Generative Adversarial Networks for CV
A Survey and Overview

How GANs are applied to Computer Vision problems

Poornapragna Vadiraj

CMPE 258 - Deep Learning
Professor Vijay Eranti

https://www.youtube.com/watch?v=Z6rxFNMGdn0

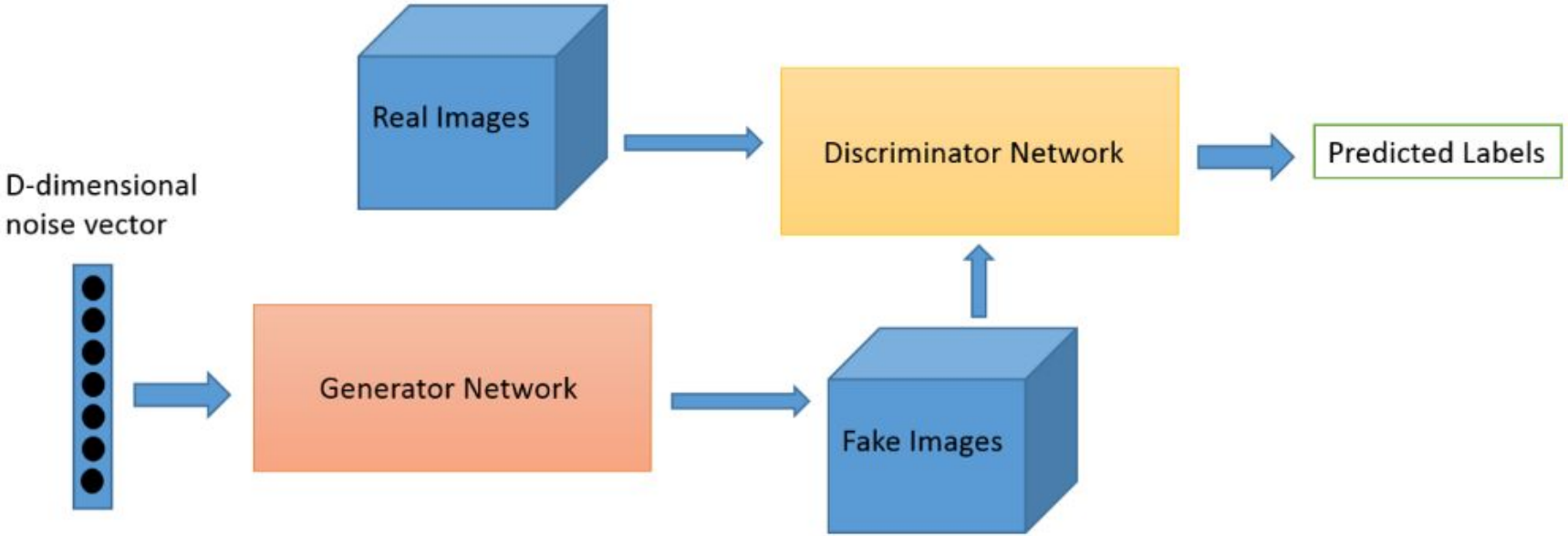
Intro

- GANs are a deep learning-based generative model that is used for unsupervised learning.
- It was first described in a paper in 2014 by Ian Goodfellow, and a standardized and much more stable model was proposed by Alec Radford et. al in 2016 called DCGAN (Deep Convolutional Generative Adversarial Network).
- A way to learn in-depth representations without extensive use of annotated training data.

Paper (contd)

- GANs are known to be hard to train, particularly when trying to generate high resolution images. This article provides a detailed overview of the cutting-edge GAN approaches in four relevant image generation fields, including Text-to-Image synthesis, Image-to-Image translation, Face Aging and 3d Image Synthesis.
- One of the interesting and difficult challenges of Computer Vision, which has many uses including picture processing and computer-aided content development, is to synthesize high-quality images from text definition.
- The job of text for image generation usually involves translating text directly into prediction of image pixels in the form of single sentence descriptions.

GAN Architecture



Source: https://www.oreilly.com/ideas/deep-convolutional-generative-adversarial-networks-with-tensorflow

GAN Architecture

A simple explanation of why we call it Adversarial is because of the fact that we are essentially pitting these two networks against each other, forcing them to improve their performance on what they're supposed to do.

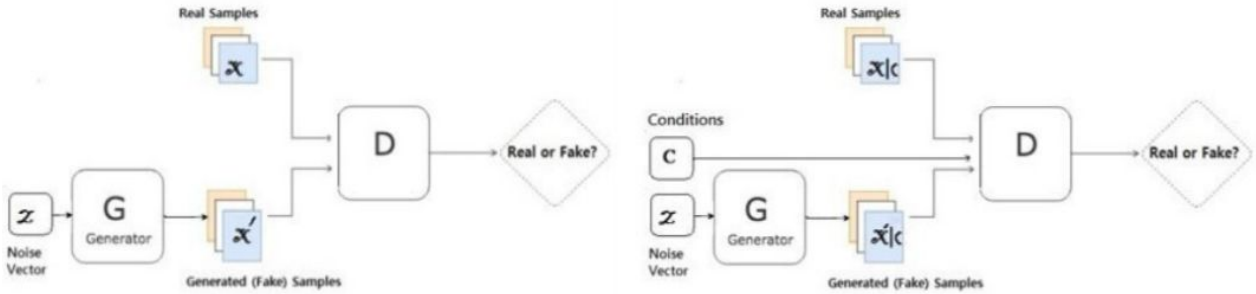
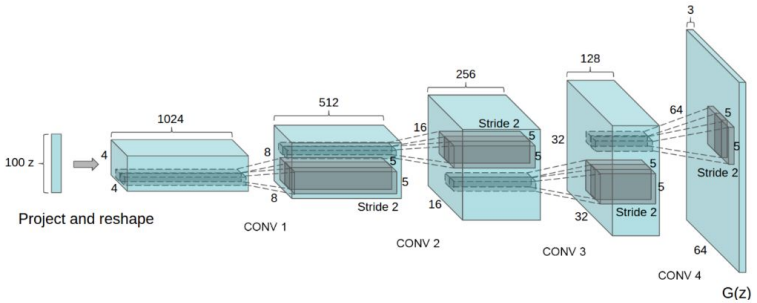


Figure 1. Structure of GANs (left) and cGANs (right)

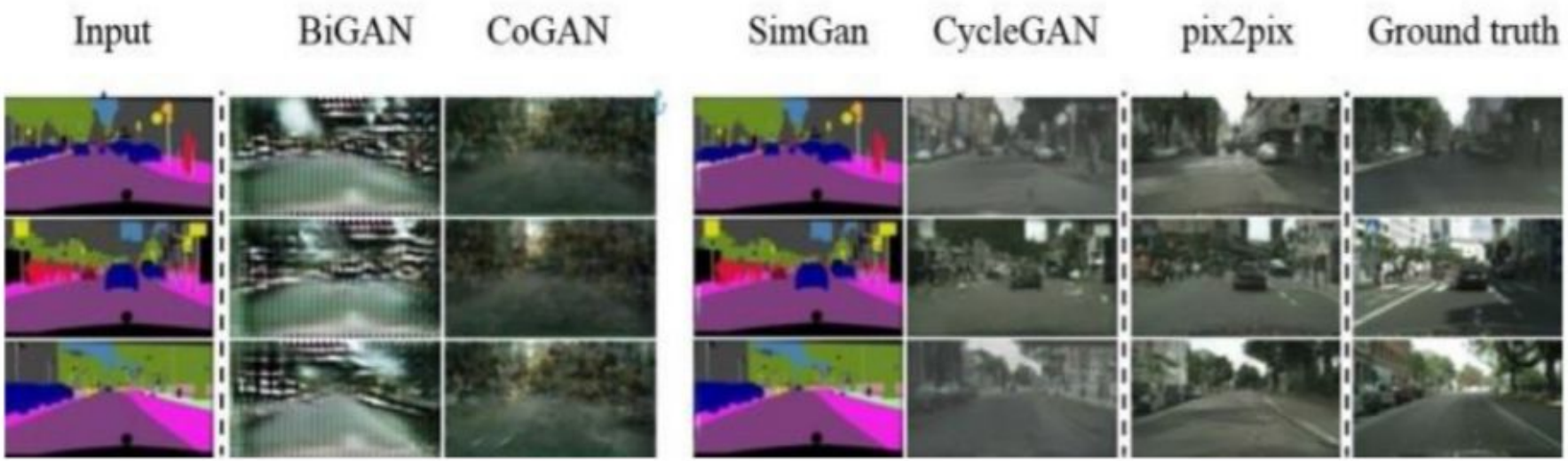
DCGAN

- Replace all max pooling with convolutional stride
- Use transposed convolution for upsampling.
- Eliminate fully connected layers.
- Use Batch normalization except the output layer for the generator and the input layer of the discriminator.
- Use ReLU in the generator except for the output which uses tanh.
- Use LeakyReLU in the discriminator.



Source

Image-to-Image Translation



Pix-2-Pix

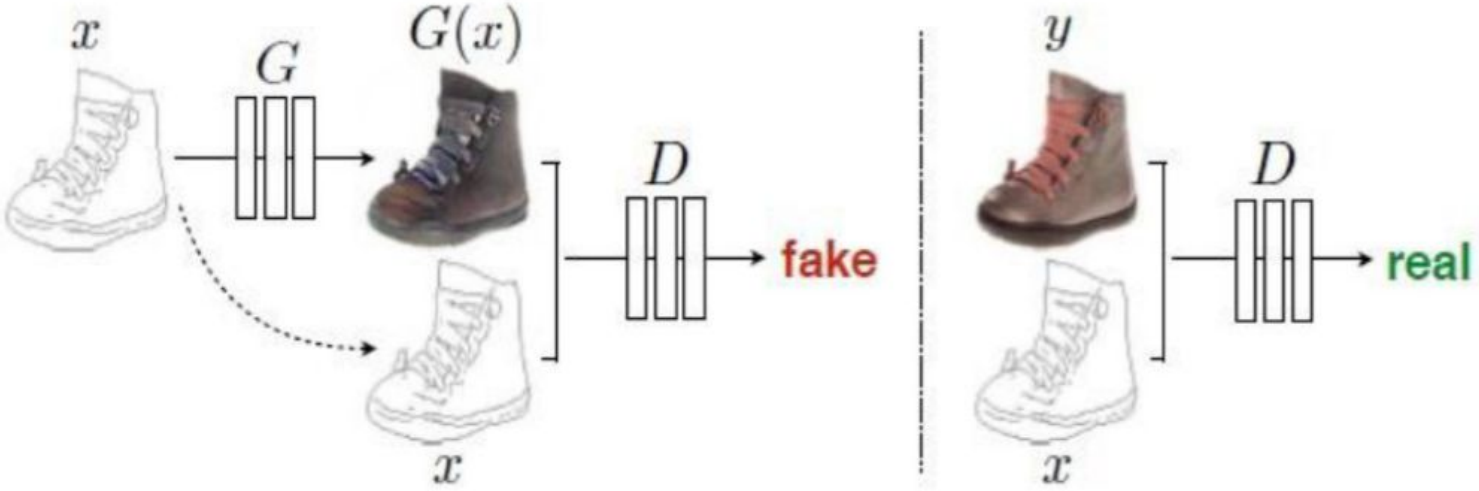


Figure 3. Training a cGANs to map edges to the photo. (Here, input map is map edges) [22]

Model	Per-pixel Accuracy	Per-class Accuracy	Class IOU
CoGAN [24]	0.45	image	0.08
BiGAN/ALI [25, 26]	0.41	image	0.07
SimGAN [9]	0.47	image	0.07
CycleGAN [23]	0.58	image	0.16
Pix2Pix [22]	0.85	image	0.32

Text to Image Models

- Synthesizing high-quality images from text descriptions, is one of the exciting and challenging problems in Computer Vision which has many applications, including photo editing and computer-aided content creation.
- The task of text to image generation usually means translating text in the form of single-sentence descriptions directly into prediction of image pixels. This can be done by different approaches.

Text to Image Models

Model	Input	Output	Characteristics	Resolution
GAN-INT-CLS [5]	text	image	---------	64 × 64
GAWWM [7]	text + location	image	location-controllable	128 × 128
StackGAN [13]	text	image	high quality	256 × 256
TAC-GAN [17]	text	image	diversity	128 × 128
ChatPainter [9]	text + dialogue	image	high inception score	256 × 256
HDGAN [10]	text	image	high quality and resolution	512 × 512
AttnGAN [20]	text	image	high quality and the highest inception score	256 × 256
Hong et al. [14]	text	image	Second highest inception score and complicated images	128 × 128

Measuring a GAN’s performance

- The Inception Score, or IS for short, is an empirical metric for assessing the quality of images produced by generative adversarial network models, specifically the synthetic image production.
- The inception score was proposed by Tim Salimans, et al. in their 2016 paper titled “Improved Techniques for Training GANs.”
- The authors in the paper use a crowd-sourcing platform (Amazon Mechanical Turk) to evaluate large numbers of images generated by GAN. They developed the inception score as an attempt to remove the subjective assessment of images by humans.
- The investigators discover their results were well associated with the quantitative evaluation.

TABLE 2: Inception scores of different models.

Model	**Inception Score**
GAN-INT-CLS [5]	7.88 ± 0.07
StackGAN [13]	8.45 ± 0.03
Hong et al. [14]	11.46 + 0.09
ChatPainter (non–current) [9]	9.43 ± 0.04
ChatPainter (recurrent) [9]	9.74 ± 0.02
AttnGAN [20]	25.89 ± 0.47

FACE AGING

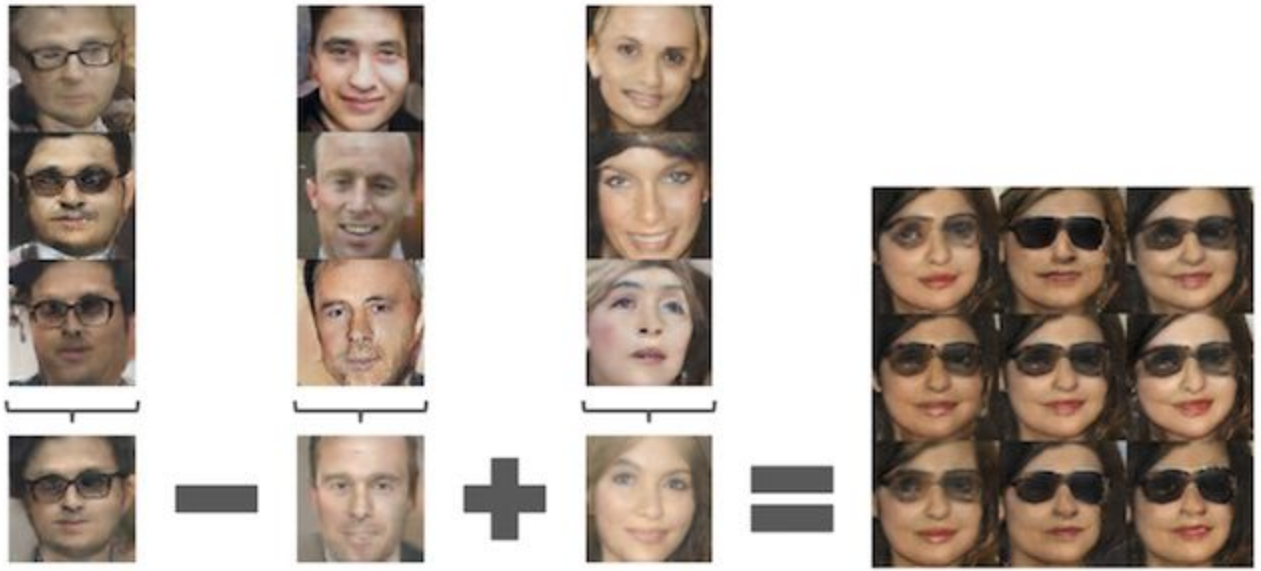
- Face aging, age synthesis or age progression (refers to future looks) and regression (refers to previous looks), are different names for a simple concept that is rendering of facial images with different ages with the same facial recognition features.
- It has many applications such as finding lost children and wanted person or entertainment.
- There have been two main traditional faceaging methods:
  - prototyping
  - modeling

Completely connected GAN (FCGAN): The initial GAN uses both generator and discriminator entirely linked neural networks. Works for basic databases, such as MNIST, CIFAR-10 etc. Does not show strong generalisation for different types of pictures.

Laplacian Pyramid of Adversarial Networks (LAPGAN): For the production of higher resolution images from lower GAN data. Uses a cascade of CNNs with a system of Laplacian pyramids to generate pictures of high quality.

Deep Convolutional GAN (DCGAN): Has shown strong CNN visualization efficiency.

SOURCE: ZHENGWEI WANG, et al. Timeline of architecture-variant GANs. Complexity in blue stream refers to the size of the architecture and computational cost such as batch size. Mechanisms refer to the number of types of models used in the architecture (e.g., BEGAN uses an autoencoder architecture for its discriminator while a deconvolutional neural network is used for the generator. In this case, two mechanisms are used).



Source: Machine Learning Mastery

Source: ML Mastery

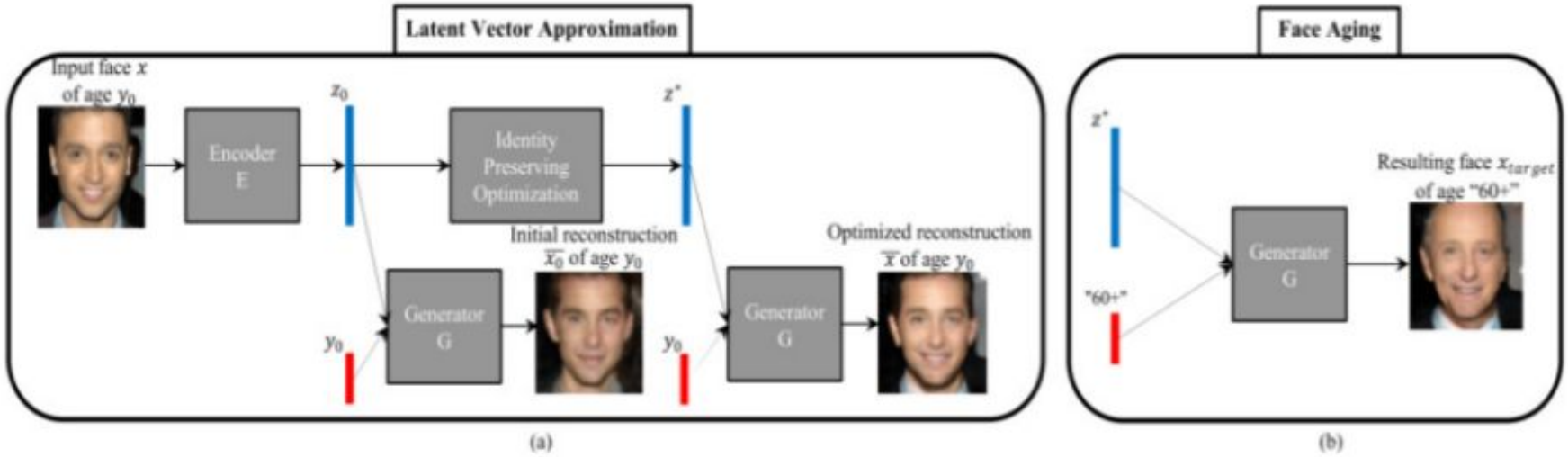
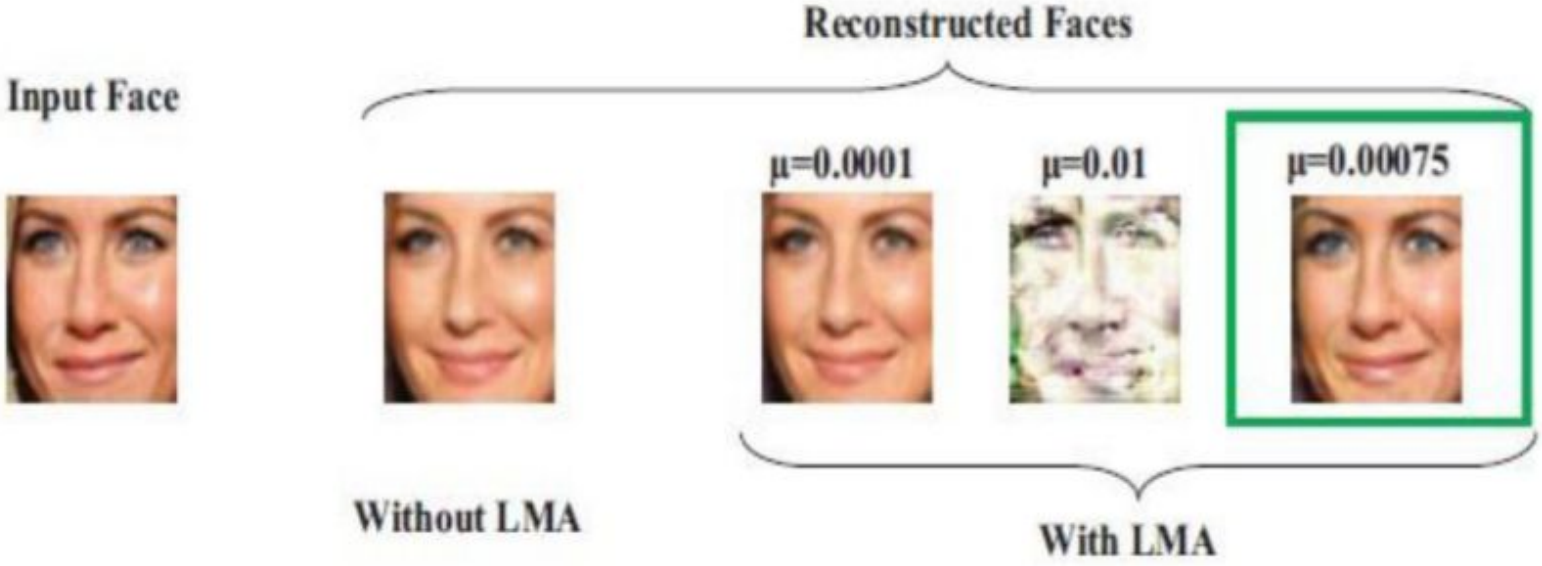


Figure 5. (a): Approximation of the latent vector to reconstruct the input image, (b): Switching the age condition at the input of the generator to perform face aging [32]


Reconstructed Faces
Input Face
μ=0.0001
μ=0.01
μ=0.00075
Without LMA
With LMA

3D IMAGE SYNTHESIS

3D object reconstruction of 2D images has always been a challenging task that try to define any object's 3D profile, as well as the 3D coordinate of every pixel.

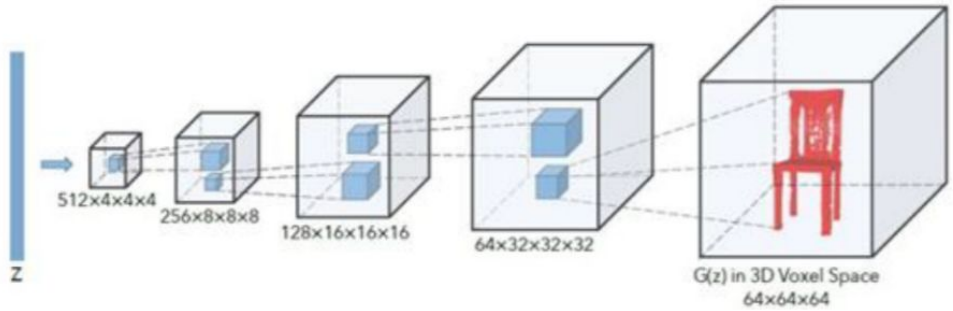


Figure 7. 3DGAN generator. The Discriminator mostly mirrors the generator

3D Encoder-Decoder GAN(3D-ED-GAN) with a Long term Recurrent Convolutional Network (LRCN)

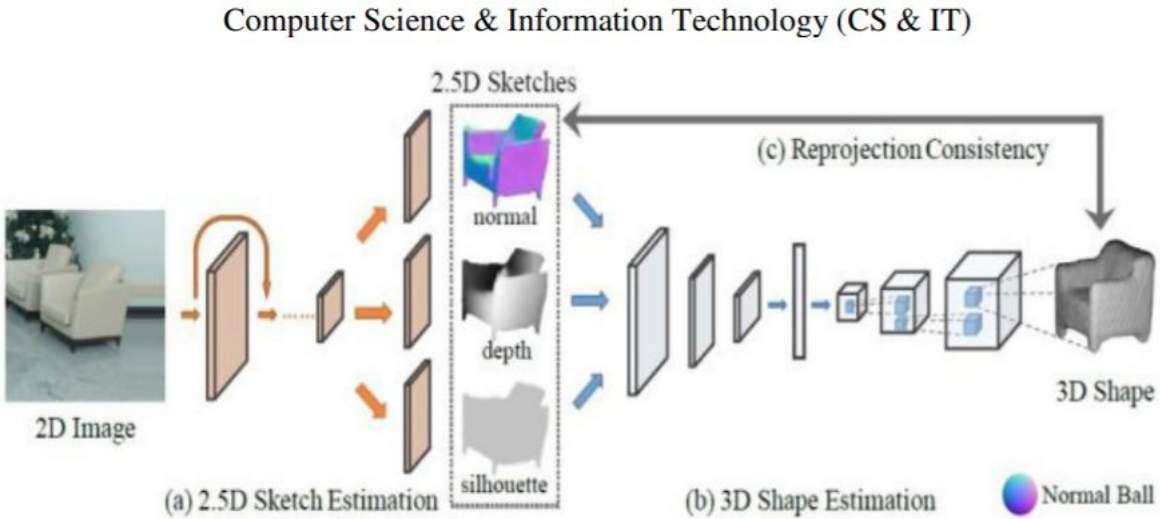


Figure 9. 3D construction of chairs on IKEA dataset. From left to right: input, ground truth, 3D estimation by 3DGAN and two view of MarrNet. [45]

Model	ModelNet40	ModelNet10
3DGAN [44]	83.3%	91.0%
3D-ED-GAN [46]	87.3%	92.6%
VoxNet [50]	92.0%	83.0%
DeepPano [51]	88.66%	82.54%
VRN [52]	91.0%	93.6%

Demos

http://ganpaint.io/demo/?project=church

https://www.nvidia.com/en-us/research/ai-playground/

https://reiinakano.com/gan-playground/

https://www.kaggle.com/summitkwan/tl-gan-demo

https://junyanz.github.io/CycleGAN/

https://thispersondoesnotexist.com/

Thank you!