Image to Image Translation using Generative Adversarial Networks

Sagie Benaim
Tel Aviv University
Image to Image Translation
Semantic label → Image

Day → Night

Winter → Summer

Artistic video gaming

Drawing → Image

Many other applications
Adversarial Nets Framework

\[ D(x) \text{ tries to be near } 1 \]

Differentiable function \( D \)

\[ x \text{ sampled from data} \]

\[ D \text{ tries to make } D(G(z)) \text{ near } 0, \]
\[ G \text{ tries to make } D(G(z)) \text{ near } 1 \]

Differentiable function \( G \)

Input noise \( z \)

\( x \) sampled from model

(Goodfellow et al., 2014)
Generative Modeling:
Sample Generation

Training Data
(CelebA)

Sample Generator
(Karras et al, 2017)
Early Approaches

Object labeling
[Long et al. 2015]

Edge Detection
[Xie et al. 2015]

Season change
[Laffont et al. 2014]

Artistic style transfer
[Gatys et al. 2016]
<table>
<thead>
<tr>
<th></th>
<th>Supervised</th>
<th>Unsupervised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unimodal</td>
<td>Pix2pix, CRN, SRGAN</td>
<td>DistanceGAN, CycleGAN, DiscoGAN, DualGAN, UNIT, DTN, StarGAN, OST</td>
</tr>
<tr>
<td>Multimodal</td>
<td>pix2pixHD, BicycleGAN</td>
<td>MUNIT, Augmented CycleGAN</td>
</tr>
</tbody>
</table>
Fully Supervised: pix2pix

Conditional GAN

\[ G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G). \]

[Isola et al., CVPR 2017]
Partially Supervised Alignment

• “Unsupervised Cross-Domain Image Generation” Taigman et al.
Unsupervised: Circular GANs

**DiscoGAN**: “Learning to Discover Cross-Domain Relations with Generative Adversarial Networks”. Kim et al. ICML’17.


Cycle Consistency Loss

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

\[ \| G(F(x)) - y \|_1 \]

See similar formulations [Yi et al. 2017], [Kim et al. 2017] [Zhu et al., ICCV 2017]
Style and Content Separation

**Paired Separation**

**Unpaired Separation**

**Adversarial Loss:** change the style

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

**Cycle Consistency Loss:** preserve the content

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

Two empirical assumptions:
- content is easy to keep.
- style is easy to change.

Separating Style and Content with Bilinear Models
[Tenenbaum and Freeman 2000’]
Collection Style Transfer

Photograph @ Alexei Efros

Monet

Van Gogh

Cezanne

Ukiyo-e
DistanceGAN

- A pair of images of a given distance are mapped to a pair of outputs with a similar distance.
- $|x_i - x_j|_1$ and $|G(x_i) - G(x_j)|_1$ are highly correlated.

$|x_1 - x_2|_1 \sim |G(x_1) - G(x_2)|_1$

Benaim et al., NIPS 2017
Motivating distance correlations

Analysis of CycleGAN’s horse to zebra results

Benaim et al., NIPS 2017
Mode Collapse

- GAN:

- Cycle:

Benaim et al., NIPS 2017
More than 2 domains

Choi et al., CVPR 2018
More than 2 domains
Modeling multiple possible outputs
**BiCycleGAN** [Zhu et al., NIPS 2017]
(c.f. InfoGAN [Chen et al. 2016])

**MAD-GAN** [Ghosh et al., CVPR 2018]
UNIT: unimodal

MUNIT: multimodal

Huang et al., ECCV 2018
Architecture

Huang et al., ECCV 2018
(a) Within-domain reconstruction

(b) Cross-domain translation
Sketch to Image Translation

(a) edges ↔ shoes

(b) edges ↔ handbags

Huang et al., ECCV 2018
Animal Image Translation

(a) house cats → big cats
(b) big cats → house cats
(c) house cats → dogs
(d) dogs → house cats
(e) big cats → dogs
(f) dogs → big cats

Huang et al., ECCV 2018
One Shot?

• Not only are we unsupervised, but we have only a single sample in the input domain!

What is Semantic?

- Potentially Infinitely many solutions preserving distance correlations

Figure 1: An illustrative example where the two domains are line segments in $\mathbb{R}^2$. There are infinitely many mappings that preserve the uniform distribution on the two segments. However, only two stand out as "semantic". These are exactly the two mappings that can be captured by a neural network with only two hidden neurons and Leaky ReLU activations, i.e., by a function $h(x) = \sigma_\alpha(Wx + b)$, for a weight matrix $W$ and the bias vector $b$.

“The role of Minimal Complexity Functions in Unsupervised Learning of Semantic Mappings”, Galanti, et al. ICLR 2018
Practical Considerations
GANs Training: Stability and Mode Collapse

• Practical Tips: https://github.com/soumith/gan hacks
• A lot of research on more stable GAN with less mode collapse:
  • Stability: WGAN/ Improved WGAN
  • Mode Collapse: Spectral Normalization, SAGAN, Many More
GAN Architecture

• DiscoGAN based (64 pixels):
  • Generator: Encoder-Decoder, Based on DCGAN
  • Discriminator: Simple Decoder

• CycleGAN based (128-256 pixels):
  • Based on “Perceptual losses for real-time style transfer and super-resolution” Johnson et al.
  • Generator: Use of additional Residual blocks
  • Discriminator: Use of 70*70 Patch-GAN
Patch GAN: Choosing the Discriminator

Rather than penalizing if output image looks fake, penalize if each overlapping patch in output looks fake.

[Li & Wand 2016]
[Shrivastava et al. 2017]
[Isola et al. 2017]
Labels → Facades

Data from [Tylecek, 2013]
Labels → Facades

Input

16x16 Discriminator

Data from [Tylecek, 2013]
Labels → Facades
Labels → Facades

Input

Full image Discriminator

Data from [Tylecek, 2013]
Replace L1 Loss with Perceptual Loss

\[ L(\hat{y}, y) = \| \phi(\hat{y}) - \phi(y) \|_2 \]

[Johnson, Alahi, Li, ECCV 2016]
[Chen & Koltun ICCV 2017]
[Zhang et al. CVPR 2018]
[Mostajabi, Maire, Shakhnarovich, arXiv 2018]
How can we measure the error in an unsupervised translation?

- Maximize the distance between two GANs
- $G_2(a)$ maximizes its distance to $G_1(a)$ ($G_1$ trains as usual)
- This distance is a proxy to the distance of $G_1(a)$ to the ground truth

Applications Beyond Computer Vision

- Many other Vision Applications: Photo Enhancement, Image Dehazing
- Medical Imaging and Biology [Wolterink et al., 2017]
- Voice conversion [Fang et al., 2018, Kaneko et al., 2017]
- Cryptography [CipherGAN: Gomez et al., ICLR 2018]
- Robotics
- NLP: Text style transfer.
- ...
Thank You! Questions?