Application of Generative Models: Image-to-Image Translation

Hao Dong

Peking University
Why we learn im2im?

• The most classical generative model application..
• The state-of-the-art methods are all based on GAN ...
• Understand GAN and the history better ...
Application of Generative Models: Image-to-Image Translation

- Problem Definition
- Image Inpainting / Reconstruction / Super Resolution
- Pix2Pix: paired data
- Discussion: ideal im2im
- UNIT and CycleGAN: unpaired data
- BiCycleGAN: multi-modality
- MUNIT and Augmented CycleGAN: unpaired data + multi-modality
- DRIT: disentangle domain-specific features
- Attention CycleGAN: maintain background
- StarGAN: label condition
- Breaking the Cycle
- GAN-CLS and SisGAN: text condition
• Problem Definition
  • Image Inpainting / Reconstruction / Super Resolution
  • Pix2Pix: paired data
  • Discussion: ideal im2im
  • UNIT and CycleGAN: unpaired data
  • BiCycleGAN: multi-modality
  • MUNIT and Augmented CycleGAN: unpaired data + multi-modality
  • DRIT: disentangle domain-specific features
  • Attention CycleGAN: maintain background
  • StarGAN: label condition
  • Breaking the Cycle
  • GAN-CLS and SisGAN: text condition
Problem Definition

• Supervised/Paired image-to-image translation

• Unsupervised/Unpaired image-to-image translation
• Problem Definition
• **Image Inpainting / Reconstruction / Super Resolution**
  • Pix2Pix: paired data
  • Discussion: ideal im2im
  • UNIT and CycleGAN: unpaired data
  • BiCycleGAN: multi-modality
  • MUNIT and Augmented CycleGAN: unpaired data + multi-modality
  • DRIT: disentangle domain-specific features
  • Attention CycleGAN: maintain background
  • StarGAN: label condition
  • Breaking the Cycle
  • GAN-CLS and SisGAN: text condition
Image Inpainting / Reconstruction / Super Resolution

• Utilising Feature Information for Medical Image Reconstruction

Encoder is a part of the generator (fully conv nets)

Image Inpainting / Reconstruction / Super Resolution

- Supervised image super resolution

Better feature reconstruction

SRGAN

Pre-trained VGG

Image Inpainting / Reconstruction / Super Resolution

- Supervised image super resolution

UV-GAN: Adversarial Facial UV Map Completion for Pose-invariant Face Recognition.

J. Deng, S. Cheng et al. CVPR. 2018.
• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• **Pix2Pix: paired data**
  • Discussion: ideal im2im
  • UNIT and CycleGAN: unpaired data
  • BiCycleGAN: multi-modality
  • MUNIT and Augmented CycleGAN: unpaired data + multi-modality
  • DRIT: disentangle domain-specific features
  • Attention CycleGAN: maintain background
  • StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

- Beyond MLE: Adversarial Learning

  - Question 1: What color are they?
    Red? Blue? Yellow? ... obviously there are more than one solution

  - Question 2: What if I train a neural net: input $x_{grey}$ output $x_{rgb}$ with MLE?

    \[ x_{grey} \rightarrow \text{Neural} \rightarrow \hat{x}_{rgb} \rightarrow L_2 \rightarrow x_{rgb} \]

    What is the problem??

Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

- Beyond MLE: Adversarial Learning

Different colors will have conflicts, (some want red, some want blue, ...) resulting “grey” outputs


Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

- Beyond MLE: Adversarial Learning

Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

Pix2Pix: paired data

- Pix2Pix: Supervised Image-to-Image Translation

• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• **Discussion: ideal im2im**
  • UNIT and CycleGAN: unpaired data
  • BiCycleGAN: multi-modality
  • MUNIT and Augmented CycleGAN: unpaired data + multi-modality
  • DRIT: disentangle domain-specific features
  • Attention CycleGAN: maintain background
  • StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
Discussion: ideal im2im

• What should the ideal image-to-image translation to be?
  • Unpaired data
  • Maintain background
  • Multi-modality
  • Disentanglement
  • Multi-domain
  • Conditional translation
• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• **UNIT and CycleGAN: unpaired data**
  • BiCycleGAN: multi-modality
  • MUNIT and Augmented CycleGAN: unpaired data + multi-modality
  • DRIT: disentangle domain-specific features
  • Attention CycleGAN: maintain background
  • StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
GAN with Encoder -- Unsupervised Image-to-Image Translation


UNIT and CycleGAN: unpaired data

- CycleGAN: Unpaired Image-to-Image Translation

UNIT and CycleGAN: unpaired data

- CycleGAN: Unpaired Image-to-Image Translation

UNIT and CycleGAN: unpaired data

- Discussion: are they unsupervised learning?
  - NO, two image domains == binary labels.

- Why the background / shape can be maintained?
  - Fully convolutional networks → inductive bias
  - Cycle-consistency loss

- Questions?
• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
  • **BiCycleGAN: multi-modality**
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background
• StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
BiCycleGAN: multi-modality

- Support diverse (multi-modal) outputs but still need paired data
BiCycleGAN: multi-modality

- Cycle on latent noises + Cycle on translated images
BiCycleGAN: multi-modality

• Result
• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background
• StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Goal: unpaired + multi-modal results

MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Latent reconstruction + Adversarial learning

MUNIT and Augmented CycleGAN: unpaired + multi-modal

- Comparison against previous methods

• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background
• StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
DRIT: disentangle domain-specific features

- Goal: Multi-modal results + Disentanglement
DRIT: disentangle domain-specific features

- Network bottleneck + Adversarial learning

DRIT: disentangle domain-specific features

- Additional losses for better disentanglement
DRIT: disentangle domain-specific features

- Results

(a) Inter-domain attribute transfer
(b) Intra-domain attribute transfer

• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• **Attention CycleGAN**: maintain background
• StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
Attention CycleGAN: maintain background

- Learn the segmentation via synthesis

Attention CycleGAN: maintain backgrounds

• Maintain backgrounds better

Attention CycleGAN: maintain backgrounds

- Learn the segmentation without segmentation labels

• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background

• **StarGAN: label condition**
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
StarGAN: label condition

- Limitation of CycleGAN

Translations between $N$ domains require $N(N-1)$ models

StarGAN: One model to rule them all !!
StarGAN: label condition

- Add a class condition into the generator and the output of the discriminator.
• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background
• StarGAN: label condition

• **Breaking the Cycle**
• GAN-CLS and SisGAN: text condition
Breaking the Cycle

- Limitation of the cycle-consistency loss

|-------|-----------|--------------|--------|--------------|----------|-------------|

Cycle loss enforces the constraint that translating an image to the target domain and back, should obtain the original image.
Breaking the Cycle

- Limitation of cycle-consistency loss

Image from: https://github.com/junyanz/CycleGAN
Breaking the Cycle

• Colleagues are all you need

Each member of the council is a triplet (Red indicates one council), whose components are one generator and two discriminators.

The task of discriminator $D_i$ is to distinguish between the generator’s $G_i$ output and real examples.

The goal of discriminator $\widehat{D}_i$ is to distinguish between images produced by $G_i$ and images produced by the other generators in the council. This discriminator is the core of the model and this is what differentiates the model from the classical GAN model. It enforces the generator to converge to images that could be acknowledged by all council.

Breaking the cycle—Colleagues are all you need. Nizan, Ori. Tal, Ayellet. arXiv 2019.11
Breaking the Cycle

- Colleagues are all you need

Breaking the cycle—Colleagues are all you need. Nizan, Ori. Tal, Ayellet. arXiv 2019.11
Breaking the Cycle

- Discussion: Why it works?

- It is not only for im2im, other distribution transformations may also benefit from this approach

- Better Methods:
  - ACL-GAN
  - XDCycleGAN

Breaking the cycle—Colleagues are all you need. Nizan, Ori. Tal, Ayellet. arXiv 2019.11
• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background
• StarGAN: label condition
• Breaking the Cycle

• GAN-CLS and SisGAN: text condition
GAN-CLS and SisGAN

• Text-to-image synthesis

  - this small bird has a pink breast and crown, and black primaries and secondaries.
  - this magnificent fellow is almost all black with a red crest, and white cheek patch.
  - the flower has petals that are bright pinkish purple with white stigma
  - this white and yellow flower have thin white petals and a round yellow stamen

• Classic multi-modal problem

  \[ P(t, z) \]

GAN-CLS and SisGAN

- Text-to-image synthesis
GAN-CLS and SisGAN

• Text-to-image synthesis

Learn to fool discriminator

1. Learn to classify matching image and text as real sample
2. Learn to classify mismatched image and text as fake sample
3. Learn to classify image from generator as fake sample
GAN-CLS and SisGAN

• Text-to-image synthesis + High resolution image

GAN-CLS and SisGAN

- Text-to-image synthesis + High resolution image

GAN-CLS and SisGAN

- Semantic image synthesis: image manipulation with natural language

GAN-CLS and SisGAN

- Semantic image synthesis: image manipulation with natural language
GAN-CLS and SisGAN

- Semantic image synthesis: image manipulation with natural language
  
  1. Learn to classify matching image and text pairs as real samples
  2. Learn to classify mismatched image and text pairs as fake samples
  3. Learn to classify samples from generator as fake samples

Learn to fool discriminator when inputting image with semantically relevant text

\[ \mathcal{L}_D = \mathbb{E}_{(x,t) \sim p_{\text{data}}} \log D(x, \varphi(t)) + \mathbb{E}_{(x, \hat{t}) \sim p_{\text{data}}} \log (1 - D(x, \varphi(\hat{t}))) + \mathbb{E}_{(x, \tilde{t}) \sim p_{\text{data}}} \log (1 - D(G(x, \varphi(\tilde{t})), \varphi(\tilde{t}))) \]

\[ \mathcal{L}_G = \mathbb{E}_{(x, \tilde{t}) \sim p_{\text{data}}} \log (D(G(x, \varphi(\tilde{t})), \varphi(\tilde{t}))) \]

GAN-CLS and SisGAN

- Semantic image synthesis: Learn the location information via synthesis

\[ \tilde{x} = G(x, \varphi(\tilde{x})) \]

A black bird with a red head.

A bird with red head and breast.

64x64 with VGG 256x256 Difference

A black bird with a red head.

64x64 with VGG 256x256 Difference

The petals are purple with no visible stamens.

The petals are white and the stamens are yellow.

Summary

• Problem Definition
• Image Inpainting / Reconstruction / Super Resolution
• Pix2Pix: paired data
• Discussion: ideal im2im
• UNIT and CycleGAN: unpaired data
• BiCycleGAN: multi-modality
• MUNIT and Augmented CycleGAN: unpaired data + multi-modality
• DRIT: disentangle domain-specific features
• Attention CycleGAN: maintain background
• StarGAN: label condition
• Breaking the Cycle
• GAN-CLS and SisGAN: text condition
Discussion: ideal im2im

• What should the ideal image-to-image translation to be?
  
  • Unpaired data
  
  • Maintain background
  
  • Multi-modality
  
  • Disentanglement
  
  • Multi-domain
  
  • Conditional translation
Thanks