ArtGAN: Artwork synthesis with Conditional Categorical GANs

Wei Ren Tan¹, Chee Seng Chan², Hernán E. Aguirre¹, and Kiyoshi Tanaka¹

¹Shinshu University, Japan. ²University 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 (~65,000 samples).
- 23 artists with > 500 paintings (~20,000 samples).

References


Overview of Architecture:

ArtGAN

- Dense code is transformed to latent code.
- Image is translated via convolutional neural network.
- Dense code is transformed to dense code.
- Data error is backpropagation.

ArtGAN

Generator:

\[ L_G = -\mathbb{E}_{z \sim p(z)} [\log p(y|G(z)) + \alpha \sum_{i,k} \log(1-p(y_i|G(z,y_i)))] \]

Discriminator:

\[ L_D = -\mathbb{E}_{(x,y) \sim \text{data}} [\log p(y|x)] - \mathbb{E}_{z \sim \text{normal}} [\log p(y|x)] - \mathbb{E}_{z \sim \text{normal}} [\log p(y|x)] \]

Architecture details:

- Strided convolution layers in D
- Strided deconvolution layers in G
- Each layer is followed by batch normalization and leaky ReLU
- Max. Iter.: 100 epochs.

Training details:

- RmsProp optimizer with decay rate of 0.9
- Mini-batch size: 128
- Learning rate: 0.001
- Learning rate reduction: factor of 10 at epoch 80
- Resized to 64 × 64 pixels
- Random horizontal reflection

Data Augmentations:

- Normalized to [−1, 1]
- Random cropped
- Random horizontal reflection

Experimental Results (CIFAR-10)

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGAN</td>
<td>2348 ± 67</td>
</tr>
<tr>
<td>GAN/VAE</td>
<td>2483 ± 67</td>
</tr>
<tr>
<td>ArtGAN</td>
<td>2564 ± 67</td>
</tr>
</tbody>
</table>

ArtGAN has the best log-likelihood.

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.