Conditional Generation
by GAN

Hung-yi Lee
Text-to-Image

- Traditional supervised approach

\[ c^1: \text{a dog is running} \] → NN → Image

Text: “train”

Target of NN output

A blurry image!

a dog is running

a bird is flying

as close as possible
Conditional GAN

\[ x = G(c, z) \]

Normal distribution \( z \)

c: train

\( x \) is real image or not

Generator will learn to generate realistic images ....

But completely ignore the input conditions.

Real images:

Generated images:
Conditional GAN

\[ x = G(c, z) \]

True text-image pairs:
- (train, ) 1
- (cat, ) 0
- (train, Image) 0

[Scott Reed, et al, ICML, 2016]
Conditional GAN - Discriminator

object x $\rightarrow$ Network $\rightarrow$ score
condition c $\rightarrow$ Network $\rightarrow$ x is realistic or not + c and x are matched or not

(almost every paper)

object x $\rightarrow$ Network $\rightarrow$ x is realistic or not
condition c $\rightarrow$ Network $\rightarrow$ c and x are matched or not

[Augustus Odena et al., ICML, 2017]
[Takeru Miyato, et al., ICLR, 2018]
[Han Zhang, et al., arXiv, 2017]
Conditional GAN

**paired data**
- blue eyes
- red hair
- short hair

Collecting anime faces and the description of its characteristics

red hair, green eyes

blue hair, red eyes

The images are generated by Yen-Hao Chen, Po-Chun Chien, Jun-Chen Xie, Tsung-Han Wu.
Stack GAN

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas, “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks”, ICCV, 2017

This bird is grey with white on its chest and has a very short beak.
Image-to-image

\[ x = G(c, z) \]
Image-to-image

- Traditional supervised approach

Testing:

It is blurry because it is the average of several images.
Image-to-image

• Experimental results

Testing:

input close GAN GAN + close
Patch GAN

Speech Enhancement

• Typical deep learning approach

Using CNN

Output

Enhancing
Speech Enhancement

- Conditional GAN

![Diagram showing the process of speech enhancement using a Conditional GAN. The input is noisy speech, which is processed by the generator (G) to produce a noisy output. The discriminator (D) then evaluates the output to determine if it is a real or fake pair, helping to improve the quality of the output.]
Video Generation

Generator

Discriminator

Last frame is real or generated

Minimize distance

Discriminator thinks it is real

Target
https://github.com/dyelax/Adversarial_Video_Generation