edge2art: Edges to Artworks Translation with Conditional Generative Adversarial Networks

Rafael Gallardo-García[1], Beatriz Beltrán Martínez [1, 2] and Carmen Cerón Garnica[1]

rafael.gallardo@alumno.buap.mx
bbeltran@cs.buap.mx
academicaceron2016@gmail.com

[1] Faculty of Computer Science, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio, Boulevard 14 Sur, Puebla, 72592, México.
[2] Language and Knowledge Engineering Lab, Benemérita Universidad Autónoma de Puebla, Avenida San Claudio, Boulevard 14 Sur, Puebla, 72592, México.

Abstract. This paper presents an application of the pix2pix model [3], which presents a solution to the image to image translation problem by using cGANs. The main objective of our research consists in the evaluation of several artificial artworks that were generated by the cGAN, taking a scribble of edges as input. This evaluation covers different artistic movements and art styles such as Rococo, Ukiyo-e, Fauvism and Cubism. The set of the trained models of these different styles is called edge2art. Each art style was trained over more than 2000 artworks examples taken from the wikiart dataset used in ArtGAN [4]-[5]. The experiments consists in giving scribbles images to the cGAN, and depending on the selected style, the network will give an colored and stylized artwork as output. Comparison between the generated artworks and the target artworks are measured by Mean Squared Error and Structural Similarity Measure.

Keywords: edge2art, pix2pix, cGAN, image to image translation.

1 Introduction

Painting is the first art, is a mode of physical expression of the human creativity and feelings, which is found in every human societies and cultures. The great diversity that has existed throughout history and worldwide made this art one of the most important and beautiful modes of expression, painting comes in different kind of forms like drawing, naturalistic, abstraction, etc. Artworks can be non-figurative nor representable, and some artwork do not follow natural shapes, this characteristic made some styles like *Cubism* or *Abstract painting,* more difficult to analyze or understand. In the philosophy of art, aesthetic judgment is always applied to artwork based on one's sentiment and taste [3]. Beauty of painting depends on who is looking it.

On the other hand, any process where and input object has a corresponding output object can be posed as a "translation" problem. This concept in applicable to many problems in image processing, computer graphics, and computer vision tasks [3], where an input image has a corresponding output image. As well as a concept in English may be expressed in Spanish too, the same scene can be rendered as an RGB Image, a gradient field, and edge map or a semantic label map. Isola et al. [3] defined *image-to-image translation* in analogy to the automatic language translation as the task of translating one possible representation of a scene into another giving sufficient training data [3]. This paper explains the setting of the dataset and the cGAN in order to train models that are able to generate artworks taking just scribbles as an input, *edge2art* realized the translation of a simple scribble image to a more complex image that will be called artificial artwork.

2 Related Work

In [4] Wei Ren Tan et al. proposed an extension to the Generative Adversarial Networks to synthetically generate more challenging and complex images such as artwork that have abstract characteristics. ArtGAN innovates allowing back-propagation of the loss function with respect to the labels(which are randomly assigned to each generated images) to the generator from the discriminator [4]. ArtGAN is capable to create realistic artwork [5], see *Fig. 1*.

Radford et al. presented the Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks [1], and robbiebarrat develop art-DCGAN based on this paper, art-DCGAN is capable of generate art by using DCGAN theory. Outputs of this art-DCGAN are available on robbiebarrat's GitHub page, an example of these outputs is presented on *Fig. 2*.

Fig. 1. *ArtGAN examples of generated artwork.*

Fig. 2. *art-DCGAN examples of portrait generated artworks.*

The main difference between the state of the art works and the edge2art functioning consists in the conditional nature of the inputs. While ArtGAN and art-DCGAN[1] generate artworks without any input condition, edge2art generates artworks based on the input scribbles, which can be draw by users or extracted with edge detection algorithms as Canny Edge Detection, these conditional inputs allows to translate drawings and photos into artworks.

3 Structure of edge2art

3.1 Structure of the Wikiart dataset.

The total Wikiart dataset[2] consists in 21 different painting styles and 79,622 artwork examples, but due to the high computational power that is required to train the cGAN, we select just 4 painting styles, structure and number of examples for each selected style is described in the following table:

Table 1. *Number of training examples for each style.*

Art Style	Artwork examples
Ukiyo-e	1167
Rococo	2089
Fauvism	934

1 https://github.com/robbiebarrat/art-DCGAN

2 http://web.fsktm.um.edu.my/~cschan/source/ICIP2017/wikiart.zip

Cubism	2235

3.2 Pre-processing of the dataset

The wikiart dataset consists in artworks of different styles and different sizes, due to the cGAN structure, is impossible to process images without resizing it to 256x256 pixels. On the other hand, the edges of the artworks were automatically extracted with an implementation of the Canny Edge Detection algorithm [7], at this point, each artwork has it corresponding scribble or edge version. The cGAN requires a specific format of training examples, which is exemplified in the *Fig. 3,* these training examples are images with a size of 512x256 pixels, and contain two images, edge version and it corresponding artwork for training.

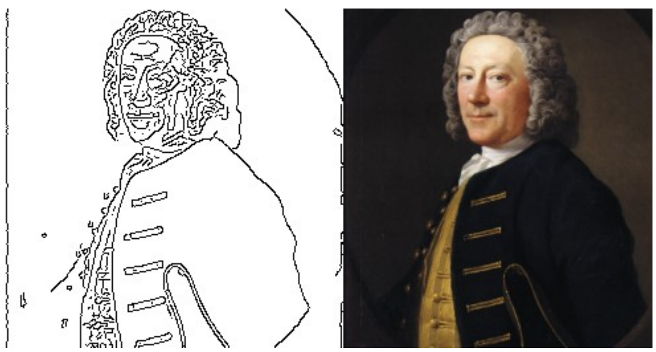

Fig. 3. *Rococo training example, edge version at the left and original artwork at the right.*

3.3 How pix2pix works

pix2pix uses a Conditional Generative Adversarial Network to learn a mapping from an input image to an output image, this is called Image-to-Image translation. This kind of networks are composed of two main pieces, the *Generator* which applies some transform to the input image to get the output image, and the *Discriminator* which compares the input image to an unknown image, this image could be a target image from the dataset or an output image from the *Generator,* and tries to guess if this was producer by the generator [3].

Fig. 4. *Input-Output example of the edge2art generator*

Discriminator is always looking at the generator's attempts and is always trying to learn to tell the difference between the images that the generator provides and the original target artworks.

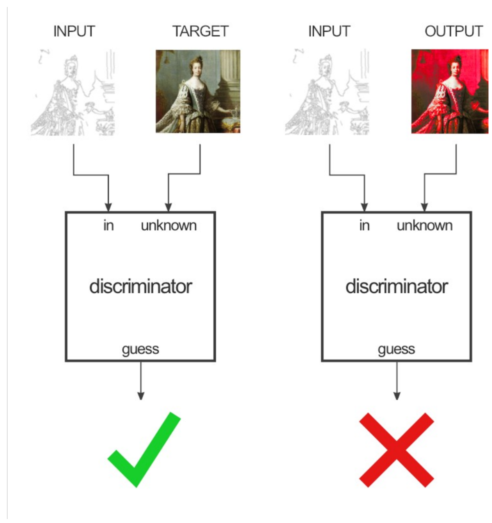

Fig. 5. *Example of how the generator-discriminator works*

One of the main points of the pix2pix paper is that the discriminator automatically provides a loss function for training the generator [3].

3.4 cGAN implementation for Image-to-Image translation problem.

The original implementation of the pix2pix model [3] were wrote in Torch, but nevertheless, there is a very good and functional TensorFlow implementation[3] of the pix2pix model developed by Christopher Hesse[4], edge2art uses this implementation to perform training and translation of images.

---

3 https://github.com/affinelayer/pix2pix-tensorflow
4 https://affinelayer.com/pix2pix/

4 Experimental Results

The experimental results were very satisfactory in visual terms. *Table 2* contains the summary of the training and testing results such as the number of parameters for each trained model, and the Mean Squared Error (smaller scores means that images are more similar) and Structural Similarity Measure (higher scores means that the images are more similar) scores obtained when comparing the outputs of the cGAN against the original artworks in order to quantify the accuracy of the models.

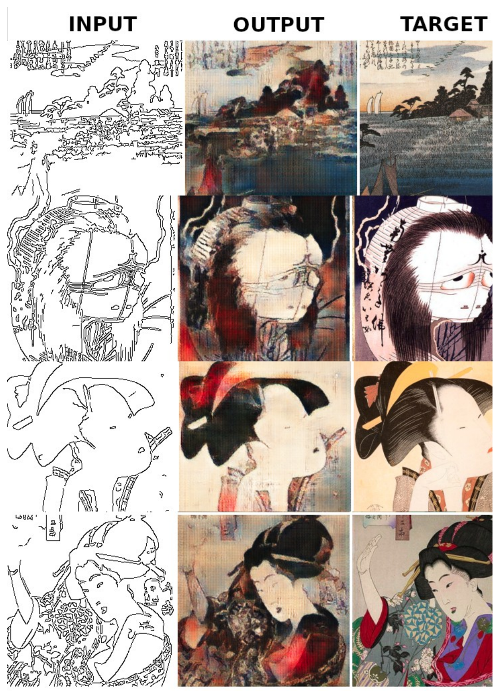

Fig. 6. *Ukiyo-e model examples of the cGAN outputs.*

Table 2. *edge2art experimental results.*

Model (Art style)	# of training examples	# of training parameters
Ukiyo-e	1167	57190084
Rococo	2089	57190084
Fauvism	934	57190084
Cubism	2235	57190084

The models obtained an MSE average of 20,006 units and an SSIM score of 0.10766, where -1 is completely different and 1 is completely similar. The full edge2art project is available on a GitHub repository, including code and pre-trained models[5].

5 Conclusions

As we can see in the experimental results, the obtained scores are not good, but in practice the output images and the target images looks very similar, this may be caused by an error in measurement, MSE and SSIM just evaluates the similarity between two images by comparing the information of the pixels and not by evaluating the content or context of the image. The similarity scores can be improved if the evaluation methods are improved too. Some proposals for future evaluations for edge2art include content analysis and classification methods with feature extraction techniques such as SIFT or HoG in order to evaluate the similarity of the images with a content approach.

If we evaluate the similarity of the images from a visual approach, many objects, colors and characters are easily distinguishable. More training examples and a better edge detection algorithm may improve the generated artworks.

Acknowledgments. Thanks to the Language and Knowledge Engineering Lab for providing the necessary computer processing power to carry out this research.

References

[1] A. Radford, L. Metz, S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434, 2015.

[2] M. Mirza and S. Osindero. Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784, 2014.

[3] P. Isola, J. Zhu, T. Zhou. Image-to-Image Translation with Conditional Adversarial Networks. arXiv preprint arXiv:1611.07004v3, 2016.

[4] W. Ren Tan, C. Seng Chan, H. Aguirre, K. Tanaka. ArtGAN: Artwork Synthesis with Conditional Categorical GANs. arXiv preprint arXiv:1702.03410v2, 2017.

[5] W. Ren Tan, C. Seng Chan, H. Aguirre, K. Tanaka. Improved ArtGAN for Conditional Synthesis of Natural Image and Artwork. arXiv preprint arXiv:1702.03410v2, 2018.

[6] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004.

[7] J. Canny. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986.

---

5 https://github.com/gallardorafael/edge2art