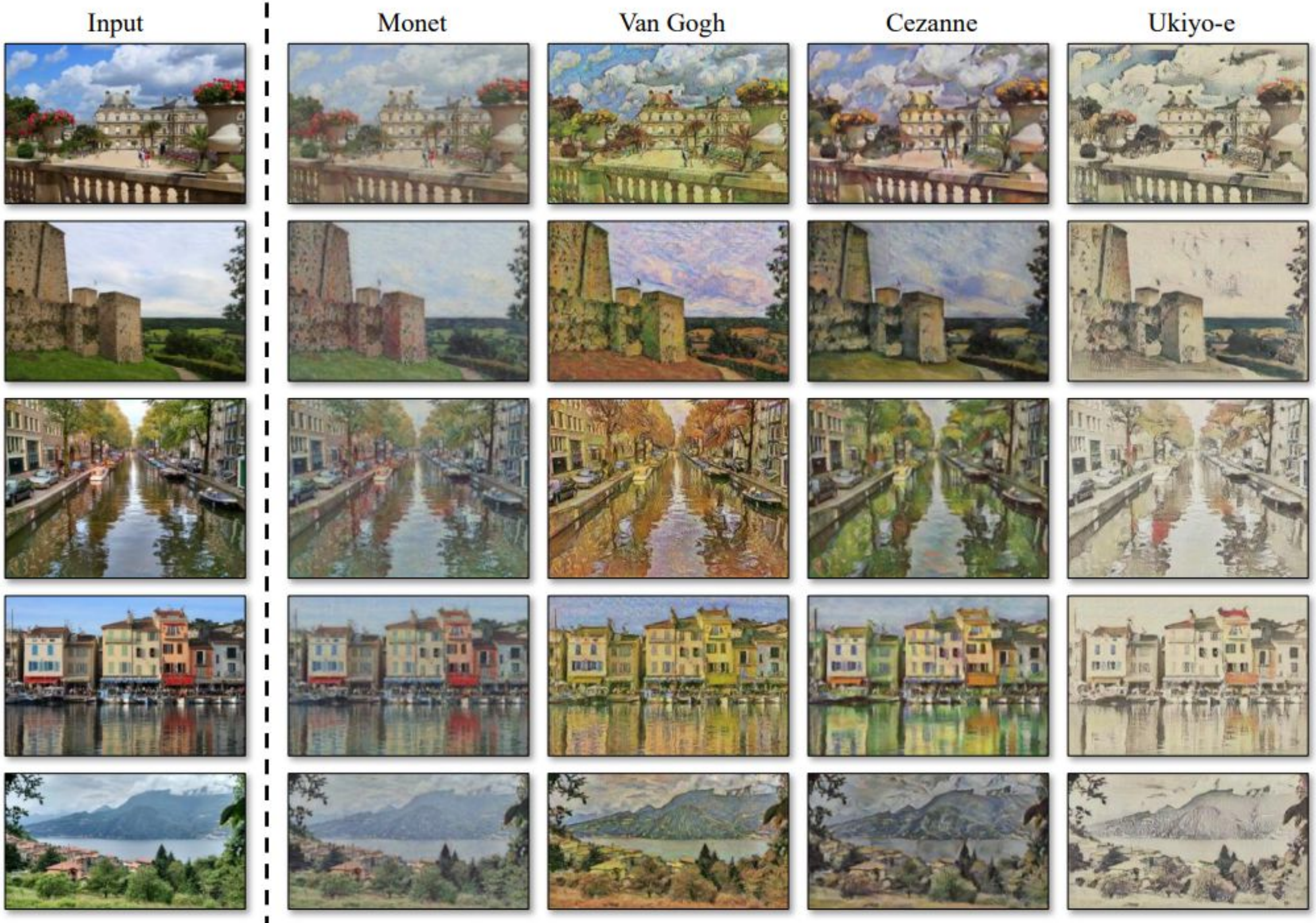
- Two kinds of loss function
  - Adversarial Loss
  - Cycle-consistence loss
- Full objective

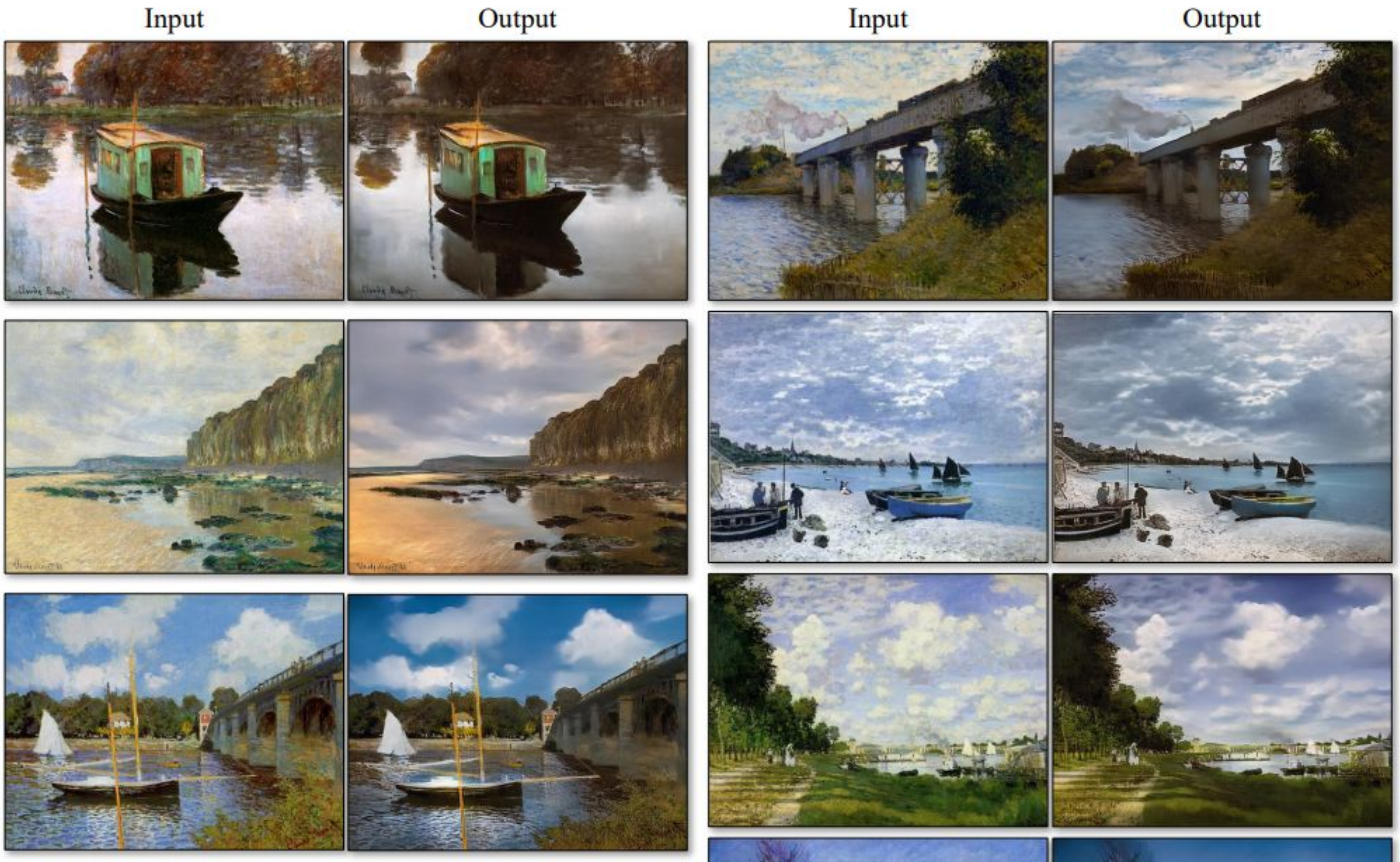
$$
\begin{aligned}
\mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\
& + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\
& + \lambda \mathcal{L}_{\text{cyc}}(G, F),
\end{aligned}
$$

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

Cycle-GAN Result




Input
Monet
Van Gogh
Cezanne
Ukiyo-e


Input
Output
Input
Output

Cycle-GAN Result

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.

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.

Cycle-GAN Result

Cycle-GAN Compare with other GAN

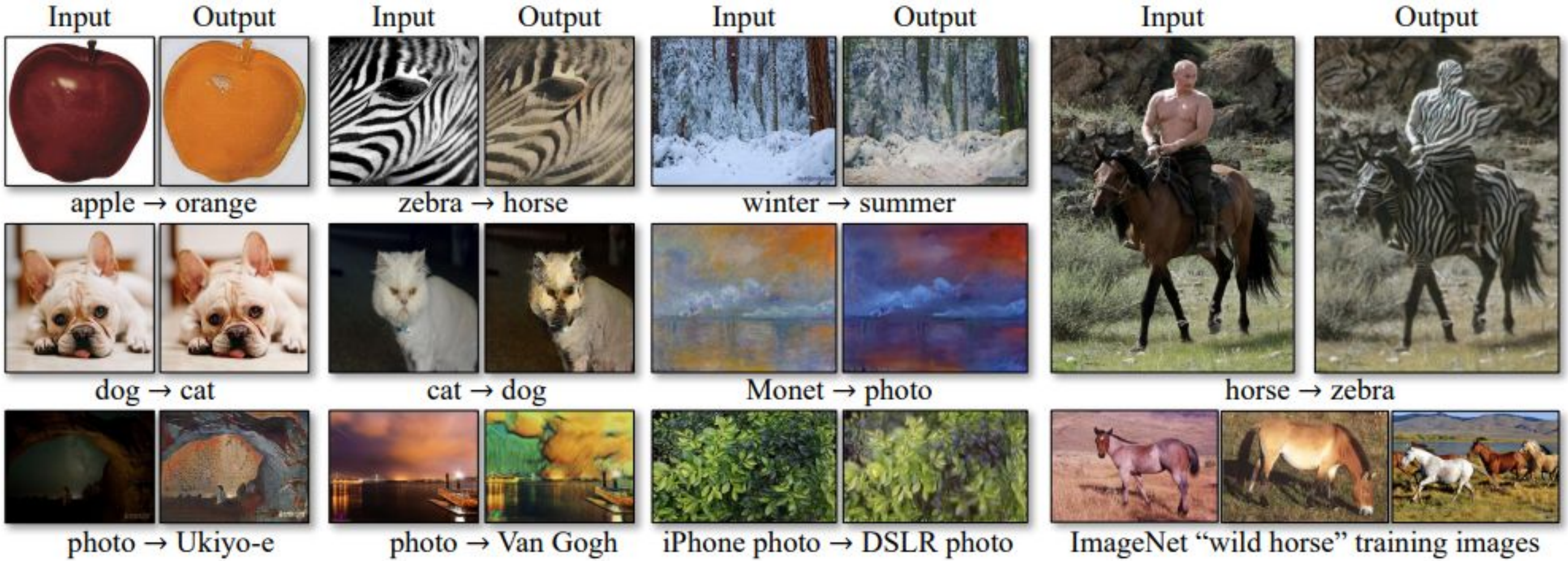
Loss	Per-pixel acc.	Per-class acc.	Class IOU
CoGAN [32]	0.40	0.10	0.06
BiGAN/ALI [9, 7]	0.19	0.06	0.02
SimGAN [46]	0.20	0.10	0.04
Feature loss + GAN	0.06	0.04	0.01
CycleGAN (ours)	**0.52**	**0.17**	**0.11**
pix2pix [22]	0.71	0.25	0.18

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

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.

Cycle-GAN Failure Case



CycleGAN Limit on Video Transfer

- Only care about one image to image
    - It may have time inconsistence problem

Re-Cycle GAN

- ECCV 2018
- CycleGAN + time constrain

Recycle-GAN: Unsupervised Video Retargeting

Aayush Bansal[1], Shugao Ma[2], Deva Ramanan[1], and Yaser Sheikh[1,2]

[1]Carnegie Mellon University [2]Facebook Reality Lab, Pittsburgh
http://www.cs.cmu.edu/-aayushb/Recycle-GAN/

Abstract. We introduce a data-driven approach for unsupervised video retargeting that translates content from one domain to another while preserving the style native to a domain, i.e., if contents of John Oliver's speech were to be transferred to Stephen Colbert, then the generated content/speech should be in Stephen Colbert's style. Our approach combines both spatial and temporal information along with adversarial losses for content translation and style preservation. In this work, we first study the advantages of using spatiotemporal constraints over spatial constraints for effective retargeting. We then demonstrate the proposed approach for the problems where information in both space and time matters such as face-to-face translation, flower-to-flower, wind and cloud synthesis, sunrise and sunset.

1 Introduction

We present an unsupervised data-driven approach for video retargeting that enables the transfer of sequential content from one domain to another while preserving the style of the target domain. Such a content translation and style preservation task has numerous applications including human motion and face translation from one person to other, teaching robots from human demonstration, or converting black-and-white videos to color. This work also finds application in creating visual content that is hard to capture or label in real world settings, e.g., aligning human motion and facial data of two individuals for virtual reality, or labeling night data for a self-driving car. Above all, the notion of content translation and style preservation transcends pixel-to-pixel operation, into a more semantic and abstract human understandable concepts.

Current approaches for retargeting can be broadly classified into three categories. The first set is specifically designed for domains such as human faces [5,41,42]. While these approaches work well when faces are fully visible, they fail when applied to occluded faces (virtual reality) and lack generalization to other domains. The work on paired image-to-image translation [23] attempts to generalize across domain but requires manual supervision for labeling and align-

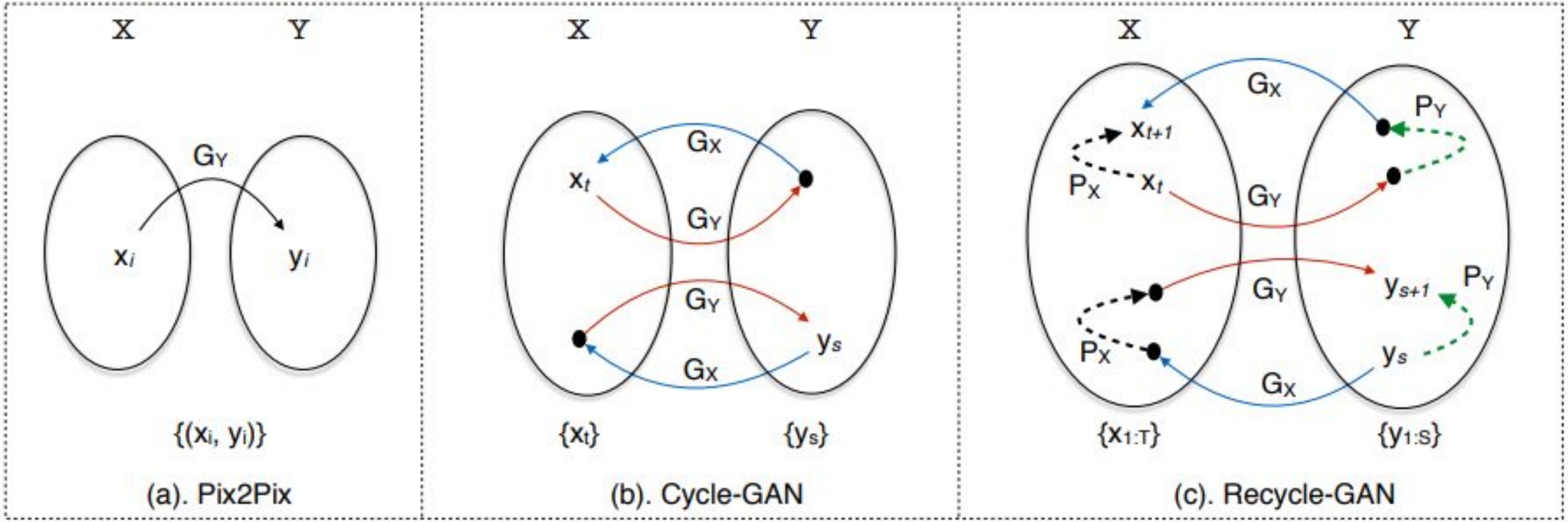
Re-Cycle GAN

- ECCV 2018
- CycleGAN + time constraint

Recycle-GAN: Unsupervised Video Retargeting

Aayush Bansal[1], Shugao Ma[2], Deva Ramanan[1], and Yaser Sheikh[1,2]

[1]Carnegie Mellon University [2]Facebook Reality Lab, Pittsburgh
http://www.cs.cmu.edu/-aayushb/Recycle-GAN/



generalize across domain but requires manual supervision for labeling and align-

Re-Cycle GAN

- ECCV 2018
- CycleGAN + time constrain

Outline

- Goal
- Related Work
- System / Method
- Result

If you want to access this slide , you can use the QR code above

Setup

- Dataset
  - Download the videos from 阿滴英文, and extract the image from the video
- Architecture
  - Based on the CycleGan official code
  - https://github.com/zivzone/pytorch-CycleGAN-and-pix2pix
- Computing Resource
  - 1080 Ti GPU * 1 + i7 8700K CPU * 1

System / Method

- Based on the official released code
  - Official released CycleGAN
  - Official released Re-CycleGAN (Based on CycleGAN code)

- Prepare my own datasets
  - Extract the image from videos downloaded from youtube
  - 16485 + for training
  - 5903 for testing

System / Method

e

on Cy

aded f

Outline

- Goal
- Related Work
- Difficulties / Uniqueness
- Plan
- System / Method
- Result

If you want to access this slide , you can use the QR code above

Compare

Figure.10 CycleGAN(left) and Re-CycleGAN(right)

CycleGAN

C

Re-CycleGAN

- Only train for 1.5 days

Thanks for listening!

If you want to access this slide , you can use the QR code above

Appendix -- April's Fool Video

App