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

J-Y Zhu, T. Park, P. Isola, A. Efros

David Albert

Korea Advanced Institute of Science and Technology

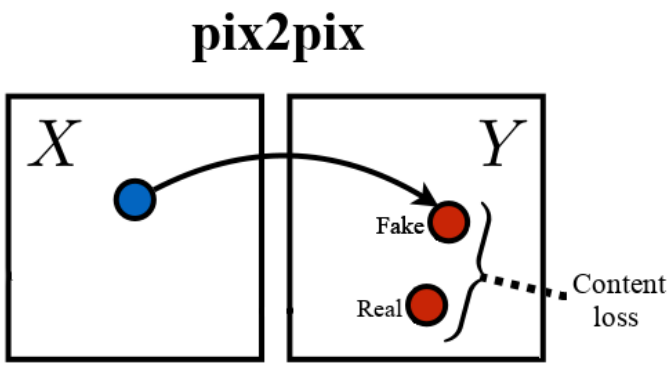
November 26, 2019

Reminder - Domain Transfer

- Learning a mapping $f : X_S \rightarrow X_T$
- Domain (or style) transfer aims to:
  - Keep content of source data
  - Change style to match with target data (or domain)
- General optimizing objective for producing outputs: $\mathcal{L}_{content} + \lambda \mathcal{L}_{style}$

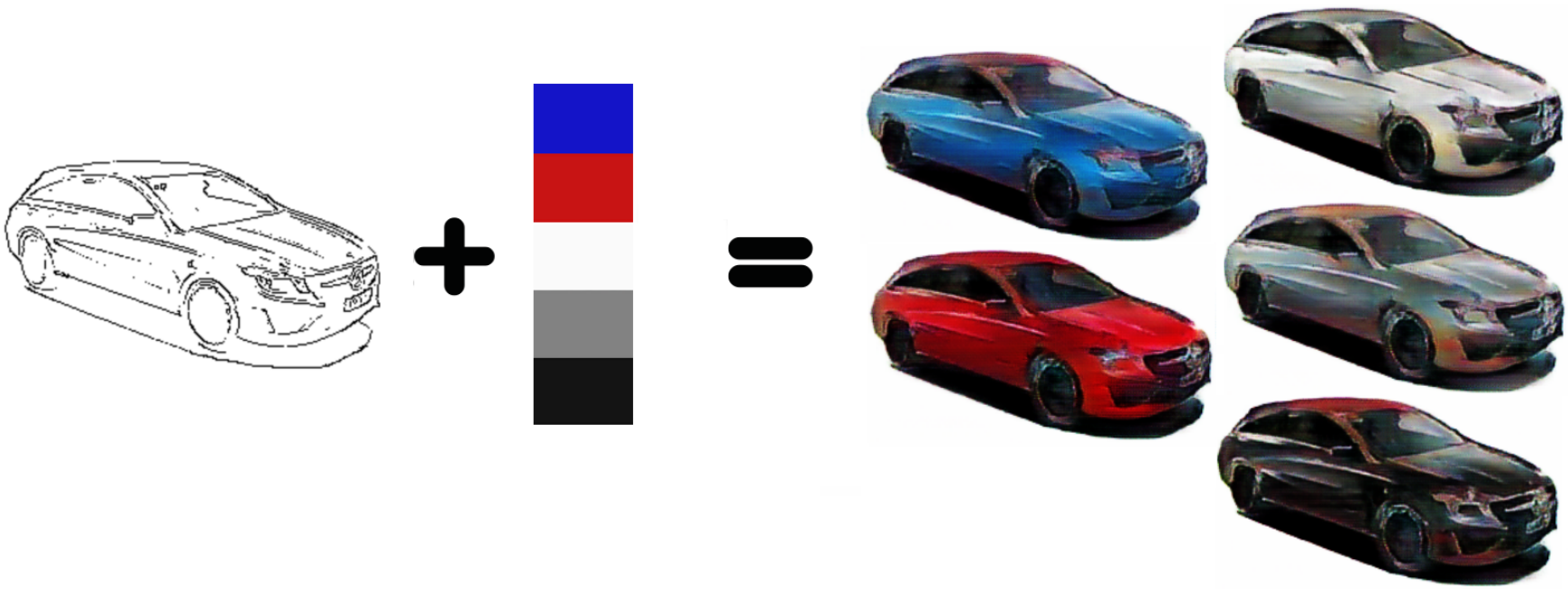
Related work - pix2pix

- **pix2pix** uses **GAN loss** to generate realistic outputs
  - **Style loss** : $\mathcal{L}_{style} = \mathbb{E}_{x_t}\left[logD(x_t)\right] + \mathbb{E}_{x_s}\left[log(1 - D(G(x_s)))\right]$
  - **Content loss** : $\mathcal{L}_{content} = \mathbb{E}_{x_s}\left[\|x_t - G(x_s)\|_1\right]$



Related work - pix2pix

- **pix2pix** uses **GAN loss** to generate realistic outputs
  - **Style loss** : $\mathcal{L}_{style} = \mathbb{E}_{x_t}\left[logD(x_t)\right] + \mathbb{E}_{x_s}\left[log(1 - D(G(x_s)))\right]$
  - **Content loss** : $\mathcal{L}_{content} = \mathbb{E}_{x_s}\left[\|x_t - G(x_s)\|_1\right]$



(-) Obtaining paired training data can be difficult and expensive

Cycle GAN [Zhu et al., 2018]

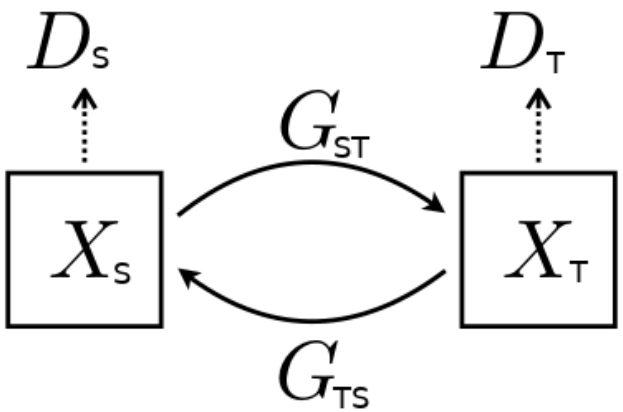
- **Motivation** : Obtaining paired training data can be difficult and expensive

- **Cycle GAN** : Allow style transfer for unpaired images.
- **Method** : Train two generators and two discriminators
  - Learn a mapping $G_{ST} : X_S \rightarrow X_T$
  - Learn a mapping $G_{TS} : X_T \rightarrow X_S$

Cycle GAN [Zhu et al., 2018]

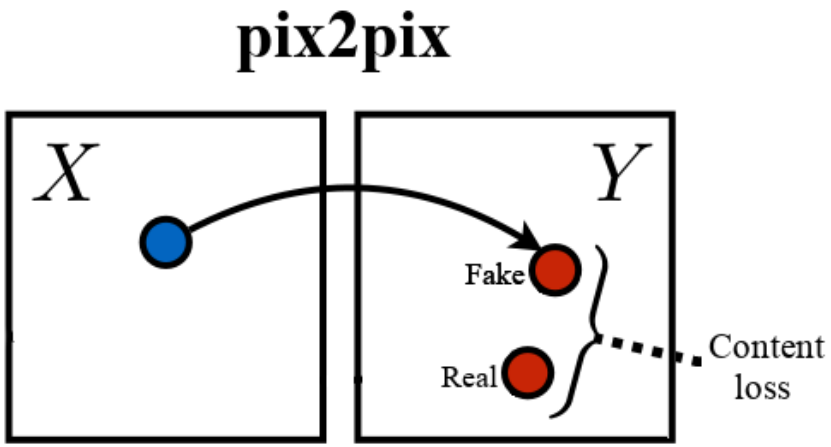
- **Motivation** : Obtaining paired training data can be difficult and expensive
- **Cycle GAN** : Allow style transfer for unpaired images.
- **Method** : Train two generators and two discriminators
  - Learn a mapping $G_{ST} : X_S \rightarrow X_T$
  - Learn a mapping $G_{TS} : X_T \rightarrow X_S$
  - **Style loss** : $\mathcal{L}_{style} = \mathcal{L}_{LSGAN}(G_{ST}, D_T, X_S, X_T) + \mathcal{L}_{LSGAN}(G_{TS}, D_S, X_T, X_S)$

    where $\mathcal{L}_{LSGAN}(G, D, X, Y) = \mathbb{E}_x\left[(D(G(x)) - 1)^2\right] + \mathbb{E}_y\left[(D(y) - 1)^2\right]$



Cycle GAN [Zhu et al., 2018]

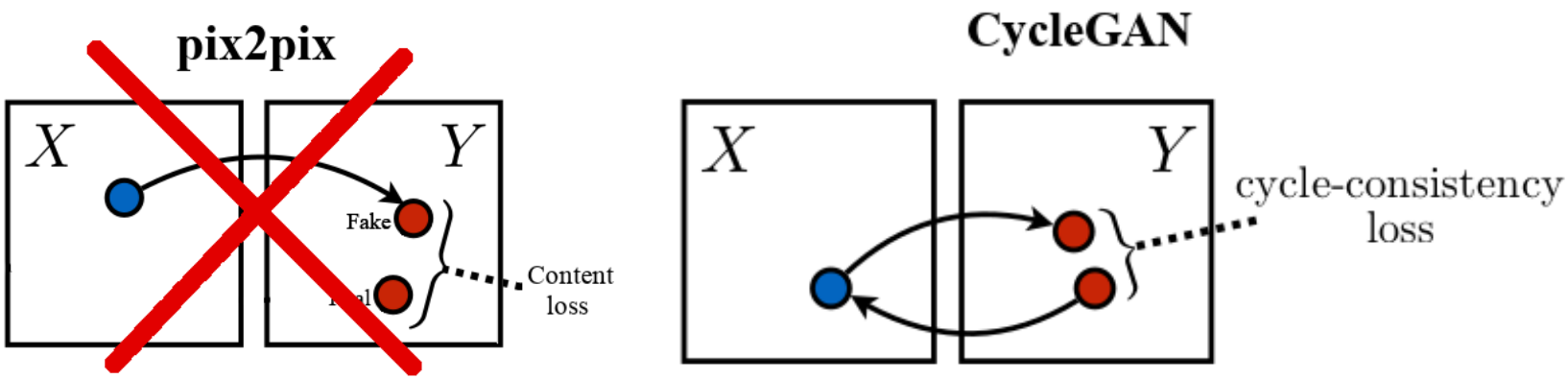
- **Motivation** : Obtaining paired training data can be difficult and expensive
- **Cycle GAN** : Allow style transfer for unpaired images.
- **Method** : Train two generators and two discriminators
  - Learn a mapping $G_{ST} : X_S \rightarrow X_T$
  - Learn a mapping $G_{TS} : X_T \rightarrow X_S$
  - **Content loss** : $\mathcal{L}_{content} = ?$



Cycle GAN [Zhu et al., 2018]

- **Motivation** : Obtaining paired training data can be difficult and expensive
- **Cycle GAN** : Allow style transfer for unpaired images.
- **Method** : Train two generators and two discriminators
  - Learn a mapping $G_{ST} : X_S \rightarrow X_T$
  - Learn a mapping $G_{TS} : X_T \rightarrow X_S$
  - **Content loss** :

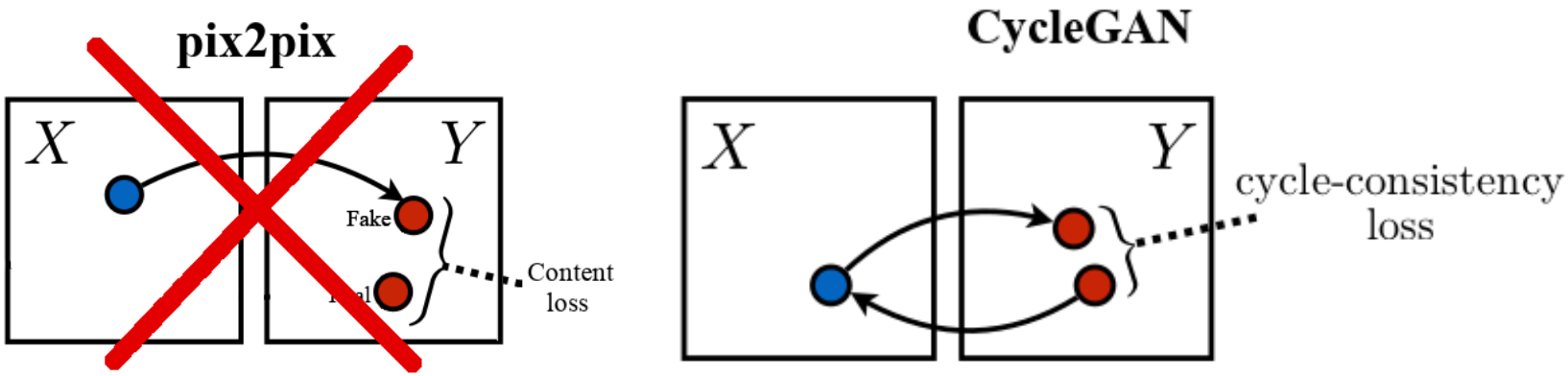
$$\mathcal{L}_{content} = \mathcal{L}_{cyc} = \mathbb{E}_{x_s} \left[\| G_{TS}(G_{ST}(x_s)) - x_s \|_1 \right] + \mathbb{E}_{x_t} \left[\| G_{ST}(G_{TS}(x_t)) - x_t \|_1 \right]$$



Cycle GAN [Zhu et al., 2018]

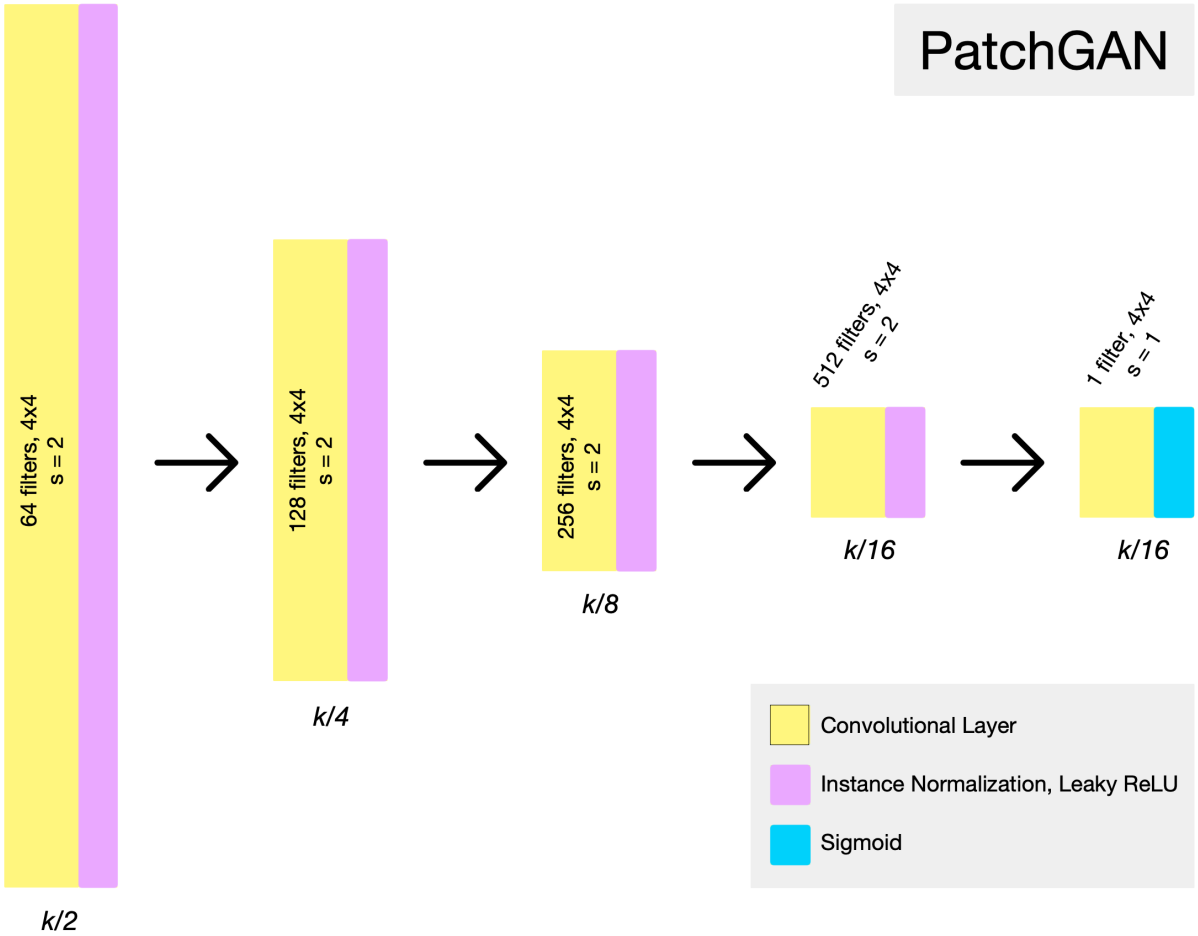
- **Motivation** : Obtaining paired training data can be difficult and expensive
- **Cycle GAN** : Allow style transfer for unpaired images.
- **Method** : Train two generators and two discriminators
  - Learn a mapping $G_{ST} : X_S \rightarrow X_T$
  - Learn a mapping $G_{TS} : X_T \rightarrow X_S$
  - **Global objective** :

$$arg \min_{G_{ST}, G_{TS}} \max_{D_S, D_T} \left[\mathcal{L}_{LSGAN}(G_{ST}, D_T) + \mathcal{L}_{LSGAN}(G_{TS}, D_S) + \lambda \mathcal{L}_{cyc} \right]$$



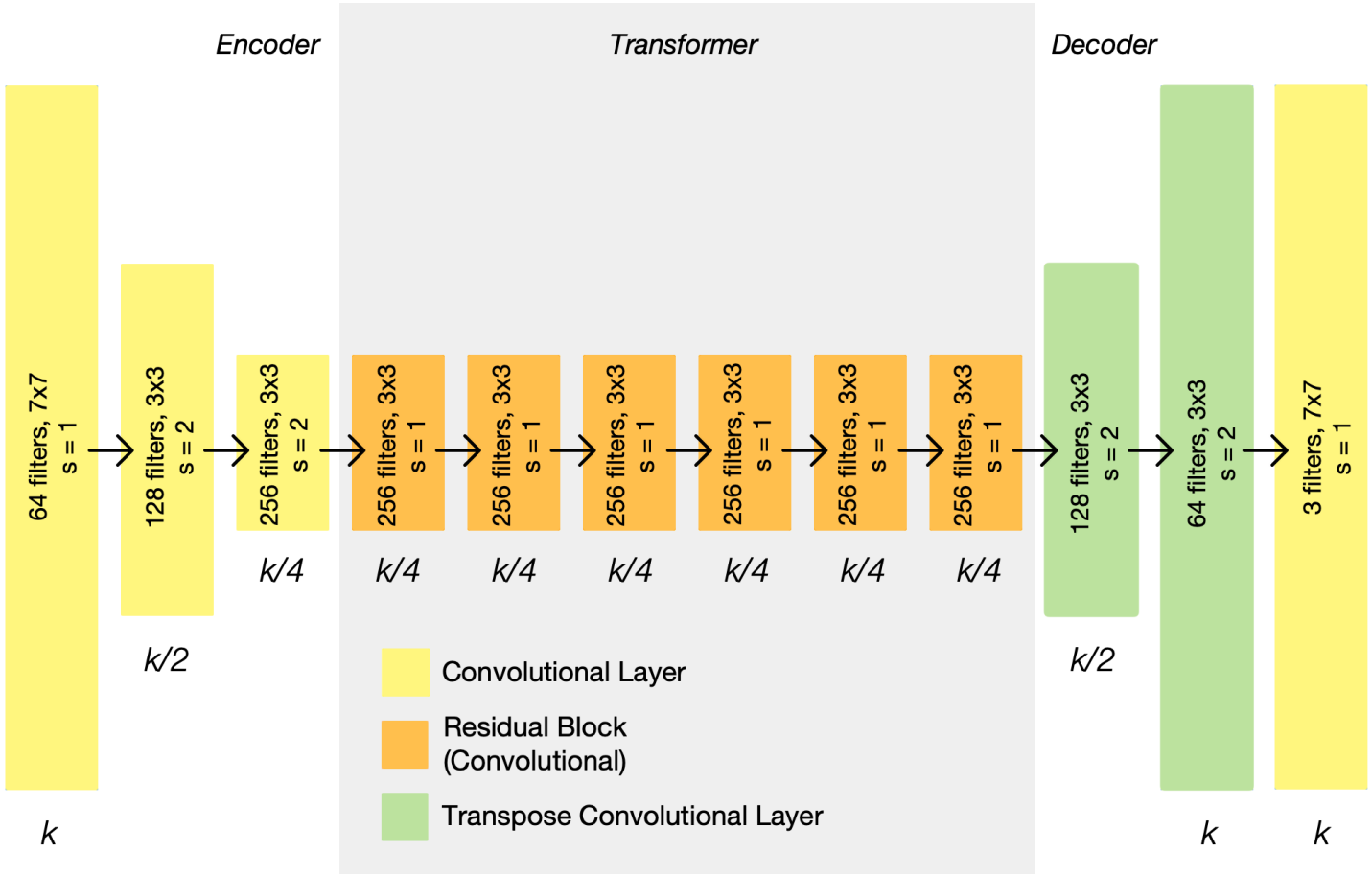
Architecture

- **Discriminator architecture**



Architecture

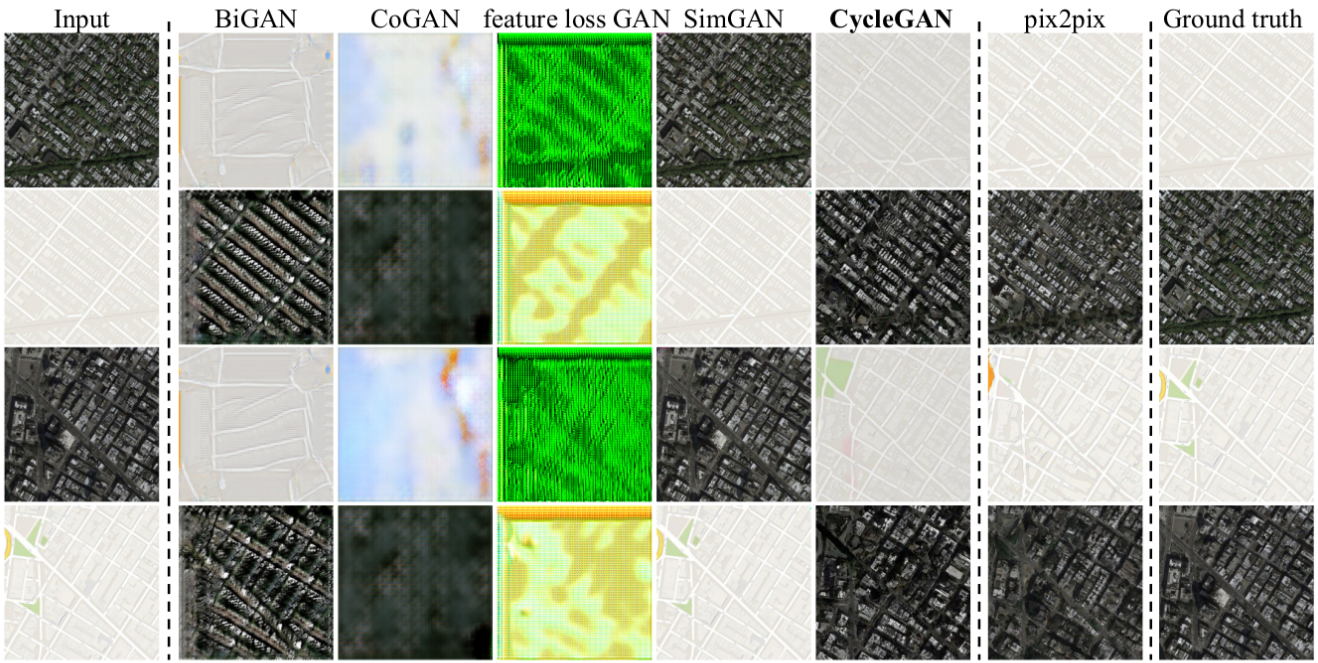
- **Generator architecture**



Evaluation - AMT Perceptual study

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

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



Results

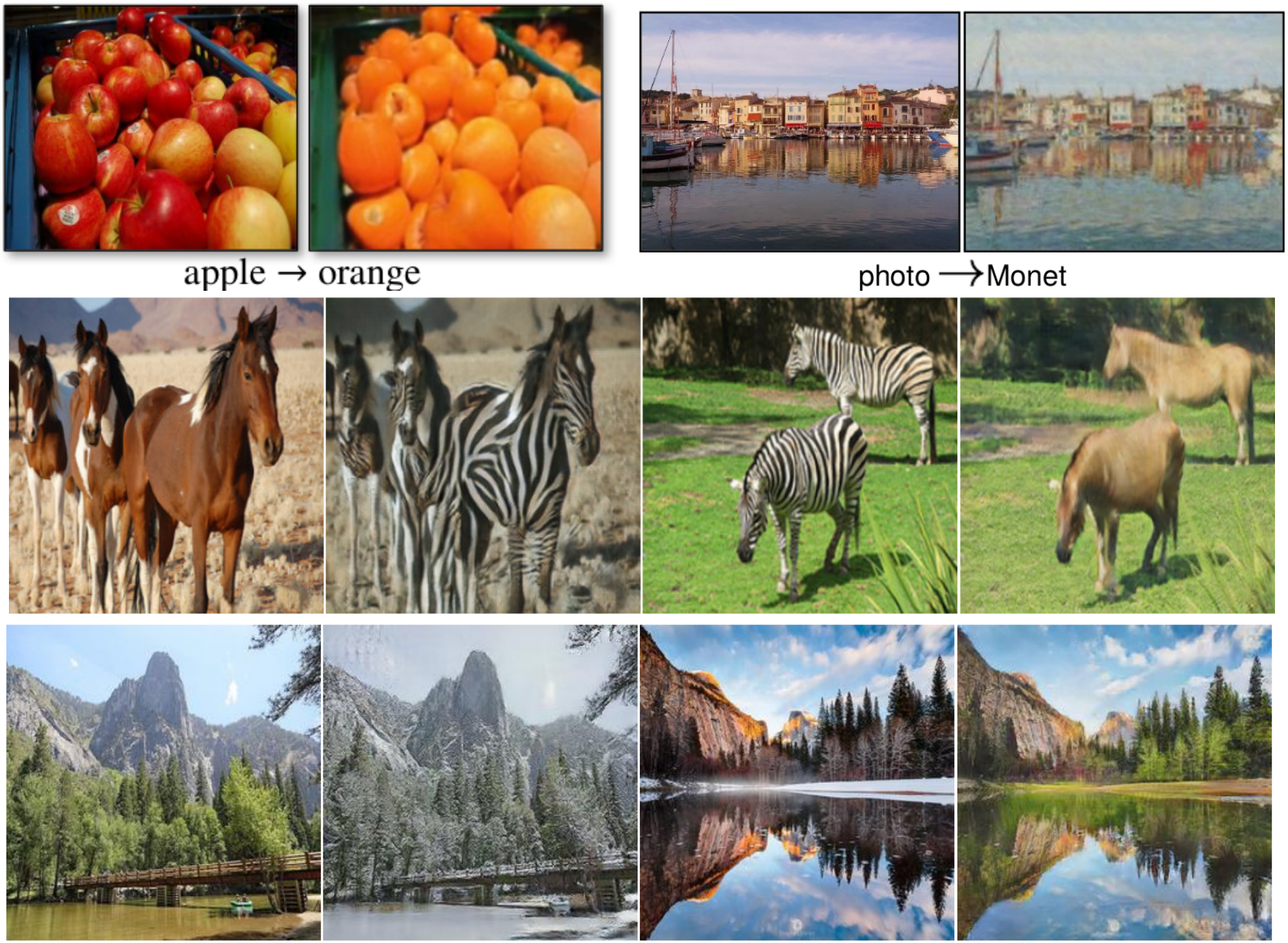


Figure 1: horse ↔ zebra and summer ↔ winter

Conclusion

- Domain transfer for unpaired images
- Train 4 neural networks jointly
  $\rightarrow$ can be instable
- Good results when good parameters

References

- *Lecture 11 : Domain Transfer and Adaptation* [Sangwoo Mo]

  → `https://alinlab.kaist.ac.kr/ai602_2019.html`

- *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks* [Zhu et al., 2018]

  → `https://arxiv.org/abs/1703.10593`

- *Image-to-Image Translation with Conditional Adversarial Networks* [Isola et al., 2016]

  → `https://arxiv.org/abs/1611.07004`

- *Least Squares Generative Adversarial Networks* [Mao et al., 2017]

  → `https://arxiv.org/abs/1611.04076`

- *Draw a car project* [G. Chen, D. Albert, 2018]

  → `https://georgezjchen.github.io/draw-a-car-build/`

Appendix - LSGAN

- **Remind**: GAN minimizes JS-divergence, WGAN minimizes Wasserstein divergence
- **LSGAN** [Mao et al., 2017]
  - LSGAN can be interpreted as minimizing the Pearson χ^2 divergence

    $\rightarrow$ i.e. : minimize $\chi^2_{pearson}(p_d + p_g \| 2p_g) = \int_{\chi} \frac{\left(2p_g(x) - (p_d(x) + p_g(x))\right)^2}{p_d(x) + p_g(x)} dx$
  - (+) Able to generate higher quality images
  - (+) Stabilize the model

Appendix - Evaluation metrics

- AMT perceptual studies
  - "real vs fake" studies on Amazon Mechanical Turk (AMT)
  - participants are shown sequence of pairs of images (1 fake and 1 real) and asked to click on the real image
- FCN score
  - uses FCN to predicts a label map for generated photo
  - compares label map against the ground truth labels using standard semantic segmentation metrics
- Semantic segmentation metrics
  - Standard metrics from Cityscapes banchmark
  - Per-pixel accuracy, Per-class accuracy, mean class Intersection-Over-Union