edge2art: Edges to Artworks Translation with Conditional Generative Adversarial Networks

Rafael Gallardo-García¹,² Beatriz Beltrán-Martínez ¹,² Carmen Cerón¹
Darnes Vilariño Ayala ¹,²

Benemérita Universidad Autónoma de Puebla¹, Language & Knowledge Engineering Lab²

May 2, 2020
Index

1 Introduction
   What is a GAN?
   What is a cGAN?

2 Image to Image Translation

3 edge2art
   Dataset
   Preprocessing Artworks
   Architecture

4 Results

5 Conclusions
Introduction

What is a GAN?
What is a cGAN?

Image to Image Translation

edge2art

Dataset
Preprocessing Artworks
Architecture

Results

Conclusions
What is a GAN?

• Introduced by Ian Goodfellow in a paper\textsuperscript{1} in 2014.
• Generative Adversarial Networks belong to the set of generative models.
• A generative model is able to produce/generate new content.
• Training is done via an adversarial process.
• GANs are Unsupervised models.

\textsuperscript{1}Generative Adversarial Networks in arXiv:1406.2661
What is a GAN?

A GAN consists in two models trained simultaneously:

- A generative model $G$ captures the distribution of the training data.
- And a discriminative model $D$ that estimates the probability that a sample came from the training data rather than $G$.
- Both models (networks) are trying to optimize a different and opposite objective or loss function.
The illustrated architecture

**Figure:** GAN Architecture
How does the generator work?

**Figure**: Illustration of the notion of a generative model.
How the generative network works?

Figure: A Deconvolutional Network for Semantic Segmentation

Figure 3 shows a Convolutional-Deconvolutional Neural Network with sample-based discretization process.

What is a cGAN?

Conditional Generative Adversarial Networks are an extension of the GANs models and share some features.

- cGANs were first proposed by Mirza Mehdi in a paper\(^3\) in 2014.
- Has two components: a Generative model G and a Discriminative model D.
- Both are neural networks.
- In cGAN, G and D both receive additional conditioning input information (a vector of features).

\(^3\)Conditional Generative Adversarial Nets in arXiv:1411.1784
What is a cGAN?

The are one main difference between GANs and cGANs:

- GANs generator outputs depends on the "random noise", which is a vector of random numbers(latent vector).
- cGANs generator outputs depends on the conditioned input vector. Similar to a labeled or supervised training.

Figure: cGAN Architecture
Introduction
What is a GAN?
What is a cGAN?

Image to Image Translation

edge2art
Dataset
Preprocessing Artworks
Architecture

Results

Conclusions
Any process where an input object has a corresponding output object can be posed as a "translation" problem. This concept is applicable to several problems in image processing where an input image has a corresponding output image.

- A concept in English may be expressed in Spanish too.
- A scene may be rendered as an RGB image, a gradient field, an edge map or a semantic label map.
In 2016, Phillip Isola and his team published the paper "Image-to-Image Translation with Conditional Adversarial Networks".

- The paper presents a solution to the image translation problem using cGANs.
Introduction
What is a GAN?
What is a cGAN?

Image to Image Translation

edge2art
Dataset
Preprocessing Artworks
Architecture

Results

Conclusions
Dataset were obtained from Wikiart⁴. The complete dataset consists in 21 different painting styles and 79,622 artworks.

- Due to the high computational power required to train the cGAN, we select just 3 painting styles, see table 1.

<table>
<thead>
<tr>
<th>Art Style</th>
<th># Artwork Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ukiyo-e</td>
<td>1167</td>
</tr>
<tr>
<td>Rococo</td>
<td>2089</td>
</tr>
<tr>
<td>Fauvism</td>
<td>934</td>
</tr>
</tbody>
</table>

Table: Number of training examples per style

⁴https://www.wikiart.org/
Steps to preprocess the dataset

1. Resize every image in dataset to a 256x256px shape.
2. Extract edges with the Canny Edge Detection Algorithm.
3. Concatenate images in order to have both conditional input and target image in the same file.
Some training examples

Figure: Training examples: Ukiyo-e, Rococo and Fauvism, respectively.
cGAN for Image to Image Translation

Figure 6 illustrates a high-level architecture of edge2art.
Introduction

What is a GAN?
What is a cGAN?

Image to Image Translation

edge2art

Dataset
Preprocessing Artworks
Architecture

Results

Conclusions
Output examples: Rococo

Figure: Edge to Artwork translation examples for Rococo style.
Output examples: Ukiyo-e

Figure: Edge to Artwork translation examples for Ukiyo-e style.
Output examples: Fauvism

Figure: Edge to Artwork translation examples for Fauvism style.
Problem 2: How to measure similarity between images?

- There are several methods to measure similarity between images.
- Some are more complex to implement, than others.
- For this paper, we selected the Mean Squared Error and the Structural Similarity Index Measure.
Mean Squared Error measures the average of the squares of the errors: that is, the average squared difference between an estimated value (edge2art output) and the actual value (target artwork). A 0 score means two completely similar images and higher values means more different images. edge2art got an MSE score of 20,006 units.
SSIM Score

Structural Similarity Index is a method for predicting the perceived quality between digital images or pictures. It quantifies image quality degradation caused by processing. SSIM is almost used for measuring the similarity between two images. edge2art got an SSIM average score of 0.10766, where -1 is completely different and 1 is completely similar.
1 Introduction
   What is a GAN?
   What is a cGAN?

2 Image to Image Translation

3 edge2art
   Dataset
   Preprocessing Artworks
   Architecture

4 Results

5 Conclusions
Conclusions

• Clearly, the obtained scores are not good enough to consider the experiment as a success in MSE and SSIM.
• MSE and SSIM evaluates the similarity between the images by comparing the information of the pixels and not by evaluating the content or the context of the images.
• edge2art scores could be improved if the measuring methods are improved too.
Future work contemplates new measuring methods for edge2art.

Future evaluations includes content analysis and classification methods for images with feature extraction techniques.

Proposed feature extraction techniques: SIFT (Scale-Invariant Feature Transform) or HoG (Histogram of Oriented Gradients).

In order to measure similarity with a content approach.