Image-to-Image Translation with Conditional Adversarial Networks

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros
Berkeley AI Research (BAIR) Laboratory, UC Berkeley

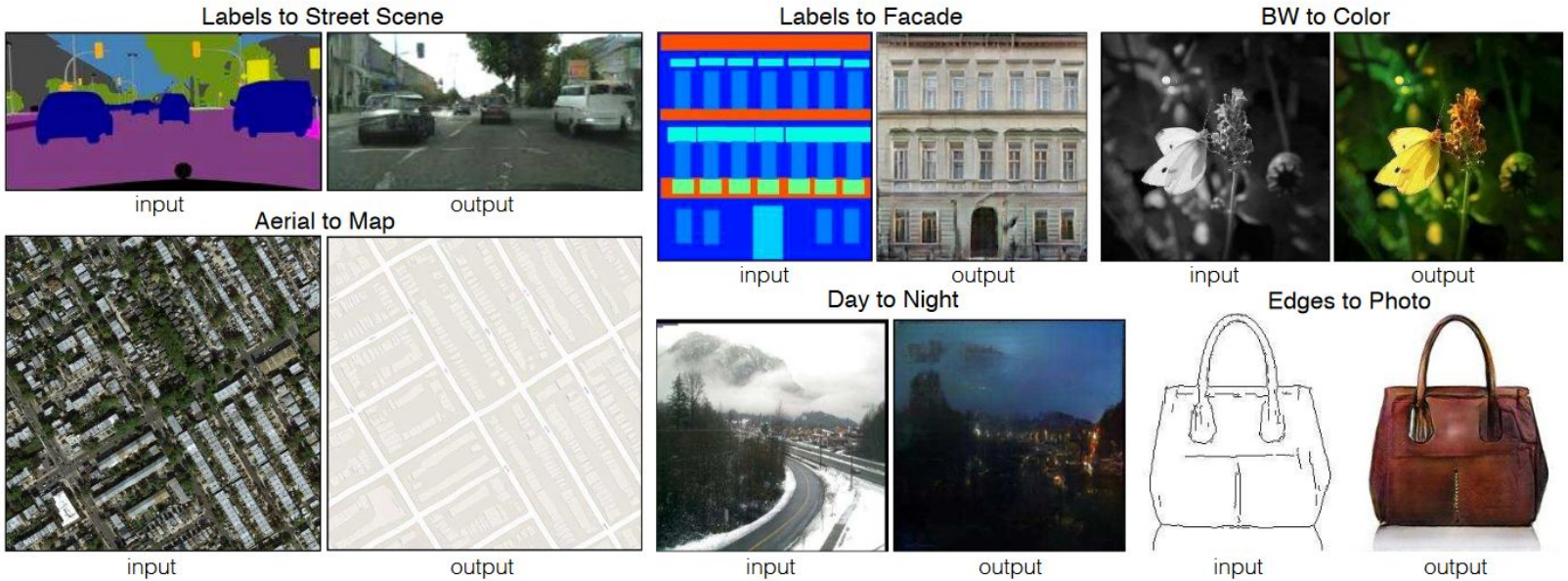
ECS 289G: Visual Recognition

Presenters: Terry Yang and Mingyang Zhou

Agenda

- Motivating the problem
- Introduction to GANs
- Method
- Dataset & tasks
- Evaluation Metrics
- Results
- Strength & Weakness
- Future Extension

Introduction: Image-to-Image Translation



Isola, Phillip, et al. "Image-To-Image Translation With Conditional Adversarial Networks." CVPR, 2017.

Previous Methods

A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. CVPR, 2015.

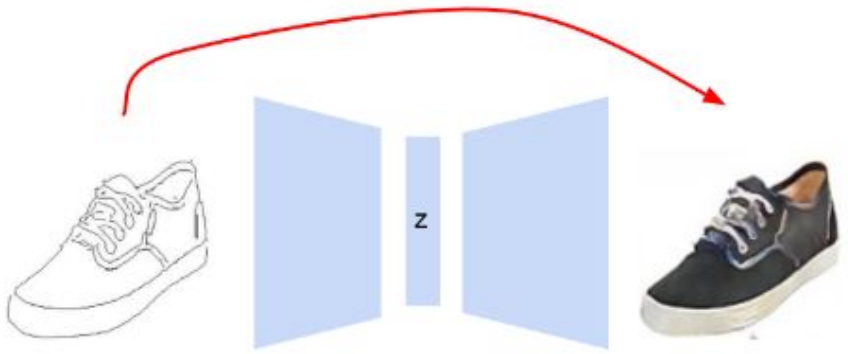
R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. ECCV, 2016.

T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. Sketch2photo: internet image montage. ACM, 2009

Required loss functions and architectures designed specifically for the task at hand.

Motivation

Goal: A general-purpose solution to image-to-image translation problems.



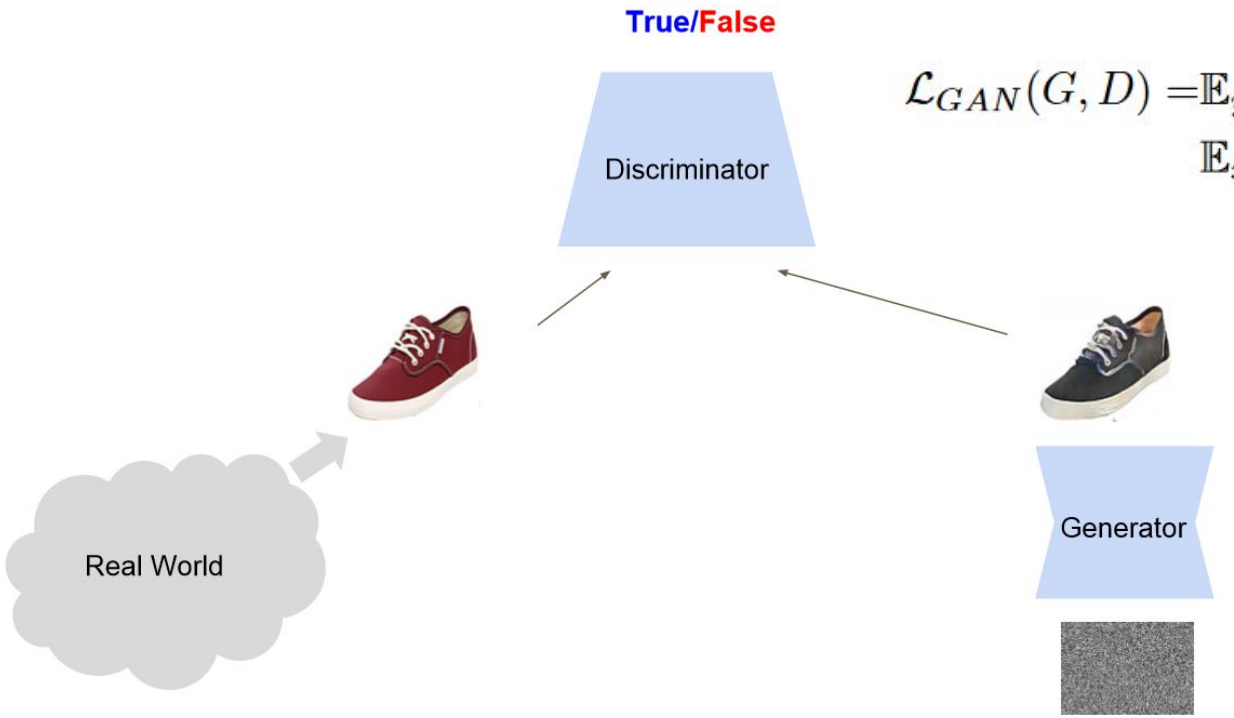
This paper: Use the same architecture and objective for each image-to-image translation task.

Garcia, Victor. "Image-to-Image Translation with Conditional Adversarial Networks." 25 Nov 2016. UPC Computer Vision Reading Group, Universitat Politècnica de Catalunya, Microsoft Powerpoint presentation

Quick Overview of Generative Adversarial Networks

©2014-2016 Ociacia

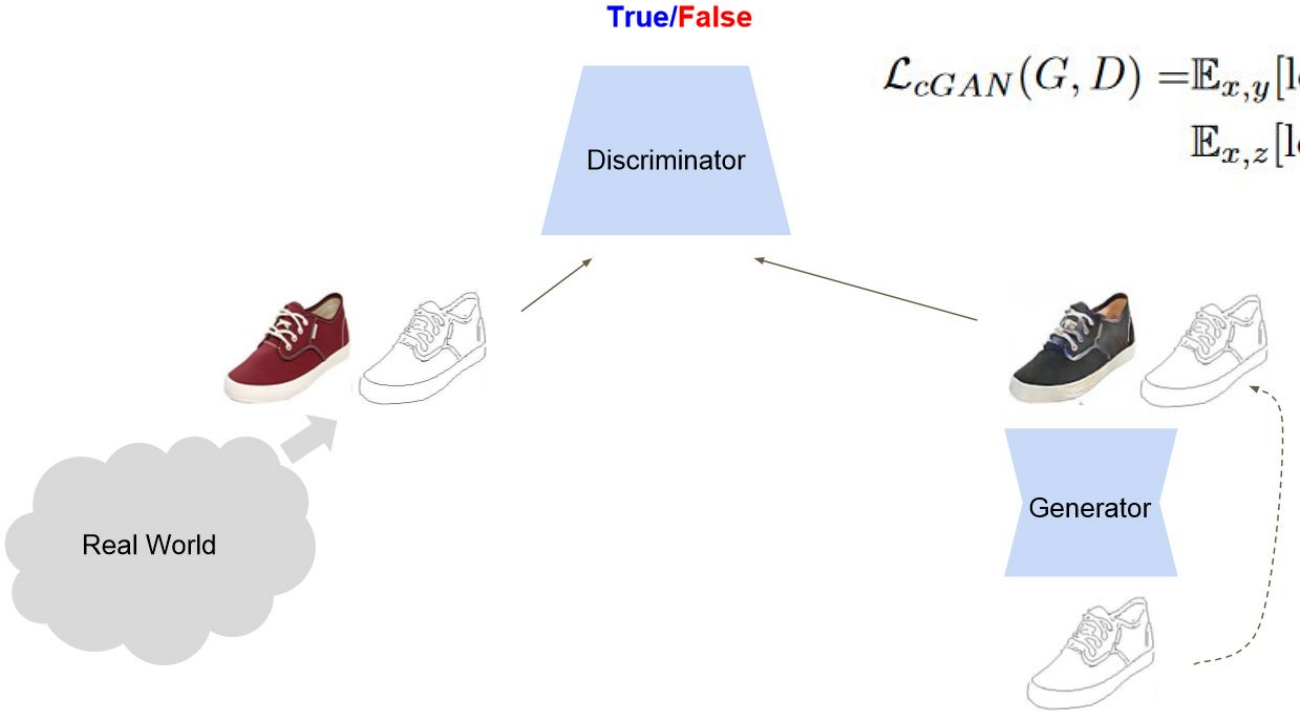
Generative Adversarial Networks (GANs)



$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_y[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x, z))].$$

Garcia, Victor. "Image-to-Image Translation with Conditional Adversarial Networks." 25 Nov 2016. UPC Computer Vision Reading Group, Universitat Politècnica de Catalunya, Microsoft Powerpoint presentation

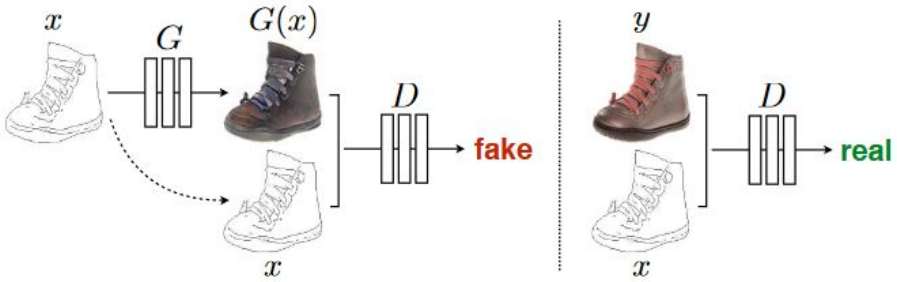
Conditional Generative Adversarial Networks (cGANs)



$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))],$$

Garcia, Victor. "Image-to-Image Translation with Conditional Adversarial Networks." 25 Nov 2016. UPC Computer Vision Reading Group, Universitat Politècnica de Catalunya, Microsoft Powerpoint presentation

Objective



$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x,z))], \quad (1)$$

$$\mathcal{L}_{GAN}(G,D) = \mathbb{E}_{y}[\log D(y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x,z))]. \quad (2)$$

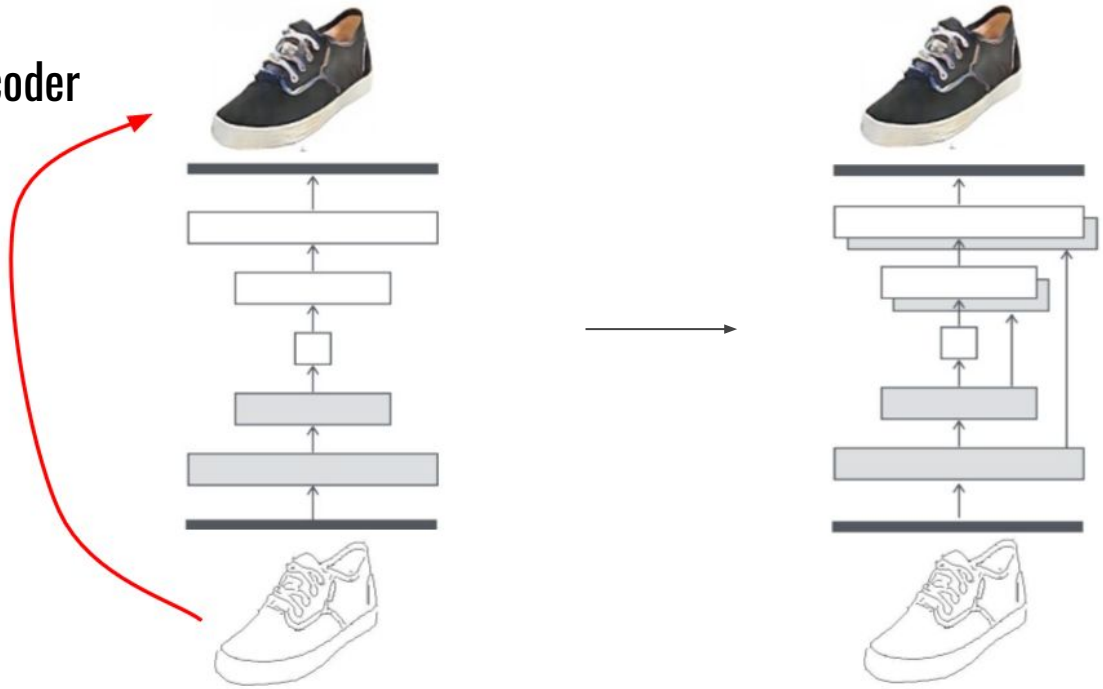
$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1]. \quad (3)$$

$$\boxed{G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G). \quad (4)}$$

Isola, Phillip, et al. "Image-To-Image Translation With Conditional Adversarial Networks." CVPR, 2017.

Generator

Encoder-Decoder Network

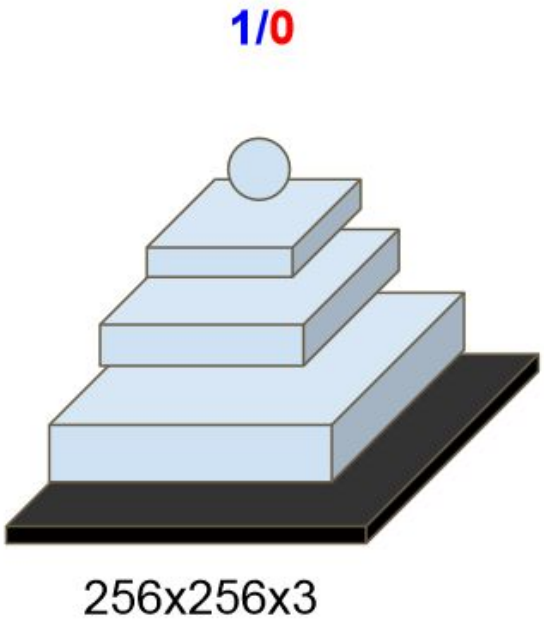


U-Net

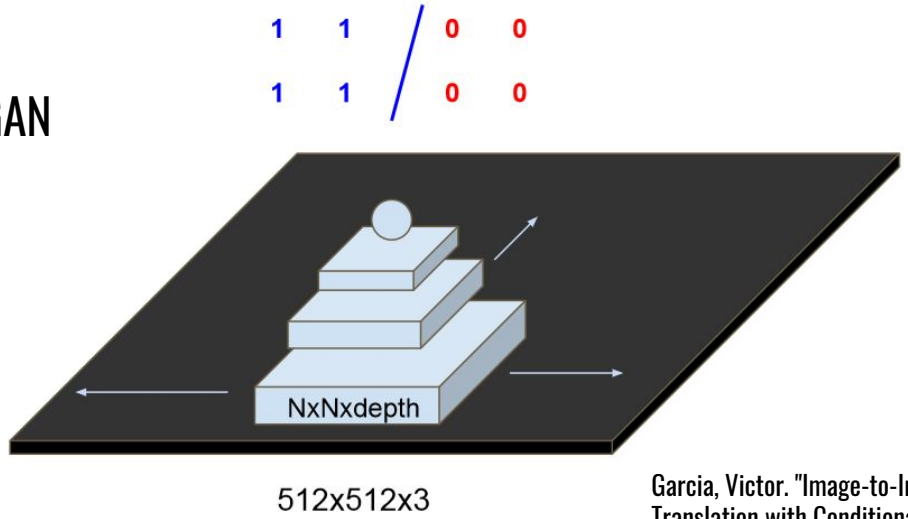
Skip Connections

Garcia, Victor. "Image-to-Image Translation with Conditional Adversarial Networks." 25 Nov 2016. UPC Computer Vision Reading Group, Universitat Politècnica de Catalunya, Microsoft Powerpoint presentation

Discriminator

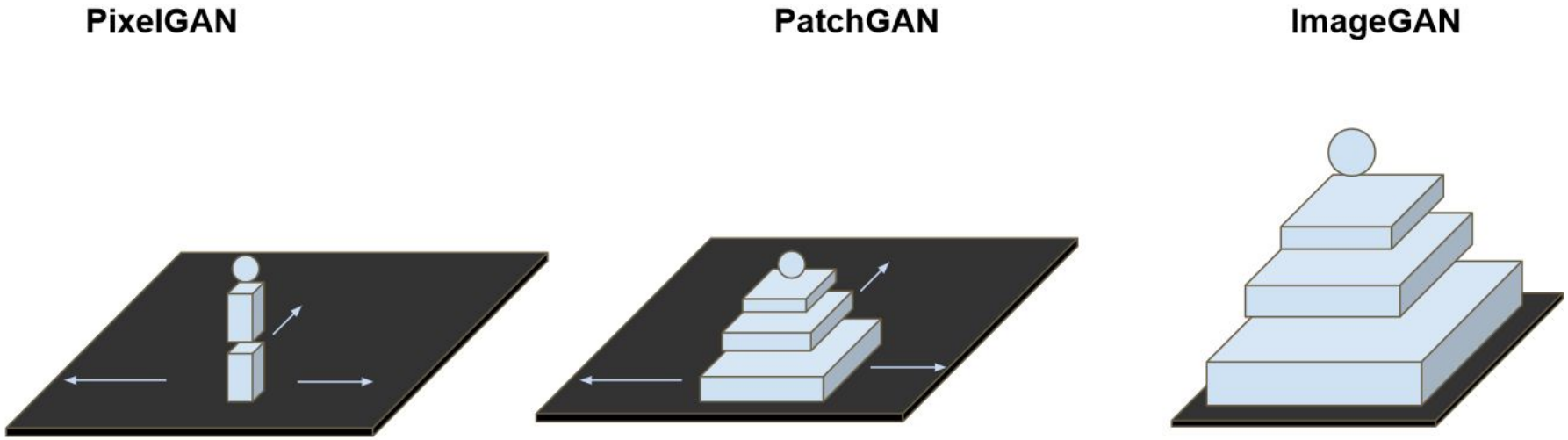


PatchGAN



Garcia, Victor. "Image-to-Image Translation with Conditional Adversarial Networks." 25 Nov 2016. UPC Computer Vision Reading Group, Universitat Politècnica de Catalunya, Microsoft Powerpoint presentation

Discriminator



Garcia, Victor. "Image-to-Image Translation with Conditional Adversarial Networks." 25 Nov 2016. UPC Computer Vision Reading Group, Universitat Politècnica de Catalunya, Microsoft Powerpoint presentation

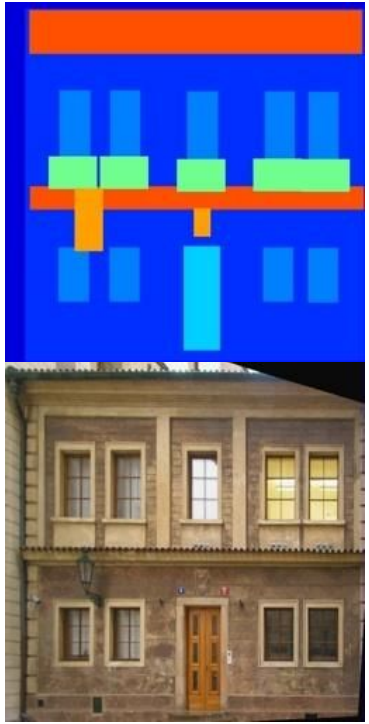
Optimization and Inference

- Alternate between one gradient descent step on D and one gradient descent step on G
- Minibatch SGD and Adam optimization
- At inference time, the generator net is run in exactly the same manner as during training phase

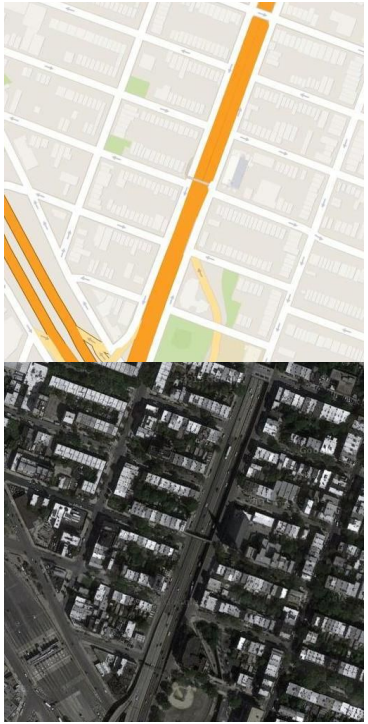
Isola, Phillip, et al. "Image-To-Image Translation With Conditional Adversarial Networks." CVPR, 2017.

Datasets & Tasks

CityScape



CMP Facades



Map to Aerial

Datasets & Tasks

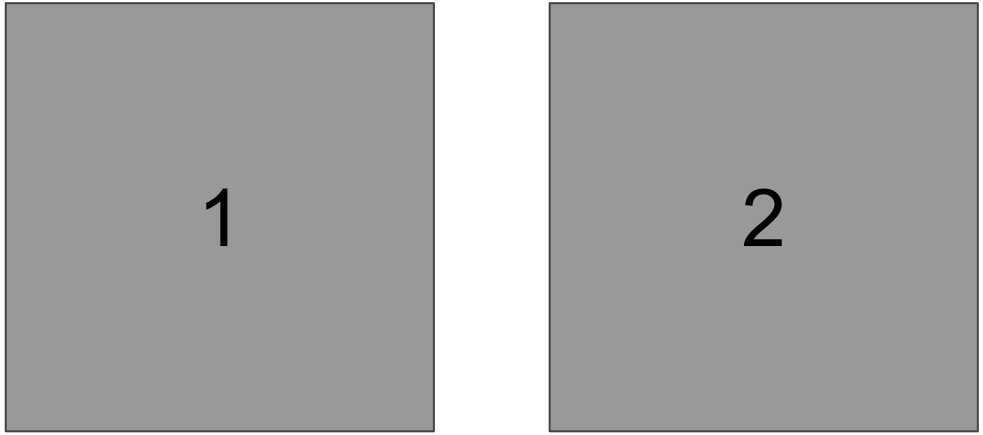
Edge to Photo

Sketch to Photo

Day to Light

Evaluation Metrics

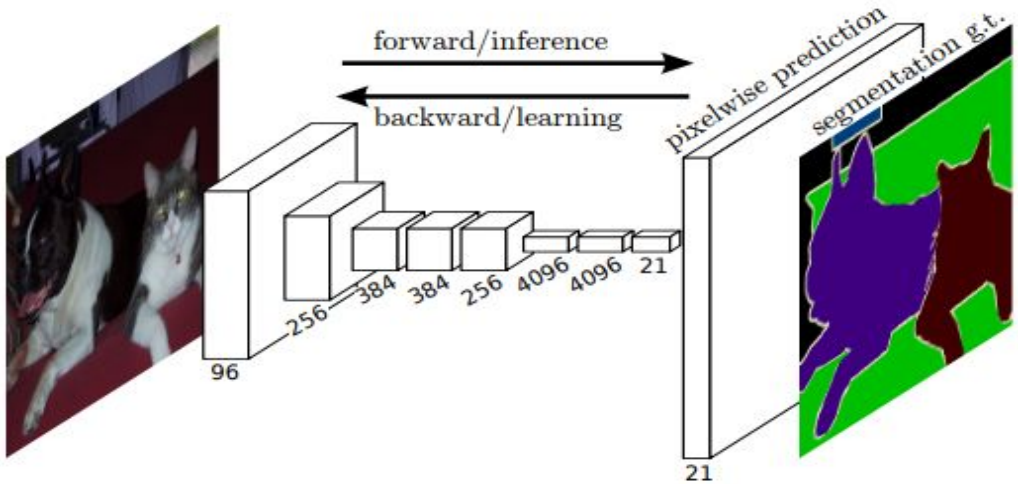
AMT Perceptual Studies



Left or Right?

Evaluation Metrics

FCN-score



pixel accuracy: $\sum_i n_{ii} / \sum_i t_i$

mean accuraccy: $(1/n_{\text{cl}}) \sum_i n_{ii}/t_i$

mean IU: $(1/n_{\text{cl}}) \sum_i n_{ii} / \left(t_i + \sum_j n_{ji} - n_{ii}\right)$

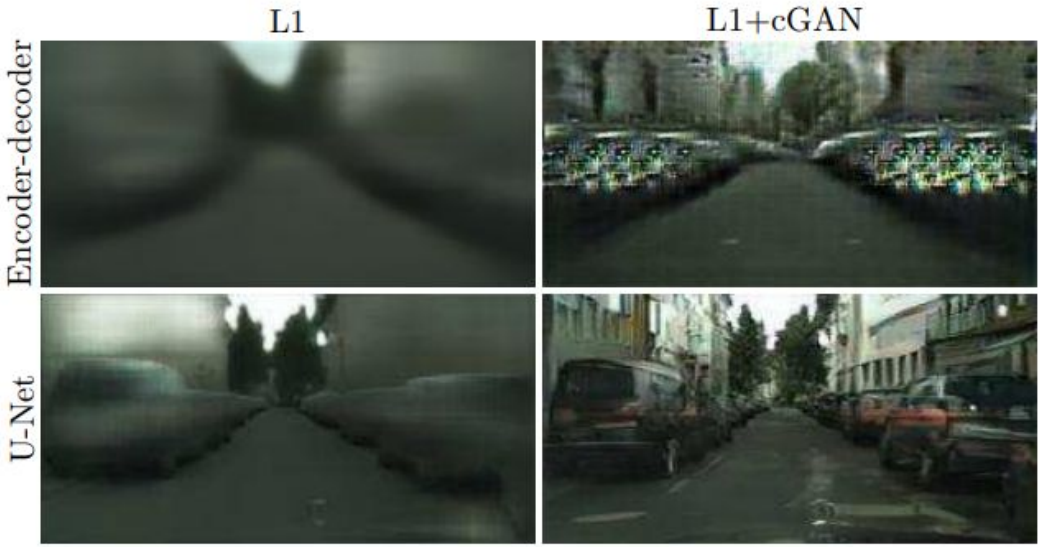
Experiment Results

Analysis of the Objective Function

Loss	Per-pixel acc.	Per-class acc.	Class IOU
L1	0.42	0.15	0.11
GAN	0.22	0.05	0.01
cGAN	0.57	0.22	0.16
L1+GAN	0.64	0.20	0.15
L1+cGAN	**0.66**	**0.23**	**0.17**
Ground truth	0.80	0.26	0.21

Experiment Results

Analysis of generator architecture



Experiment Results

Receptive Field of Discriminator

Discriminator receptive field	Per-pixel acc.	Per-class acc.	Class IOU
1×1	0.39	0.15	0.10
16×16	0.65	0.21	**0.17**
70×70	**0.66**	**0.23**	**0.17**
286×286	0.42	0.16	0.11

Experiment Results

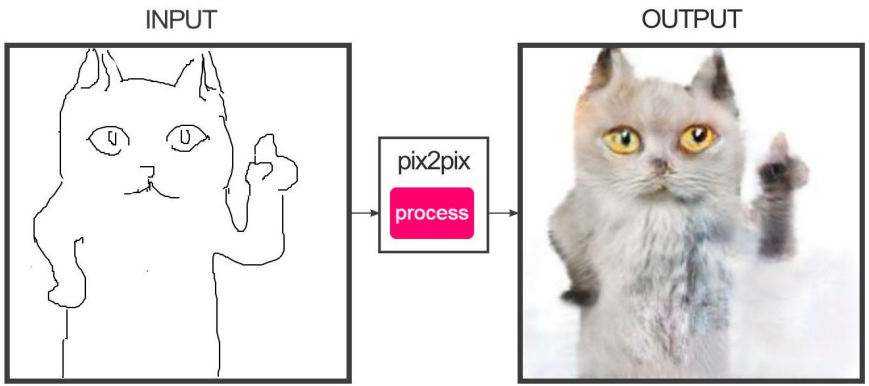
Perceptual Validation

Loss	Photo → Map % Turkers labeled *real*	Map → Photo % Turkers labeled *real*
L1	2.8% ± 1.0%	0.8% ± 0.3%
L1+cGAN	6.1% ± 1.3%	**18.9% ± 2.5%**

Method	% Turkers labeled *real*
L2 regression from [58]	16.3% ± 2.4%
Zhang et al. 2016 [58]	**27.8% ± 2.7%**
Ours	22.5% ± 1.6%

Experiment Results

Rich Community Driven Results



Edges2Cats

Background Masking

Strength & Weaknesses

Strength:

1. An easily generalizable structure to conduct image-to-image translation tasks.
2. A good combination of existing techniques and well-designed loss function to ensure the generation of high-quality synthesized image.

Weakness:

1. Required large number of 1-to-1 paired images for training, which are expensive to collect.
2. Does not provide decent comparisons between this network with the other task-specific model to defend its claim for good generalizability.

Future Extension

---

1. Modify the structure such that it can learn the translation mapping with unpaired images in two different domains.

2. Further improvement to generate photorealistic image.