ArtGAN: Artwork synthesis with Conditional Categorical GANs

Wei Ren Tan [1], Chee Seng Chan[2], Hernán E. Aguirre[1], and Kiyoshi Tanaka[1]

[1]Shinshu University, Japan. [2]University of Malaya, Malaysia.

Motivations

- Recently, GAN showed significant promise in generating natural images, e.g. from MNIST, CIFAR-10, CUB-200, or LFW datasets.
- Subjects in these images have structured shape such as numeric, vehicles, birds, face, etc.
- Can GAN learn to draw based on more abstract classes?

Challenges

- Many paintings are non-representative nor figurative
- Some paintings are classified based on non-visual background knowledge, e.g. Renaissance paintings are artworks from Renaissance period.

Inspiration

- An artist teacher [1] pointed out that an effective learning in art domain requires one to focus on a particular skill.
- For instance, practice to draw one kind of movement at a time.
- To imitate this learning skill, ArtGAN is trained with additional information by training a classifier and backpropagate the class errors.

Goals

- End-to-end train ArtGAN to synthesize artworks based on *style*, *genre*, or *artist*.
- Investigate what kind of features GAN learnt from fine-art paintings.

Wikiart Paintings Dataset[2]

- Collection of $> 80,000$ fine-art paintings ranging from 15th century to modern times.
- 27 *styles* from **all** paintings.
- 10 *genres* with $> 1,500$ paintings ($\sim 65,000$ samples).
- 23 *artists* with > 500 paintings ($\sim 20,000$ samples).

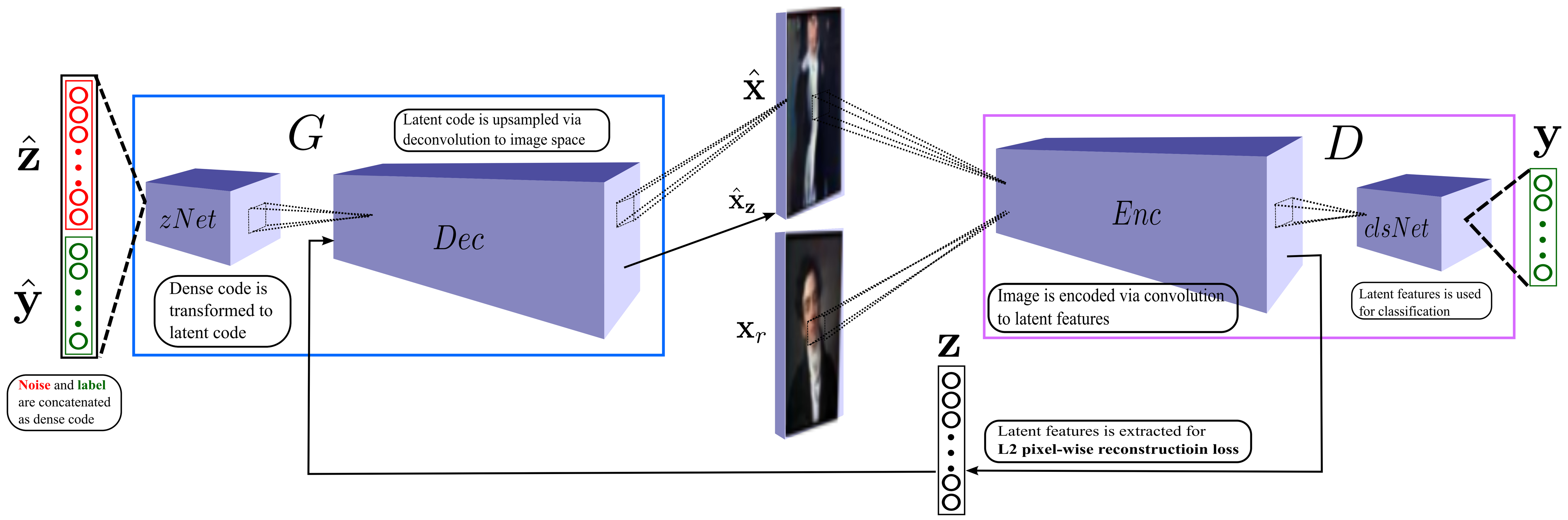
References

[1] Paul Foxton.
How to practise drawing effectly, 2011.

[2] Babak Saleh and Ahmed Elgammal.
Large-scale classification of fine-art paintings: Learning the right metric on the right feature.
arXiv preprint arXiv:1505.00855, 2015.

ArtGAN

Overview of Architecture:



Loss functions:

- Discriminator

$$\mathcal{L}_D = -\mathbb{E}_{(\mathbf{x}_r,k)\sim p_{data}}\left[y_k \log p(y_k|\mathbf{x}_r)+\sum_{i\neq k}\log(1-p(y_i|\mathbf{x}_r))\right]-\mathbb{E}_{\hat{\mathbf{z}}\sim p_{noise},\hat{k}\sim\mathbf{K}}\left[\log p(y_{K+1}|G(\hat{\mathbf{z}},\hat{\mathbf{y}}_{\hat{k}}))+\sum_{i<K+1}\log(1-p(y_i|G(\hat{\mathbf{z}},\hat{\mathbf{y}}_{\hat{k}})))\right] \quad (1)$$

- Generator

$$\mathcal{L}_G = -\mathbb{E}_{\hat{\mathbf{z}}\sim p_{noise},\hat{k}\sim\mathbf{K}}\left[\log p(y_{\hat{k}}|G(\hat{\mathbf{z}},\hat{\mathbf{y}}_{\hat{k}}))+\sum_{i\neq\hat{k}}\log(1-p(y_i|G(\hat{\mathbf{z}},\hat{\mathbf{y}}_{\hat{k}})))\right]+\mathbb{E}_{\mathbf{x}_r\sim p_{data}}\left[||Dec(Enc(\mathbf{x}_r))-\mathbf{x}_r||_2^2\right] \quad (2)$$

Architecture details:

- Strided convolution layers in D
- Strided deconvolution layers in G
- Each layer is followed by batch normalization and leaky ReLU ($\alpha = 0.2$), except first layer in D and last layer in D and G.

Training details:

- RmsProp optimizer with decay rate of 0.9
- Mini-batch size: 128
- Init. learning rate: 0.001
- Learning rate reduction: factor of 10 at epoch 80
- Max. Iter.: 100 epochs.

Data Augmentations:

- Normalized to $[-1, 1]$
- Random cropped
- Resized to 64×64 pixels
- Random horizontal reflection

Experimental Results (CIFAR-10)

DCGAN

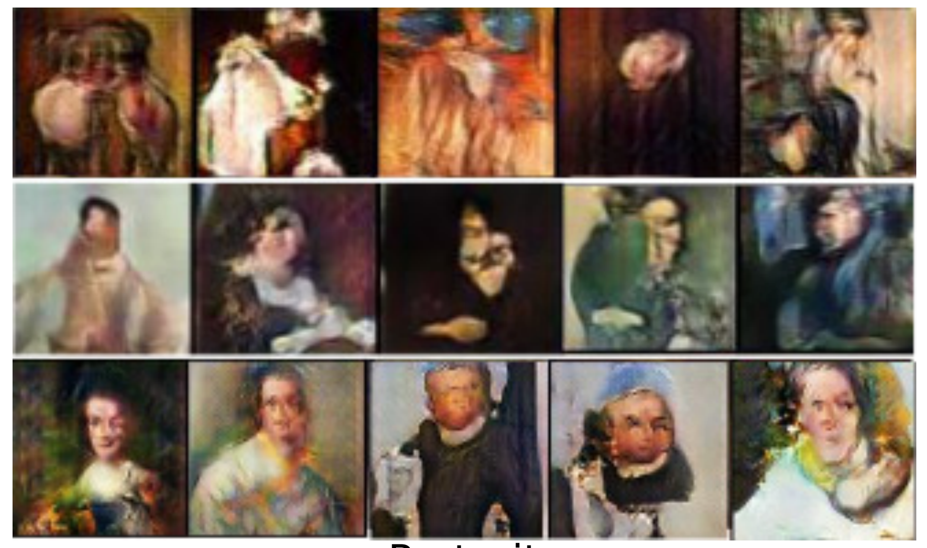
ArtGAN

Model	Log-likelihood
DCGAN	2348 ± 67
GAN/VAE	2483 ± 67
ArtGAN	$\mathbf{2564 \pm 67}$

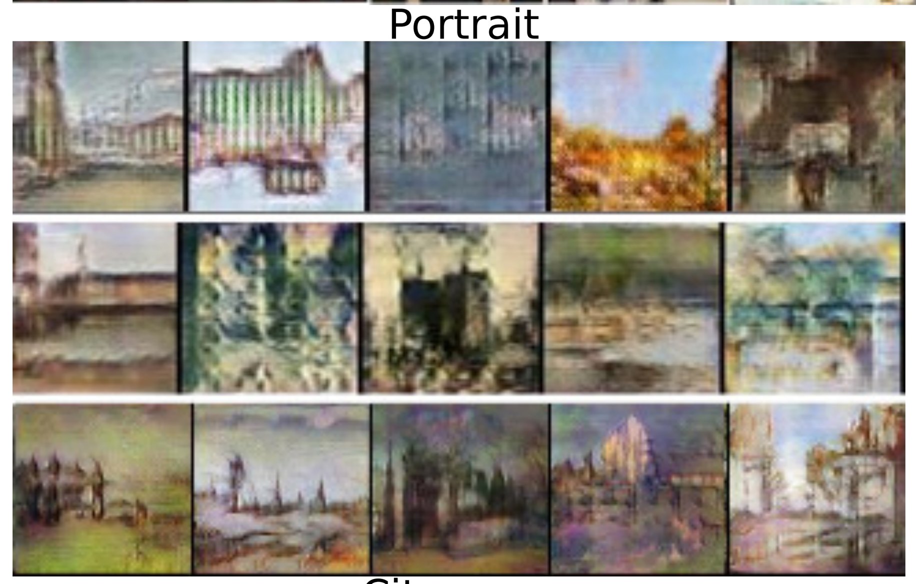
- ArtGAN has the best log-likelihood.
- Objects generated are more recognizable compared to artworks.
- ArtGAN is able to generate much compelling images compared to DCGAN.

Synthetic Artworks

Genre:



Portrait



Cityscape

Still Life

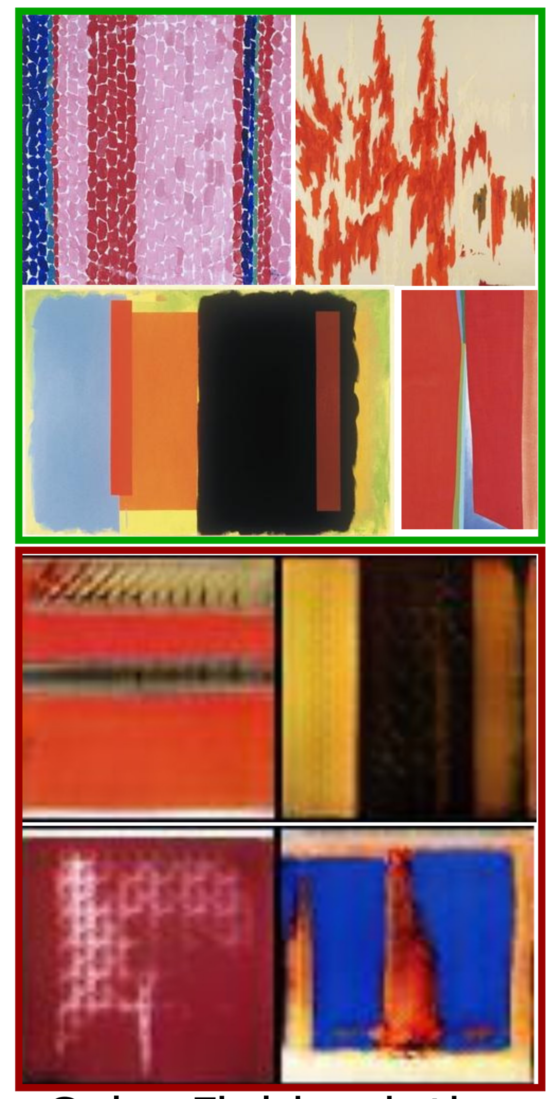
Artist:

Ivan Aivazovsky (Marine art)

Gustave Dore (Engraving)

Vincent van Gogh (Sketch)

Style:



Color Field painting

Ukiyo-e

- For *genre*, from top to bottom are: DCGAN, GAN/VAE, ArtGAN.
- *Genre*: ArtGAN generates a lot more natural artworks.
- *Artist*: Generated artworks show the painters' preferences.
- *Style*: Most difficult to train because class definition is confusing.

Future works

- A natural extension is to train better model to generate artworks with better details and higher resolution.
- Jointly learn the modes from *genres*, *artists*, and *styles*, such that ArtGAN can create artwork based on the combination of several modes.