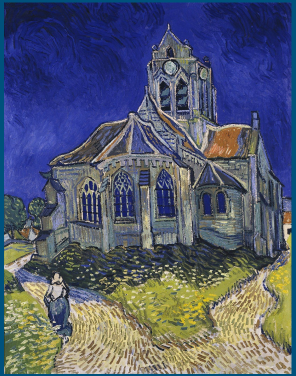
GAN Computer Generate Art: A GANs Survey

by
Bharath Gunasekaran

What is Art?

- An essentially part of human culture
- A way to express imagination, thoughts, memories, and ideas
- Categorized at Visual Arts and Performing Arts



Images from https://en.wikipedia.org/wiki/Visual_arts, https://www.stac.edu/academics/school-arts-sciences/school-arts-sciences-academic-programs/performing-arts-minor

Early Computer Art

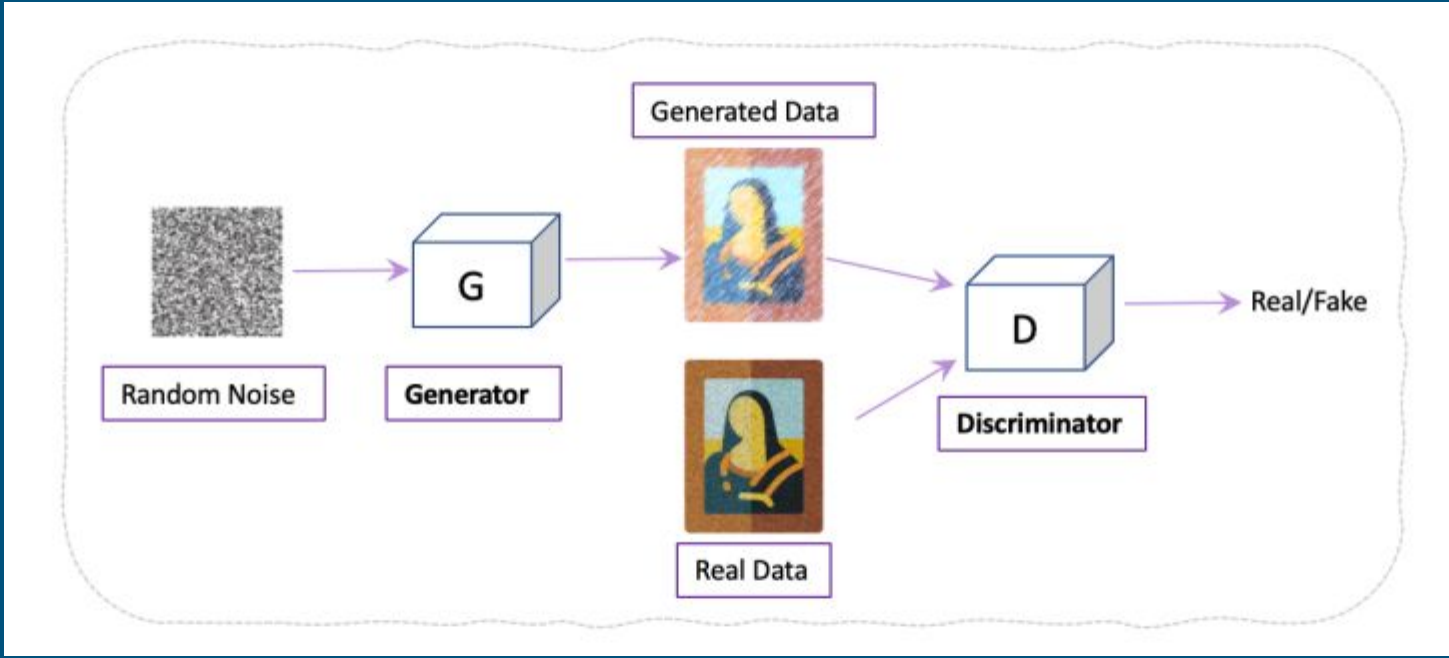
- Very Uniform and lifeless
- Require human intervention

Age of Deep Learning

- Advancement of Deep Learning
  - Allow for computer to generate content without human intervention

- GANS
  - Visual Art
  - Musical Art
  - Literary text

What are GANs?



GANs Architectures

All GAN architectures are based on generator and discriminator

- Conditional GAN
- Deep Convolutional GAN (DCGAN)
- Recurrent Adversarial Networks

Conditional GAN

- allowing the generator and discriminator to access auxiliary information such as class labels
- Allow generator to generate content based on a condition

Deep Convolutional GAN

- The generator and the discriminator are made up of convolution networks (CNN)
- Convolution Layers are used to apply up sampling and down sampling
- Good at extract features from images

Recurrent Adversarial Network

- Suitable for sequential or time dependent data such as text or audio generation
- Consists of a encoder and decoder

Common Loss Functions

- Cross-Entropy Loss
- Binary Cross-Entropy Loss
- Mean Square Loss
- KL divergence Loss
- Wasserstein distance

$$H(p,q) = -\sum_{\forall x} p(x)\log\left(q(x)\right)$$

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(p(y_i)\right) + (1-y_i)\cdot\log\left(1-p(y_i)\right)$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$\text{KL}(P \parallel Q) = \sum_{x\in\mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

$$l_1(u,v) = \inf_{\pi\in\Gamma(u,v)} \int_{\mathbb{R}\times\mathbb{R}} |x-y|\,\mathrm{d}\pi(x,y)$$

Visual Arts with GANs

- Generate cartoon images from a sketch
    - Conditional GANs
    - U-net Generator
    - CNN discriminator
    - Outperformed existing solutions
- Unconditional Art generation
    - modified DCGANs
    - Deep CNN Generator
    - CONV discriminator
    - 53% better than DCGANs

Source	Task	GAN Type	Loss Function	Generator-Discriminator Architecture	Result
[24]	Generate cartoon image from sketch	Conditional GAN	Cross entropy, L1 distance for pixel-level loss	U-Net Generator, CNN-FC discriminator	Qualitative only, outperformed existing works
[25]	Generate shoe image from sketch	Conditional GAN	Binary cross-entropy loss for discriminator	U-Net Generator, Deep CNN discriminator	No evaluation
[26]	Generate fully colored synthetic images from sketch	Vanilla GAN with Encoder for image style recognition	Auxiliary Losses, discriminator is trained on style loss & content loss. MSE Loss.	U-Net Generator, DNN discriminator	FID 4.18, classification score 0.57. 63% evaluators preferred images from proposed model
[27]	Generate painting by brushstrokes	Conditional GAN	Wasserstein Loss	Configurations not provided	No evaluation
[28]	Unconditional art generation	DCGAN	Cross entropy loss and added classification, style ambiguity losses	Deep CNNs, CONV followed by LeakyReLU	53% of evaluators believe that synthesized images were by an artist.
[29]	Generate pre-modern Japanese art facial expression	StyleGAN	WGAN-GP	Configurations not provided	No evaluation
[31]	Generate face photo from a sketch and vice versa	Conditional GAN	Cross entropy	Configurations not provided	No evaluation, the loss values were reported
[32]	Generate digits and character strokes	Modified DCGAN + agent	MSE	Generator contains CONV+LeakyReLU, Agent is VGG	Generated images not evaluated, classification accuracy 91% on MNIST
[33]	Generate calligraphy and handwritten digits	Modified conditional GAN with multiple encoders for style & content-encoding	Binary Cross entropy discriminator, cross-entropy style loss & Kullback-Leibler content loss	Two residual blocks then 4 CONV modules generator, 1 CONV layer then 6 residual blocks discriminator	FID 120.1, 49.3% of the images were identified to be synthesized

Music Generation with GANs

- Melody Generation
  - LSTM-based GAN
  - Bi-directional LSTM generator
  - LSTM discriminator
- Generate Pop Music
  - Modified DCGAN
  - 6 layer with 2 Dense and 4 Convolution Layers generator
  - 3 layer Discriminator

Source	Task	GAN Type	Loss Function	Generator-Discriminator Architecture	Result
[34]	Generate melody for a specific genre	Hybrid VAE and GAN	KL Divergence	Four DeCONV layers for both generator & discriminator, VAE encoder used three Conv2D	No quantitative evaluation, concluded that consistency of generated melodies was not up to the same level as human composition
[36]	Melody Generation	LSTM-based GAN	Bayesian	Bi-LSTM generator and LSTM discriminator	Average score of 3.27 on the three qualitative metrics, 48% likely to be detected as synthetic
[37]	Generate pop music monophonic melodies	Modified DCGAN	Cross entropy	Two dense layers followed by four transposed CONV for generator; 2 CONV layers followed by a dense layer discriminator	Mean score around 3 for being pleasant & realistic, 4 for interesting people with musical backgrounds, 3.4 for people without musical backgrounds
[39]	Generate melodies based on lyrics	Conditional GAN	Cross entropy	LSTM generators and discriminators	No evaluation was provided
[41]	Generate melodies based on lyrics	Conditional LSTM GAN	Cross entropy	Dense layer followed by 2 LSTM followed by a dense layer for generator, 2 LSTM followed by dense for discriminator	BLEU-2 score of 0.735, scores of about 3.8, 3.5, 4.1 respectively out of 5 for lyrics, rhythm, and melody by evaluators
[42]	Generate single voice polyphonic music	RNN GAN	Cross entropy and Squared error loss	2 LSTM layers for generator, 2 Bi-directional LSTM layers followed by a dense for discriminator	No evaluation was provided
[43]	Generate multi-track, polyphonic music	Conditional GAN	Wasserstein	Generators contain 1D transposed CONV, discriminators 5 contain 1D Conv layers followed by one dense layer	The highest score for conditional generation was 3.1 and non-conditional was 3.16 out of 5 by 'non-pro' evaluators. For intra-track metrics, jamming model performed best
[44]	Generate folk music	RL GAN	Cross entropy and policy gradient	RNN generators, CNN discriminators	BLEU score of 0.94 and MSE of 20.6 outperformed baseline maximum likelihood estimation

Literary Text Generation

- Chinese Poetry Generation
  - RL GAN
  - RNN Generator
  - CNN discriminators
  - BLEU-2 score 0.74
- Generate Poetry from Image
  - GAN encoder and decoder
  - RNN Generator
  - CNN-RNN agent for encoding and decoding painting
  - LSTM for text generation

Source	Task	GAN Type	Loss Function	Generator-Discriminator Architecture	Dataset	Result
[44]	Chinese Poetry generation	RL GAN	Cross entropy and policy gradient	RNN Generators and CNN discriminators	16394 Chinese quatrains	BLEU-2 score of 0.74, overall score of 0.54 by human evaluators
[45]	Generate poetry from an image	Multiadversarial GAN with an embedding model	Cross entropy and policy gradient	RNN Generators, GRU-based discriminator, CNN image encoder and RNN poem decoder	Novel dataset with paired image and poetry	Overall BLEU score of 0.77, 7.18 out of 10 overall score by human evaluators
[46]	Generate Shakespearean prose from a painting	Multiadversarial GAN with encoder and decoder	Cross entropy and policy gradient	RNN Generator, CNN-RNN agent for encoding and decoding painting, LSTM encoder and decoder for generating prose	Two datasets for generating English poem from an image, and Shakespeare plays and their English translations for text style transfer	Average scores of 3.7, 3.9, and 3.9 out of 5 by evaluators for content, creativity, and similarity to Shakespearean style respectively
[47]	Chinese Poetry generation	RL GAN	Maximum-likelihood	A single layer LSTM generator, two-layer Bi-directional LSTMs discriminator	Poem-5 and Poem-7 Chinese Poem dataset	BLEU-2 scores of 0.76 and 0.55 for the two datasets respectively
[48]	Chinese Poetry generation	RNN GAN	Wasserstein distance	LSTM generator and discriminator with WGAN-GP training	Poem-5 and Poem-7 Chinese Poem dataset	BLEU-2 scores of 0.88 and 0.67 for the two datasets respectively
[50]	Chinese Poetry generation	GAN with a ranking function	Ranking objective	LSTM generator, CNN-based ranker	Over 13,000 Chinese quatrains	BLEU-2 score of 0.81, 4.6 out of 10 overall score by human evaluators
[50]	Learn rare words from Romeo and Juliet play	GAN with a ranking function	Ranking objective	LSTM generator, CNN-based ranker	Over 3000 sentences from Romeo and Juliet play	BLEU-2 score of 0.914
[51]	Poetry and lyrics generation	GAN with language model generator	Cross entropy	AWD-LSTM [52] and TransformerXL [53] language model for generator, discriminator encoder-decoder pair.	740 classical and contemporary English poems and 1500 song lyrics across various genres	Perplexity score of 42.5 for poetry and 9.02 for lyrics generations

Challenges

- Image generation require a lot of data to train
- Music generation is time dependent which leads to complex GAN architectures
- Literary GAN research is limited to poetry generation
    - More research need to be done to draw a conclusion

Future Work

- Experiment with smaller dataset and GAN architectures for visual arts generation.
- Implement GANs to generate music in raw audio format as opposed to MIDI file format.
- Implement longer text literary work generations including novels and dramas.
- Propose a well-defined and comprehensive qualitative validation for visual and performing arts

Reference

https://arxiv.org/ftp/arxiv/papers/2108/2108.03857.pdf

All images are originally from the paper