Photo by Swapnll Dwivedi

Larger resolution with better quality in the first 2 years.

GAN — Some cool applications of GANs.

Jonathan Hui Follow
Jun 22 · 8 min read

We make impressive progress in the first few years of GAN developments. No more stamp-size facial pictures like those in horror movies. In 2017, GAN produced 1024 × 1024 images that can fool a talent scout. In coming years, we will probably see high-quality videos generated from GANs. The commercial applications will come! As part of the GAN series, we look into some cool applications and hope that they become the inspiration for your GAN application.

Create Anime characters

Game development and animation production are expensive and hire many production artists for relatively routine tasks. GAN can auto-generate and colorize Anime characters.

Figure 7: Generated samples

Towards the automatic Anime characters creation with Generative Adversarial Networks

The generator and the discriminator composes of many layers of convolutional layers, batch normalization and ReLU with skip connections.

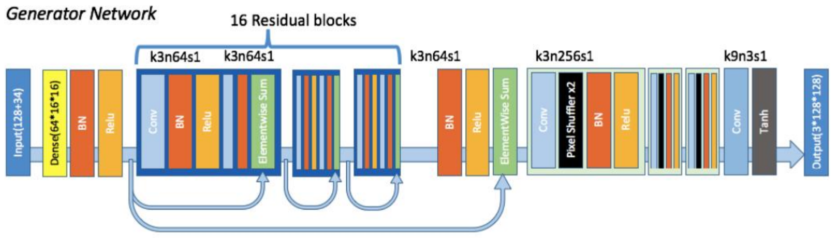


Figure 3: Generator Architecture

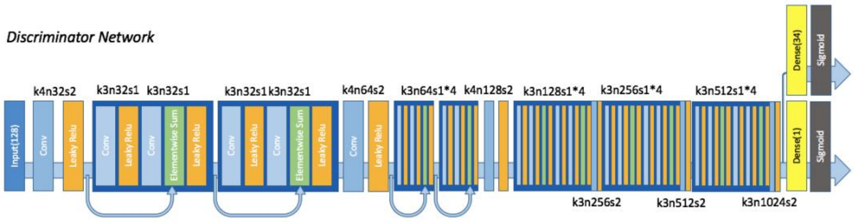
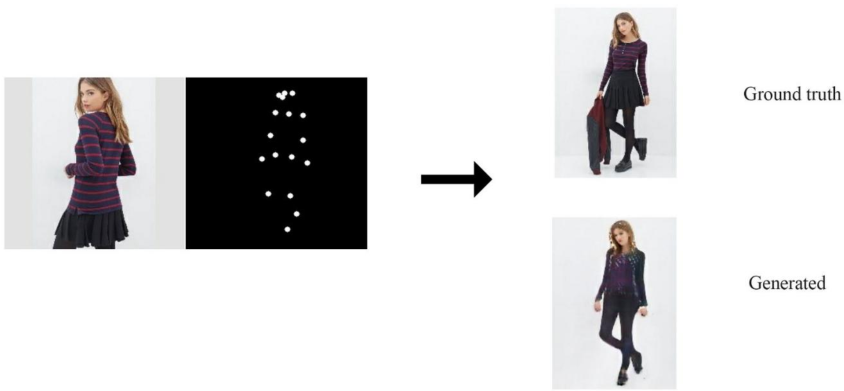


Figure 4: Discriminator Architecture

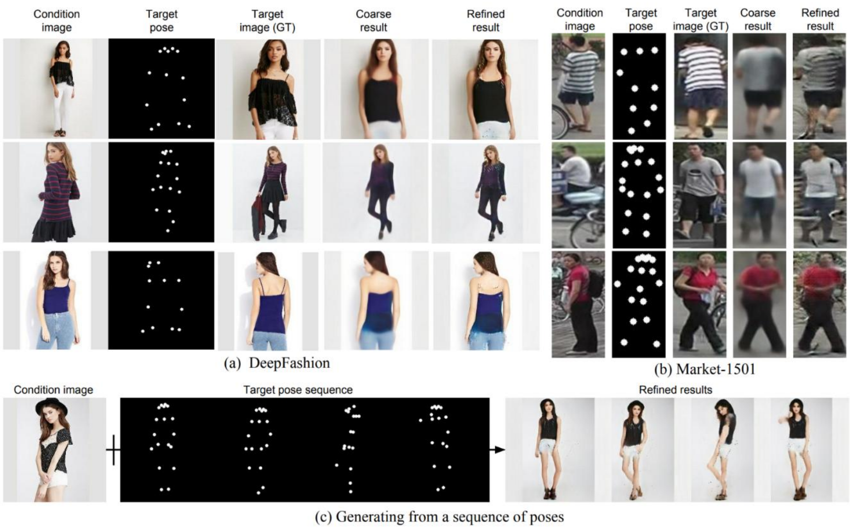
Source

Pose Guided Person Image Generation

With an additional input of the pose, we can transform an image into different poses. For example, the top right image is the ground truth while the bottom right is the generated image.

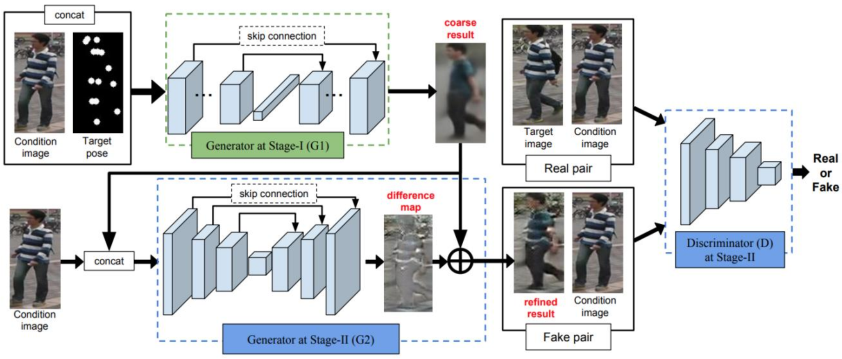


The refined result column below is the generated images.



Pose Guided Person Image Generation

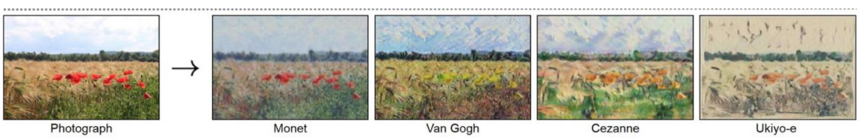
The design composes of a 2-stage image generator and a discriminator. The generator reconstruct an image using the meta-data (pose) and the original image. The discriminator uses the original image as part of the label input to a CGAN design.



Pose Guided Person Image Generation

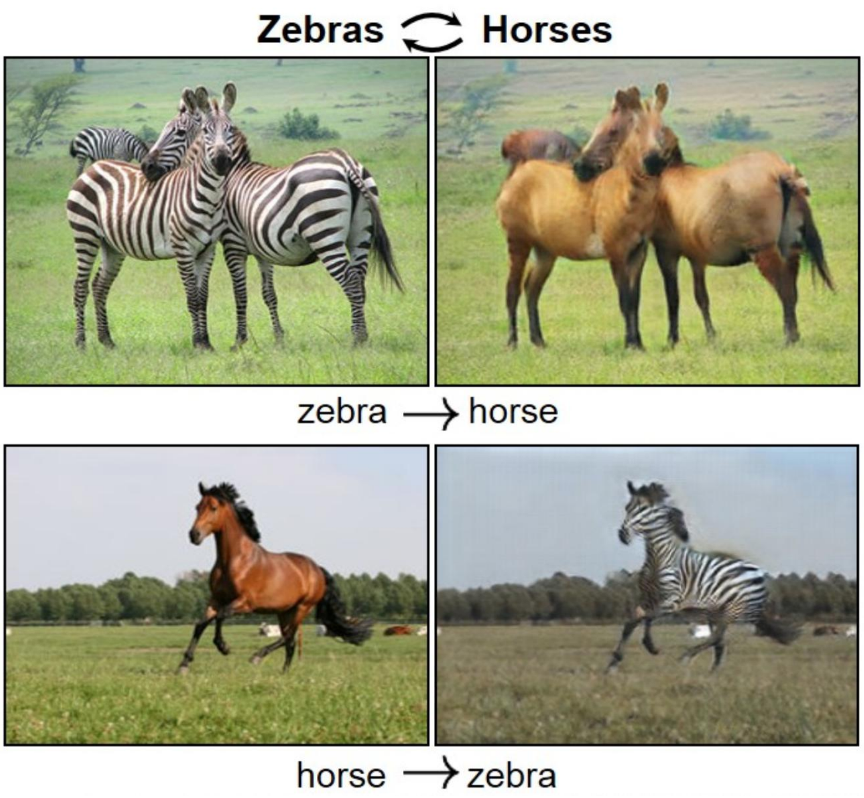
CycleGAN

Cross-domain transfer GANs will be likely the first batch of commercial applications. These GANs transform images from one domain (say real scenery) to another domain (Monet paintings or Van Gogh).



CycleGAN

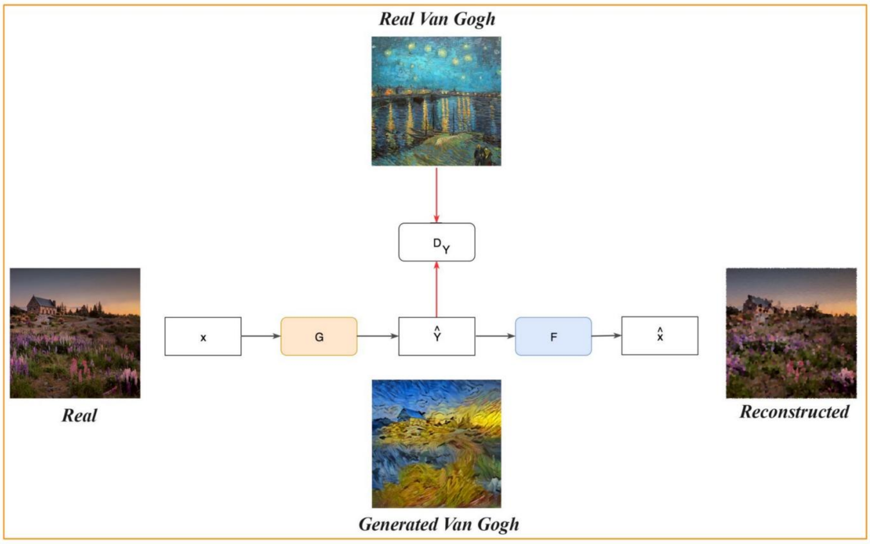
For example, it can transform pictures between zebras and horses.



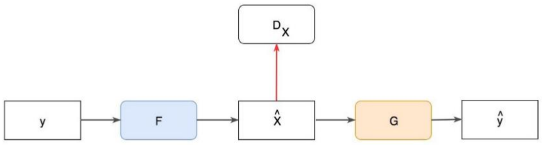
CycleGAN

CycleGAN builds 2 networks ***G*** and ***F*** to construct images from one domain to another and in the reverse direction. It uses discriminators *D* to critic how well the generated images are. For example, ***G*** converts real images to Van Gogh style painting and *Dy* is used to distinguish whether the image is real or generated.

Domain A → Domain B:

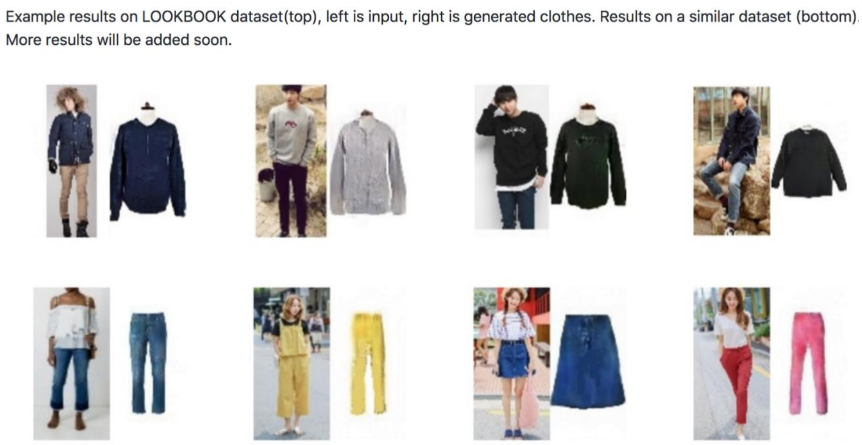


We repeat the process in the reverse direction Domain B→ Domain A:



PixelDTGAN

Suggesting merchandise based on celebrity pictures has been popular for fashion blogger and e-commerce. PixelDTGAN creates clothing images and styles from an image.



PixelDTGAN

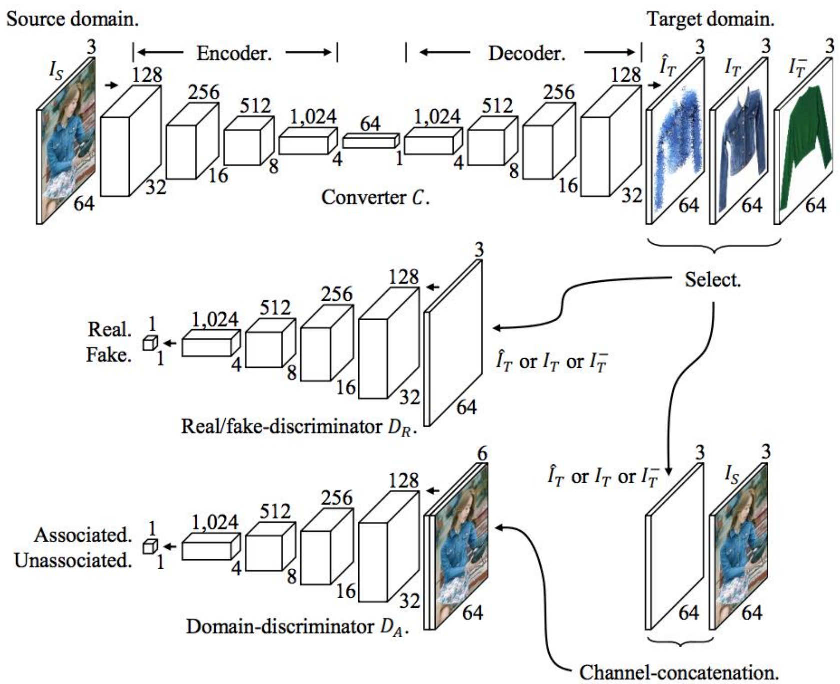


Fig. 2. Whole architecture for pixel-level domain transfer.

PixelDTGAN

Super resolution

Create super-resolution images from the lower resolution. This is one area where GAN shows very impressive result with immediate commercial possibility.

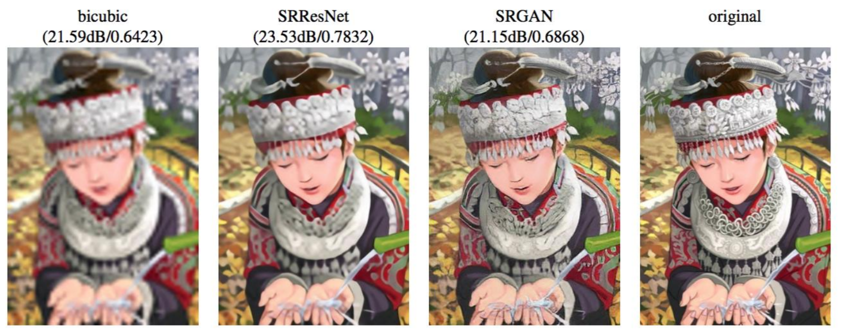
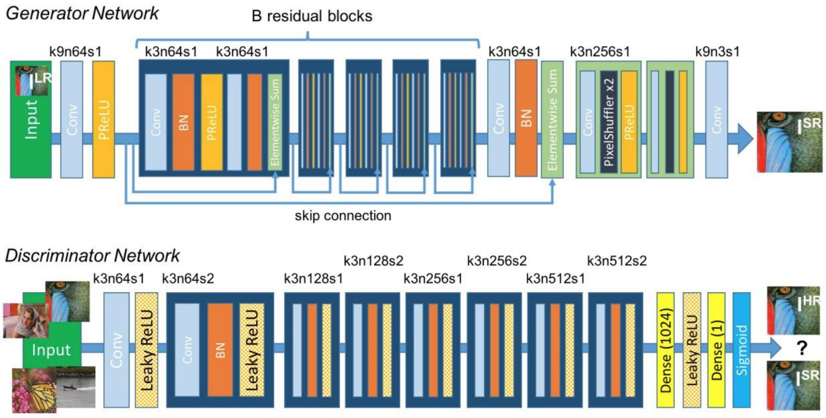


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

SRGAN

Similar to many GAN designs, it composes of many layers of convolutional layer, batch normalization, advanced ReLU and skip connections.



SRGAN

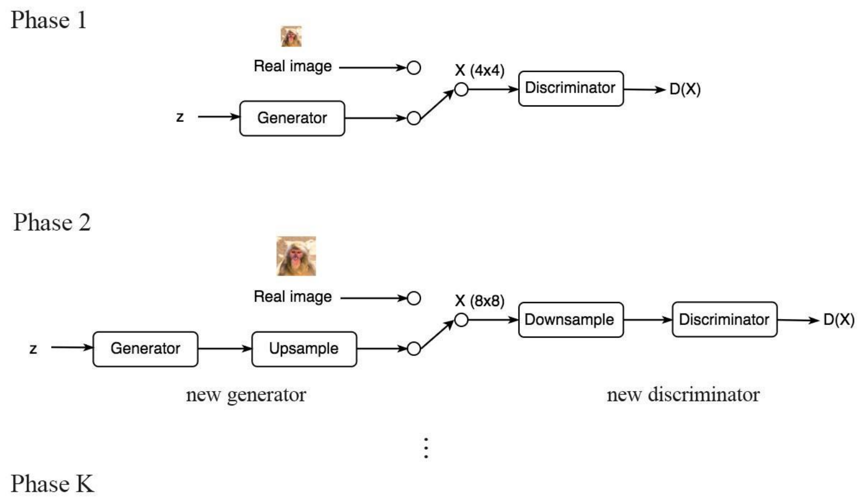
Progressive growing of GANs

Progressive GAN is probably one of the first GAN showing commercial-like image quality. Below is 1024 × 1024 celebrity look images created by GAN.

Figure 5: 1024 × 1024 images generated using the CELEBA-HQ dataset. See Appendix F for a larger set of results, and the accompanying video for latent space interpolations.

Progressive growing of GANs

It applies the strategy of divide-and-conquer to make training much feasible. Layers of convolution layers are trained once at a time to build images of 2× resolution.



In 9 phases, a 1024 × 1024 image is generated.

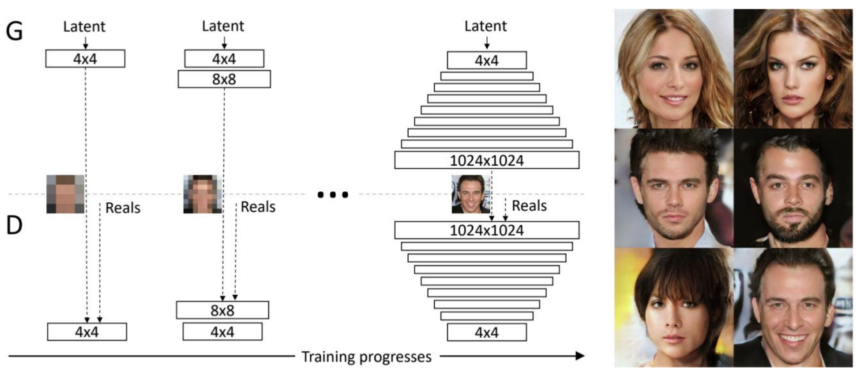


Figure 1: Our training starts with both the generator (G) and discriminator (D) having a low spatial resolution of 4×4 pixels. As the training advances, we incrementally add layers to G and D, thus increasing the spatial resolution of the generated images. All existing layers remain trainable throughout the process. Here $\boxed{N \times N}$ refers to convolutional layers operating on $N \times N$ spatial resolution. This allows stable synthesis in high resolutions and also speeds up training considerably. One the right we show six example images generated using progressive growing at 1024×1024.

Progressive growing of GANs

High-resolution image synthesis

This is not image segmentation! It is the reverse, generating images from a semantic map. Collecting samples are very expensive. We have trying to supplement training dataset with generated data to lower development cost. It will be handy to generate videos in training autonomous cars rather than see them cruising in your neighborhood.

pix2pixHD

Network design:

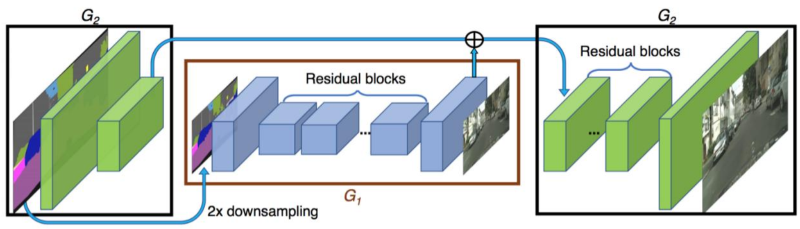


Figure 2: Network architecture of our generator. We first train a residual network G_1 on lower resolution images. Then, another residual network G_2 is appended to G_1 and the two networks are trained jointly on high resolution images. Specifically, the input to the residual blocks in G_2 is the element-wise sum of the feature map from G_2 and the last feature map from G_1.

pix2pixHD

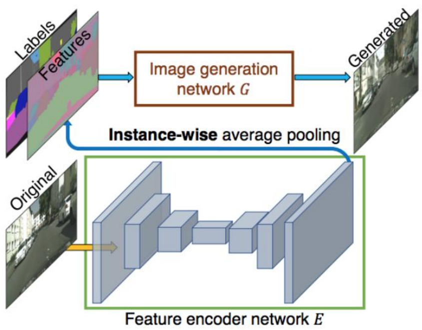


Figure 5: Using instance-wise features in addition to labels for generating images.

pix2pixHD

Text to image (StackGAN)

Text to image is one of the earlier application of domain-transfer GAN. We input a sentence and generate multiple images fitting the description.

StackGAN

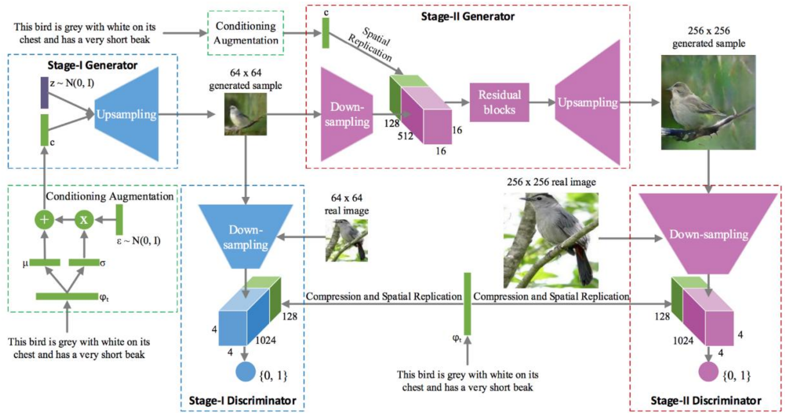


Figure 2. The architecture of the proposed StackGAN. The Stage-I generator draws a low resolution image by sketching rough shape and basic colors of the object from the given text and painting the background from a random noise vector. The Stage-II generator generates a high resolution image with photo-realistic details by conditioning on both the Stage-I result and the text again.

Source

Text to Image Synthesis

Another popular implementation:

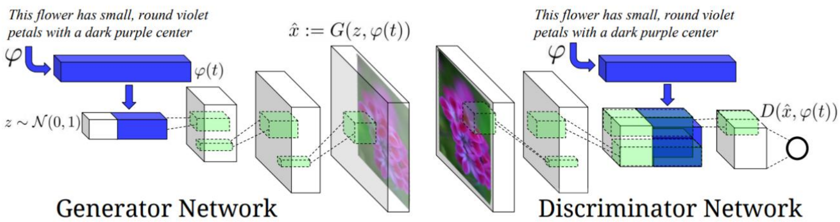


Figure 2. Our text-conditional convolutional GAN architecture. Text encoding $\varphi(t)$ is used by both generator and discriminator. It is projected to a lower-dimensions and depth concatenated with image feature maps for further stages of convolutional processing.

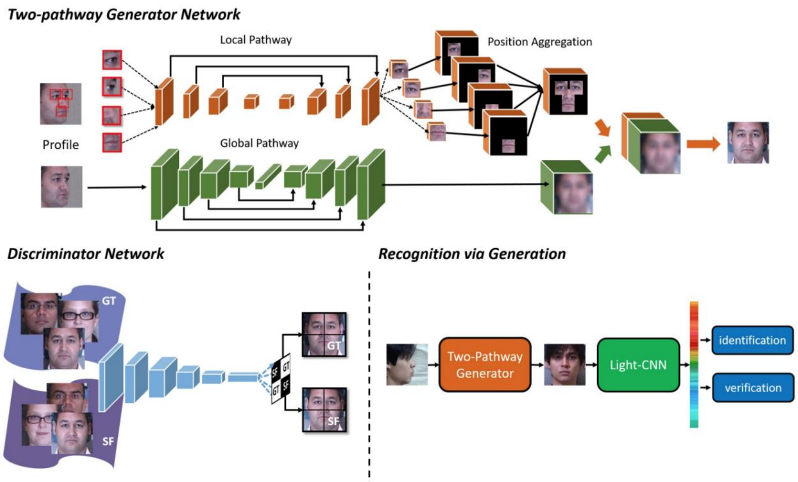
Generative Adversarial Text to Image Synthesis

Face synthesis

Synthesis faces in different poses: With a single input image, we create faces in different viewing angles. For example, we can use this to transform images that will be easier for face recognition.

. Synthesis results under various illuminations. The first row is the synthesized image, the second row is the input.

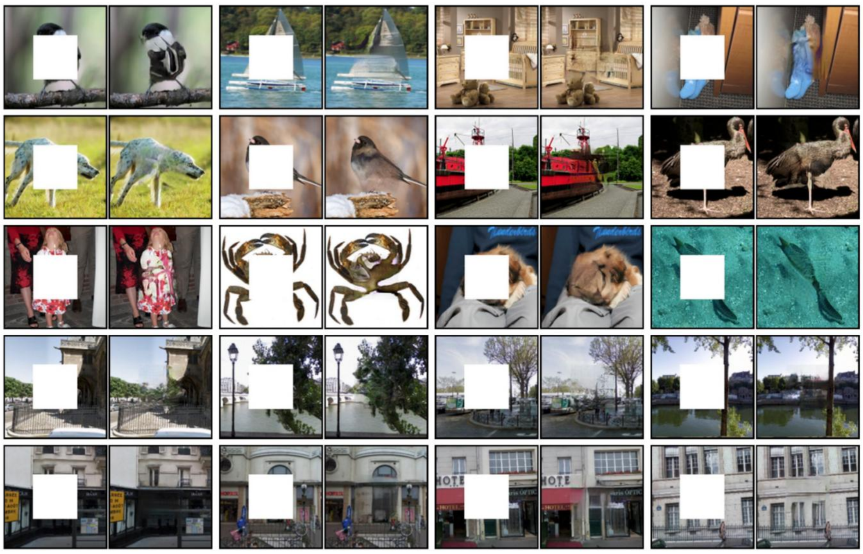
TP-GAN



TP-GAN

Image inpainting

Repair images have been an important subject decades ago. GAN is used to repair images and fill the missing part with created "content".



Context encoder

Learn Joint Distribution

It is expensive to create GANs with different combinations of facial characters *P(blond, female, smiling, with glasses)*, *P(brown, male, smiling, no glasses)* etc…The curse of dimensionality makes the number of GANs to grow exponentially. Instead, we can learn individual data distribution and combine them to form different distributions. i.e. different attribute combinations.

CoGAN learns the joint probability $P(x_1, x_2)$ by sampling $x_1 \sim P(x_1)$ and $x_2 \sim P(x_2)$

Figure 4: Generation of face images with different attributes using CoGAN. From top to bottom, the figure shows pair face generation results for the blond-hair, smiling, and eyeglasses attributes. For each pair, the 1st row contains faces with the attribute, while the 2nd row contains corresponding faces without the attribute.

CoGAN

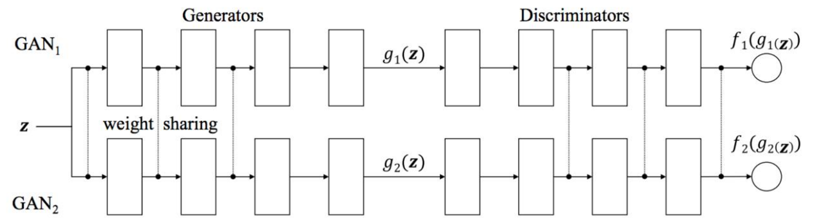


Figure 1: CoGAN consists of a pair of GANs: GAN_1 and GAN_2. Each has a generative model for synthesizing realistic images in one domain and a discriminative model for classifying whether an image is real or synthesized. We tie the weights of the first few layers (responsible for decoding high-level semantics) of the generative models, g_1 and g_2. We also tie the weights of the last few layers (responsible for encoding high-level semantics) of the discriminative models, f_1 and f_2. This weight-sharing constraint allows CoGAN to learn a joint distribution of images without correspondence supervision. A trained CoGAN can be used to synthesize pairs of corresponding images—pairs of images sharing the same high-level abstraction but having different low-level realizations.

CoGAN

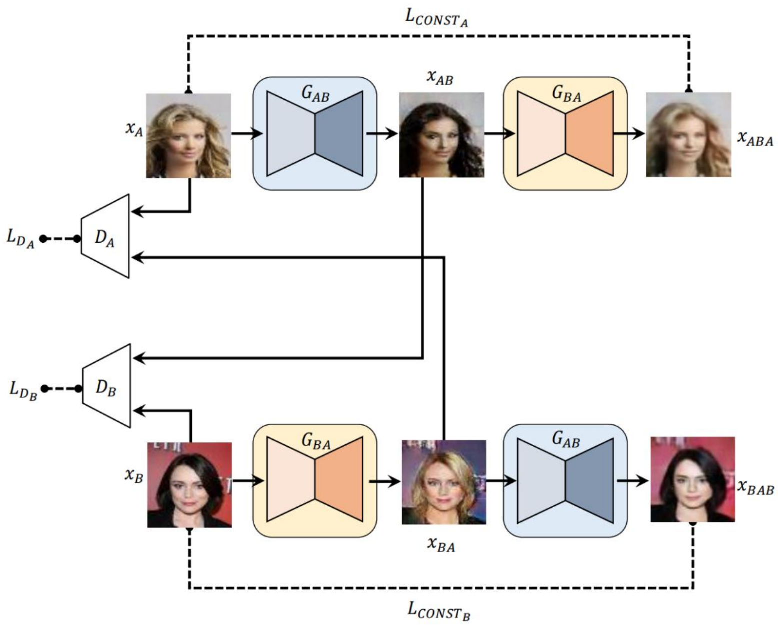
DiscoGAN

DiscoGAN provides matching style: many potential applications. DiscoGAN learns cross domain relationship without labels or pairing. For example, it successfully transfers style (or patterns) from one domain (handbag) to another (shoe).

(b) Handbag images (input) & **Generated** shoe images (output)

DiscoGAN

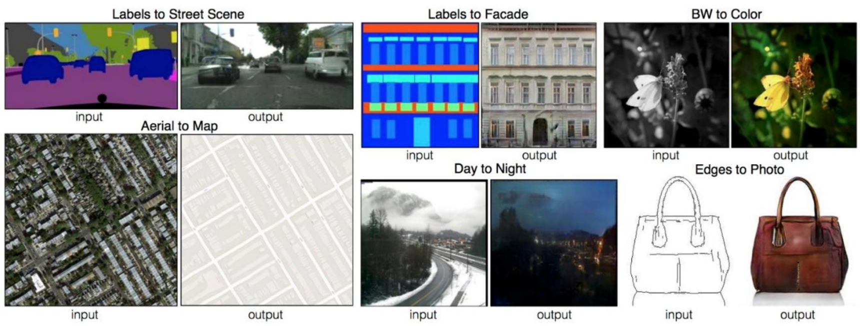
DiscoGAN and CycleGAN are very similar in the network design.



DiscoGAN

Pix2Pix

Pix2Pix is an image-to-image translation that get quoted in cross-domain GAN's paper frequently. For example, it converts a satellite image into a map (the bottom left).



pix2pix

DTN

Creating Emoji from pictures.

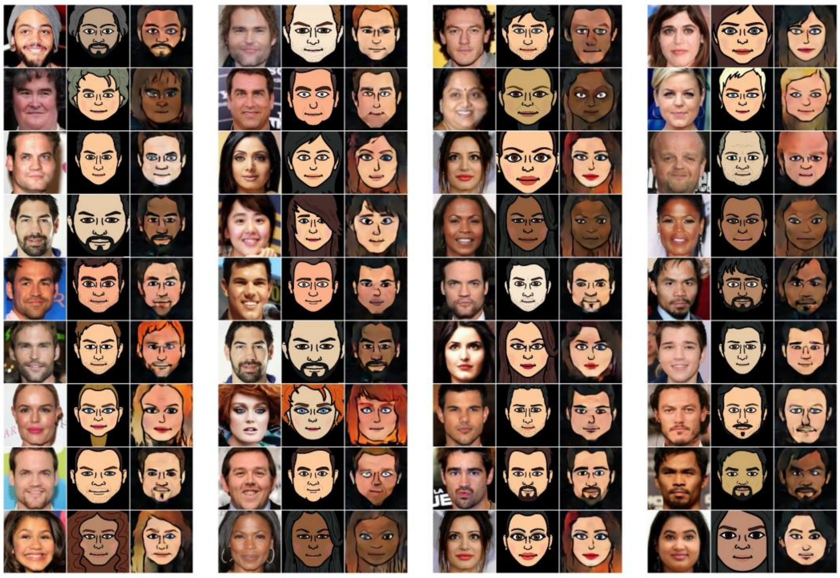


Figure 4: Shown, side by side are sample images from the CelebA dataset, the emoji images created manually using a web interface (for validation only), and the result of the unsupervised DTN. See Tab. 4 for retrieval performance.

DTN

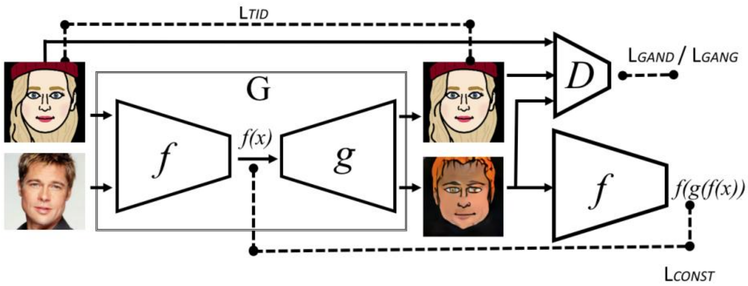
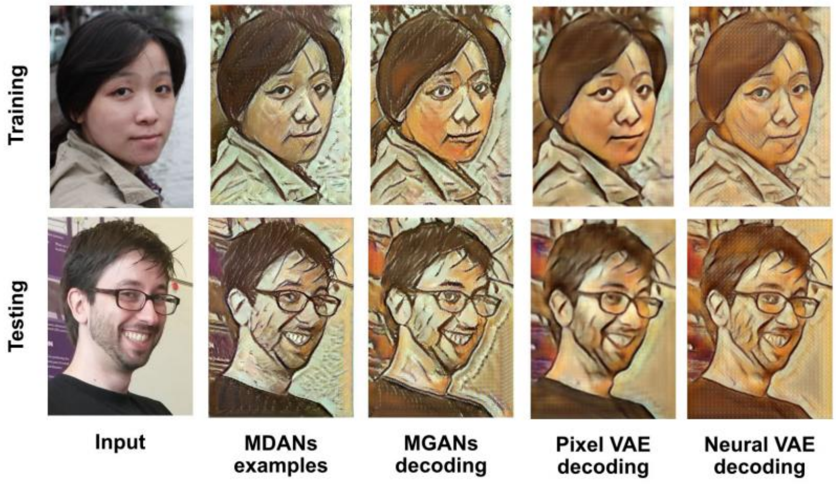


Figure 1: The Domain Transfer Network. Losses are drawn with dashed lines, input/output with solid lines. After training, the forward model G is used for the sample transfer.

DTN

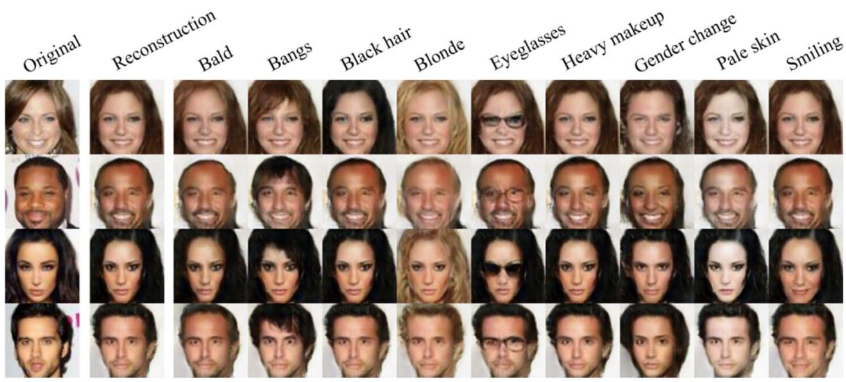
Texture synthesis



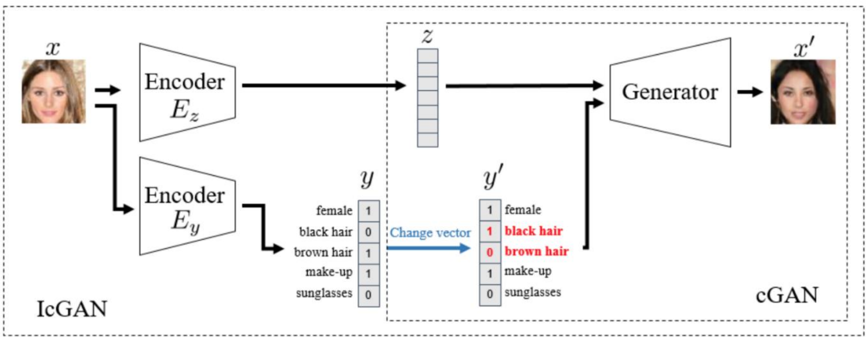
MGAN

Image editing (IcGAN)

Reconstruct or edit images with specific attributes.

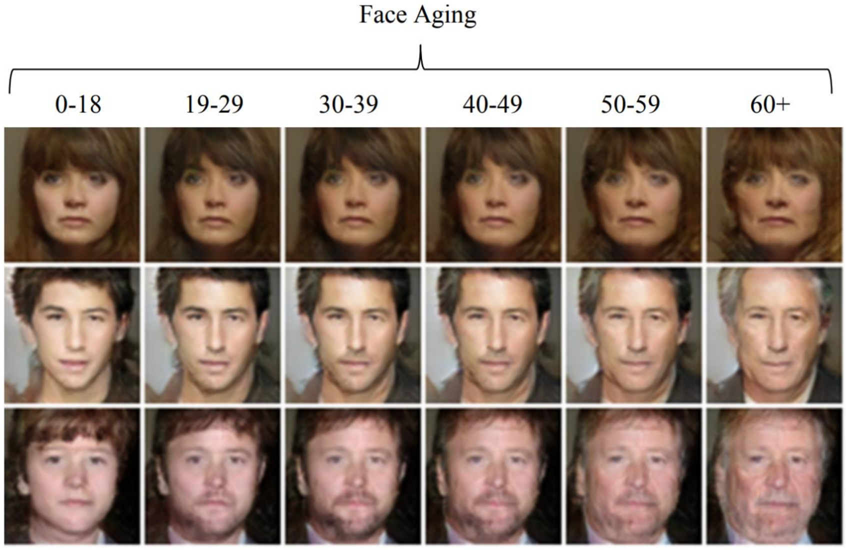


IcGAN

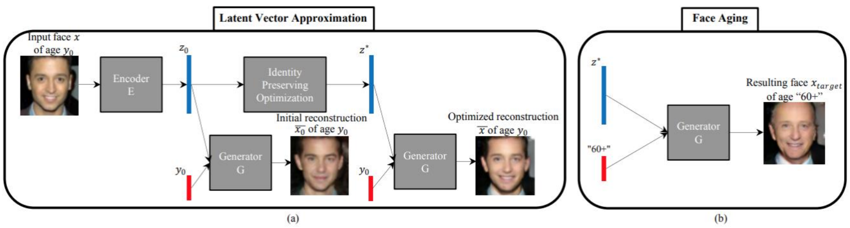


IcGAN

Face aging (Age-cGAN)



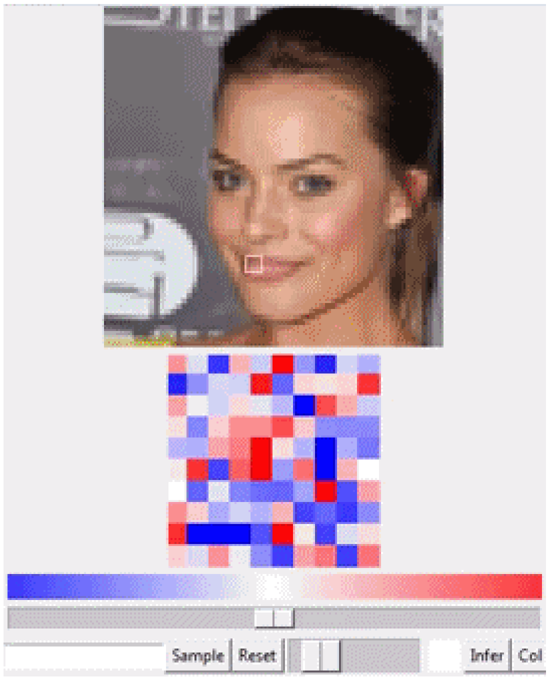
Age-cGAN



Age-cGAN

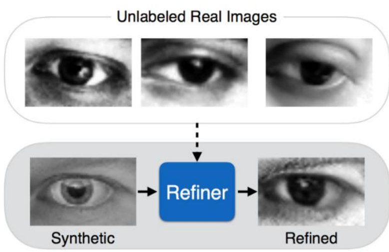
Neural Photo Editor

Content based image editing: for example, extend the hairband.



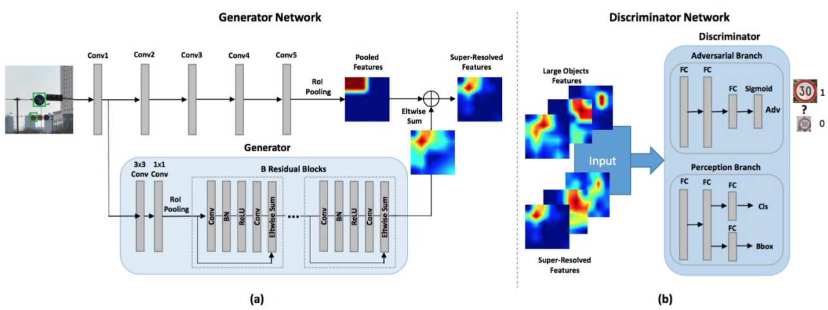
Neural Photo Editor

Refine image



Object detection

This is one application in enhancing an existing solution with GAN.



Perceptual GAN

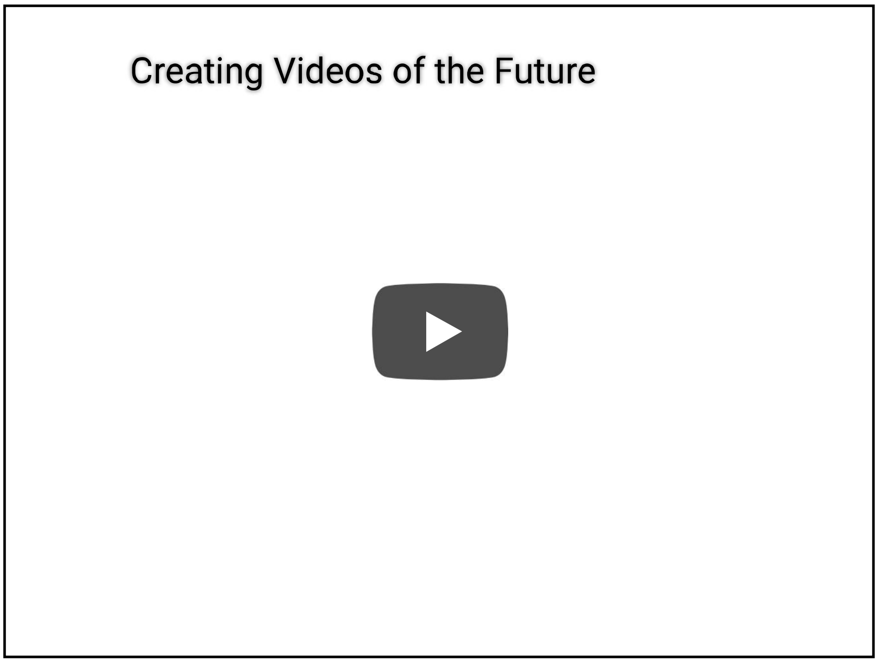
Image blending

Blending images together.

GP-GAN

Video generation

Create new video sequence. It recognizes what is background and create new time sequence for the foreground action.



Generate 3D objects

This is one often quoted paper in creating 3D objects with GAN.

3DGAN

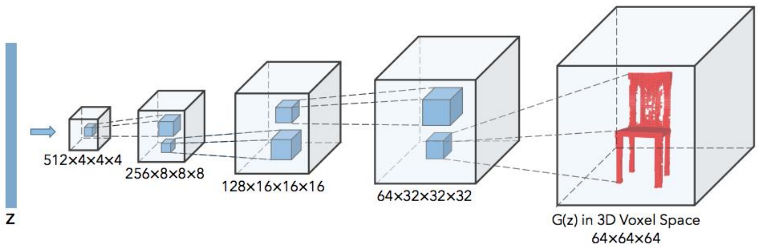


Figure 1: The generator in 3D-GAN. The discriminator mostly mirrors the generator.

3DGAN

Music generation

GAN can be applied to non-image domain, like composing music.

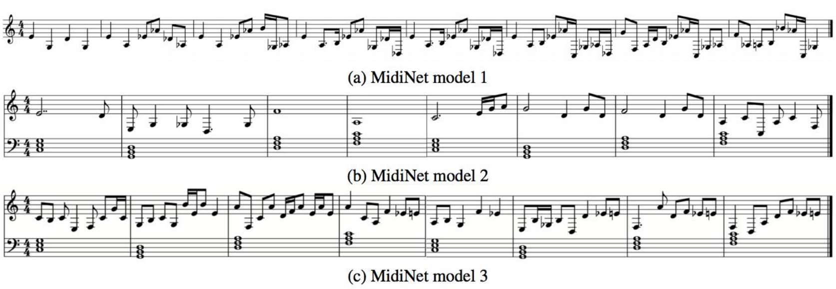
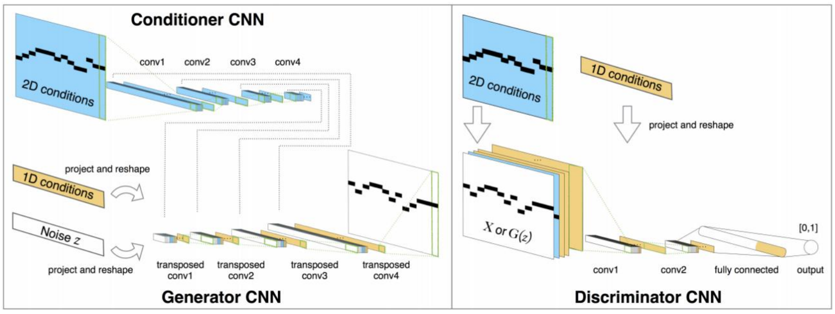


Figure 3. Example result of the melodies (of 8 bars) generated by different implementations of MidiNet.

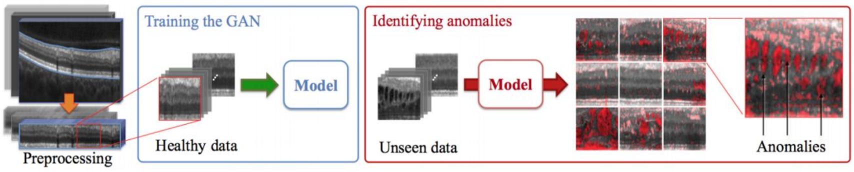
MidiNet



MidiNet

Medical (Anomaly Detection)

GAN can also extend to other industry, for example medical in tumor detection.



AnoGAN

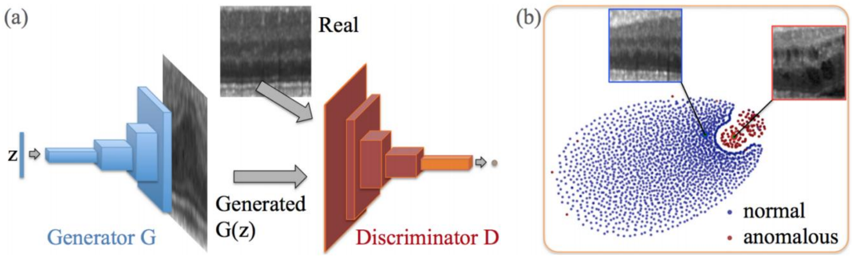


Fig. 2. (a) Deep convolutional generative adversarial network. (b) t-SNE embedding of normal (blue) and anomalous (red) images on the feature representation of the last convolution layer (orange in (a)) of the discriminator.

AnoGAN

Further Readings

This article shows some of the GAN application. For those interested in further study of GAN:

Part 1: Focus on how GANs are applied to solve deep learning problems, and an overview of why it is so hard to train GANs.

GAN — A comprehensive review into the gangsters of GANs (Part 1)

Are we there yet? In this GAN series, we identify a general pattern on how GAN is applied to deep…

medium.com

Part 2: An overview of solving the training problems in GAN.

GAN — A comprehensive review into the gangsters of GANs (Part 2)

This article studies the motivation and the direction of the GAN research in improving GAN…

medium.com

All the articles in this series:

GAN — GAN Series (from the beginning to the end)

A full listing of our articles covers the applications of GAN, the issues, and the solutions.

medium.com