Conditional GANs, Image-to-Image Translation
Jun-Yan Zhu
16-726, Spring 2022

Many slides from Phillip Isola, Ming-Yu Liu, Xun Huang, etc.
Logistics

• HW2 gather town party Mon 8-10 pm
• No class next week (due to Spring break)
• HW 1 Class Choice Award:
  – Vote by the end of Wed.
  – Winner will be announced on 03/14 (Mon)
Problem Statement

**Goal**: synthesize a realistic photograph given an input image
Learning objective

$$\min_G \max_D \mathbb{E}_x [\log (1 - D(G(x)))] + \mathbb{E}_y [\log D(y)]$$

Pix2pix [Isola et al., 2016]
Learning objective

$$\min_G \max_D \mathbb{E}_x[\log(1 - D(G(x)))] + \mathbb{E}_y[\log D(y)]$$

Pix2pix [Isola et al., 2016]
Learning objective

\[
\min_G \max_D \mathbb{E}_x [\log (1 - D(x, G(x))] + \mathbb{E}_{x,y} [\log D(x, y)]
\]

Pix2pix [Isola et al., 2016]
#edges2cats [Christopher Hesse]

@ivymyt

Ivy Tasi

@matthematician

Vitaly Vidmirov @vvid

https://affinelayer.com/pixsrv/
Input: Sketch $\rightarrow$ Output: Photo

Input: Grayscale $\rightarrow$ Output: Color

Discriminator

Real or fake pair?

Pix2pix [Isola et al., 2016]
Automatic Colorization with pix2pix

Data from [Russakovsky et al. 2015]
Learning vs. Exemplar-based
Hybrid Method

Output from exemplar-based method

Semi-Parametric Image Synthesis [Qi et al., 2018]
Learning-based

[Isola et al], [Wang et al]
[Park et al], SEAN [Zhu et al]

Hybrid method

SIMS [Qi et al]

Exemplar-based

[Johnson et al], [Lalonde et al]
[Tao et al], [Bansal et al]

<table>
<thead>
<tr>
<th>Speed</th>
<th>Learning-based</th>
<th>Hybrid method</th>
<th>Exemplar-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>☀️</td>
<td>☠️</td>
<td>☠️</td>
<td>☠️</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Local realism</th>
<th>Learning-based</th>
<th>Hybrid method</th>
<th>Exemplar-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>☹️</td>
<td>☀️</td>
<td>☀️</td>
<td>☀️</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Global realism</th>
<th>Learning-based</th>
<th>Hybrid method</th>
<th>Exemplar-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>☀️</td>
<td>☠️</td>
<td>☠️</td>
<td>☠️</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Match Input</th>
<th>Learning-based</th>
<th>Hybrid method</th>
<th>Exemplar-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>☀️</td>
<td>☠️</td>
<td>☠️</td>
<td>☠️</td>
</tr>
</tbody>
</table>
Discussion

Summary

- Intuitive user inputs.
- Realistic outputs.
- Used by visual artists.

Challenges

- Fine-grained controls (texture, 3D, and lighting).
- High-resolution output (4K).
- Model efficiency on mobile devices.
- Video Control.
Input: **Text**  →  Output: **Photo**

**Text-to-Image Synthesis**

StackGAN, StackGAN++ [Zhang et al. 2016 and 2017], AttnGAN [Xu et al., 2018]
Input: **Class** → Output: **Photo**

Class-conditional GANs

- cGANs [Mirza and Osindero. 2014]
- SAGAN [Zhang et al., 2018]
- BigGAN [Brock et al., 2019]
- StyleGAN-XL [Sauer et al., 2022]
Multimodal Conditional Image Synthesis with Product-of-Experts GANs [Huang et al., 2021]
Multimodal Conditional Image Synthesis with Product-of-Experts GANs [Huang et al., 2021]
GauGAN2 Demo

http://gaugan.org/gaugan2/
Supervised Learning Approach

Edges2cats

Image colorization

Street view images

Natural outdoor images
Supervised Learning Approach

User Input → Learning algorithm + Labeled data → Visual Content

Expensive labor → Artistic authoring → Infeasible

$10-12 → ML Model

Artistic authoring is infeasible due to expensive labor.
Supervised

\[ x_i \quad y_i \]

Unsupervised

\[ X \quad Y \]
Unsupervised Learning of $p(y \mid x)$
Unsupervised Learning of $p(y \mid x)$

$\mathbb{E}_x \log(1 - D(G(x))) + \mathbb{E}_y \log D(y)$
Unsupervised Learning of $p(y \mid x)$

- artifacts
- ignore inputs

[Goodfellow et al. 2014]
Additional Constraint: Identity Mapping

Input image \( x \) \( \xrightarrow{G} \) Output image \( G(x) \) \( \xrightarrow{D} \) Real (1) or fake (0)?

Adversarial loss

\[
E_x \log(1 - D_Y(G(x))) + E_y \log D_Y(y)
\]

Self-Regularization loss

\[
E_x ||G(x) - x||_1
\]

SimGAN [Shrivastava et al., 2017]
Additional Constraint: Feature Loss

Adversarial loss

\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

Feature loss

\[ \mathbb{E}_x \| F(G(x)) - F(x) \| \]

Requires F to work across two domains

DTN [Taigman et al., 2017]
Additional Constraint: Cycle-Consistency

\[ X \quad \leftrightarrow \quad Y \]

CycleGAN [Zhu*, Park* et al., ICCV 2017]
Cycle-Consistent Adversarial Networks

Adversarial loss
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

Cycle-consistency loss
\[ \mathbb{E}_x \| F(G(x)) - x \|_1 \]

CycleGAN [Zhu*, Park* et al., ICCV 2017]
Cycle-Consistent Adversarial Networks

Adversarial loss: $D_Y(G(x))$

Cycle-consistency loss: $\|F(G(x)) - x\|_1$

Adversarial loss: $D_X(F(y))$

Cycle-consistency loss: $\|G(F(y)) - y\|_1$

CycleGAN [Zhu*, Park* et al., ICCV 2017]
Results
Horse → Zebra
Orange → Apple
Monet’s paintings → photographic style
Monet’s paintings → photographic style
Collection Style Transfer

Photograph ©Alexei Efros

Monet

Van Gogh

Cezanne

Ukiyo-e
Improving the Realism of CG Rendering

CG Game: Grand Theft Auto

Street view images in German cities

Data from [Richter et al., 2016], [Cordts et al, 2016]
Improving the Realism of CG Rendering

Output image with CG image street view style
Domain Adaptation with CycleGAN

CG images
Free segmentation labels

Data and labels from [Richter et al. 2016]
Domain Adaptation with CycleGAN

Train on CG data

Test on real images

Class-weighted Accuracy

47.4

Train on CG
Domain Adaptation with CycleGAN

Train on CycleGAN images

Test on real images

Train on CycleGAN images

Class-weighted Accuracy

47.4
63.8
67.4
72.4

CycleGAN
CycleGAN+
SOTA adaptation
SOTA adaptation

CycleGAN [Zhu*, Park* et al., ICCV 2017]
CycADA [Hoffman et al., ICML 2018]
Why CycleGAN works
Why CycleGAN works

Adversarial loss
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]
Why CycleGAN works

Adversarial loss
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

Cycle-consistency loss
\[ \mathbb{E}_x \|F(G(x)) - x\|_1 \]
Why CycleGAN works

**Adversarial loss**
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

**Cycle-consistency loss**
\[ \mathbb{E}_x \|F(G(x)) - x\|_1 \]
Why CycleGAN works

Adversarial loss

\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

Cycle-consistency loss

\[ \mathbb{E}_x \| F(G(x)) - x \|_1 \]

Auto-encoder w/ domain prior

Constraint:

\[ \mathbb{L}_1(G((G(x))_{\text{data}}(Y))) \]

[Hinton and Salakhutdinov. Science 2006]
Why CycleGAN works

Adversarial loss
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

Under-constrained problem

Cycle-consistency loss
\[ \mathbb{E}_x ||F(G(x)) - x||_1 \]

Prior of \( G \)

A strong regularizer

Assumption: simple invertible function

Probabilistic Interpretation: Upper bound of conditional entropy \( H(y|x) \)

[Li et al. 2017]
Why CycleGAN works

**Adversarial loss**
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

**Cycle-consistency loss**
\[ \mathbb{E}_x \| F(G(x)) - x \|_1 \]

Flip the image

\[ P \circ G \]

Invertible Perturbation

**Adversarial loss:** images are horizontally symmetric

**Cycle-consistency loss:**
\[ \| F \circ P^{-1}(P \circ G(x)) - x \| \]
Applications of CycleGAN
Image Dehazing

Foggy image

Clear image

[Engin et al. CVPRW 2018]
Computer Graphics

Computer Vision

Natural images

Other Image data

Natural images

Natural language (NLP)

Computer music

Audio processing

Cryptography

Remote Sensing

Medical Imaging

Robotics

Biology

Other Image data

Non-image data

Art

Piano→Harpsichord

[Huang et al. 2019]

[Bartha et al. 2018]
Artistic Applications

The Electronic Curator

“The Electronic Curator” © Eran Hadas and Eyal Gruss, israel
Latest from #CycleGAN

Input dog → Output cat → Input cat → Output dog

CycleGAN with modified architectures © itok_msi
Style and Content

Disentanglement
Style and Content Separation

A

Classification

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>A</td>
<td>E</td>
<td>D</td>
</tr>
</tbody>
</table>

Domain Adaptation

B

Extrapolation

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
</tbody>
</table>

Paired Image-to-Image Translation

C

Translation

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unpaired Image-to-Image Translation

Separating Style and Content
[Tenenbaum and Freeman 1996]

\[ y_{k}^{sc} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ijk}a_{i}^{c}b_{j}^{c}. \]
Style and Content

Adversarial loss
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

\[ p(x) \rightarrow p(y) \] change style

Cycle-consistency loss
\[ \mathbb{E}_x \| F(G(x)) - x \|_1 \]

Bidirectional: preserve content

Separating Style and Content
[Tenenbaum and Freeman 1996]
Style and Content

Adversarial loss (change style)
\[ \mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y) \]

L1 loss (preserve content in pixel space)
\[ \mathbb{E}_x ||G(x) - x||_1 \]

SimGAN [Shrivastava et al., 2017]
**Style and Content**

Adversarial loss (change style)

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Feature loss (Preserve content in feature space)

$$\mathbb{E}_x \| F(G(x)) - F(x) \|$$

DTN [Taigman et al., 2017]
CycleGAN and UNIT

- **CycleGAN (cycle consistency)**

  ![CycleGAN Diagram](image)

- **UNIT (shared latent space) [Liu et al. 2017]**

  ![UNIT Diagram](image)
Disentangling the Latent Space

• UNIT
  – A single **shared, domain-invariant** latent space $\mathcal{Z}$
Disentangling the Latent Space

- Multimodal UNIT (MUNIT)
  - A content space $C$ that is shared, domain-invariant
  - Two style spaces $S_1, S_2$ that are unshared, domain-specific
Unimodality
Towards Multimodality
Training

- Notations:
  - $x$: images
  - $c$: content
  - $s$: style

- Loss:
  - Bidirectional reconstruction loss
    - Image reconstruction loss
    - Latent reconstruction loss
  - GAN loss

With cross-domain reconstruction
Bidirectional Reconstruction Loss: Image Reconstruction

Notations:
– \( x \): images
– \( c \): content
– \( s \): style

\[
\begin{align*}
\hat{x}_1 & \rightarrow s_1 \rightarrow c_1 \rightarrow x_1 \\
\hat{x}_2 & \rightarrow s_2 \rightarrow c_2 \rightarrow x_2 \\
\end{align*}
\]

\[ \mathcal{L}_1 \text{ loss} \]
Notations:
- $x$: images
- $c$: content
- $s$: style
GAN Loss

Notations:
- $x$: images
- $c$: content
- $s$: style
AdaIN in a Generative Network

\[ \text{AdaIN}(c, s) = \gamma \left( \frac{c - \mu(c)}{\sigma(c)} \right) + \beta \]
AdaIN in a Generative Network

$$\text{AdaIN}(c, s) = \gamma \left( \frac{c - \mu(c)}{\sigma(c)} \right) + \beta$$

AdaIN in a generative network
Sketches <-> Photo

Input

Outputs
Cats ↔ Dogs

Input

Output
Example-guided Translation
Example-guided Translation

<table>
<thead>
<tr>
<th>Content</th>
<th>Style</th>
<th>Ours</th>
<th>Gatys et al.</th>
<th>AdaIN</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Thank You!

16-726, Spring 2022

https://learning-image-synthesis.github.io/sp22/