

Art Generation using Small Data Complacent Generative Adversarial Networks

Adit Whorra, Ayush Nath, Koushik Verma

Ashoka University, [adit.whorra, ayush.nath, koushik.verma]_asp19@ashoka.edu.in

Abstract – The new class of neural network architectures introduced in 2014, called generative adversarial networks (GANs), can be used to mimic any distribution in any domain: images, music, speech, prose etc. We propose a pipeline SDCGAN to produce paintings from extremely small datasets using a combination of VAE, DCGAN and image translation frameworks.

Index Terms – VAE, DCGAN, CycleGAN, pix2pix, painting.

1. INTRODUCTION

GANs [1] are deep neural net architectures comprised of two networks pitted against each other. The first network, called the *generator*, generates new data samples from noise, while the other network, the *discriminator*, classifies images as authentic or fake. As we feed the discriminator real data and data generated by the generator, the loss function ensures that after every iteration the generator learns how to *fool* the discriminator, and the discriminator to predict with greater accuracy.

In this process, the generator learns to produce data from the *distribution* (dataset) that was feeded to the discriminator, and these generated samples are considered as new instances because the generator is trained to produce them from noise (random numbers).

We decided to use GANs to generate paintings, something that are usually derived from humans' innate creative processes, and see how well they perform given a style of paintings. We chose a small dataset, to ensure smallest variation in style, of 78 landscape oil paintings by Albert Bierstadt, and trained a GAN pipeline on a dataset derived from these 78 images to produce new paintings.

2. PROBLEM DESCRIPTION

The abilities of GANs to generate new images from a given distribution has been thoroughly tested since they were introduced in 2014. Various techniques have allowed them to adapt to increasingly complex distributions.

However, the objects in a painting still comprise of a very complex distribution which the network might have difficulty replicating. Hence, we need to use techniques that will allow us to generate paintings with a simpler

distribution, the outputs of which can later be used to obtain images of higher complexity. Moreover, isolating paintings that are continuous in their composition of forms and consistent in their textures results in extremely small datasets which is detrimental to the functioning of GAN architectures given their large data dependency.

3. RELATED WORK

In the recent years, immense progress has been made in applying GANs to images in both generation of new images as well as Neural Style Transfer (NST), an optimization technique that manipulates digital images to adopt the appearance or visual style of another image.

Variational autoencoders (VAE) [2] have had great success mimicking simpler distributions like the MNIST [3] dataset but often produces blurry outputs when generating more complicated distributions like that of natural images. Moreover, the original GAN architecture also resulted in noisy and incomprehensible outputs while generating natural images.

Therefore, even though generating images has been explored extensively, generating natural images of the real world has not made much progress until recently. The conception of the DCGAN architecture [4] introduced a set of constraints on the architectural topology of Convolutional GANs that made them stable to train in most settings and was able to produce realistic images of bedrooms. Another contribution to natural image generation was made by ACGAN [5] by leveraging the labelled ImageNet [6] data and using conditional image synthesis to generate high resolution natural images.

Much of the work in art generation in deep learning has centered around style, and specifically the style of particular art pieces. [7] focused on differentiating the style of a piece of art from its content and then applying that style to other images using NST. However, this does not facilitate novel art generation and simply transfers artistic style to existing images. To create novel art, many recent attempts have built upon the architectures described above as they have proved to work well on natural images.

Our approach tries to combine the power of DCGAN to generate good quality images with the style transfer abilities of NST models like CycleGan [8] and pix2pix [9] and propose a novel pipeline for art generation.

3. ADOPTED APPROACH

With the problem description in mind, we first wanted to create a large dataset from a relatively easier distribution. To do so, we used a VAE to create 31,875 black and white images from the 78 Bierstadt paintings. Theoretically, all these images should follow a similar distribution as the original paintings. The VAE also causes some distortion in the outputs, which reduces the image complexity.

We used DCGAN to generate *new* paintings from this distribution. DCGAN was used because the presented architecture displayed good results in (indoor) scene generation and is proved to be a stable model otherwise.

After obtaining the generated paintings, we used CycleGAN, an unpaired image-to-image translation network to add natural colours to the paintings. The colourized outputs were still of low resolution which was caused by the VAE.

To add natural textures to the coloured paintings, we used pix2pix, a paired image-to-image translation network. We hypothesized that texturizing the distorted, coloured landscape images, using high resolution paintings, will give result in similar looking paintings.

4. IMPLEMENTATION DETAILS

I. Curating the Initial Dataset

We downloaded a dataset of 46,000 oil paintings by 1,400 artists. We narrowed down to 1,186 images of landscape paintings (Dataset A). To further narrow down on the distribution of the images, we selected 78 images (Dataset B) from Albert Bierstadt, who had several similar landscape paintings.



Figure 1. Seed data for our painting generation pipeline (Dataset A)

II. Expanding the Dataset

We run these 78 images through a simple VAE to generate more images while adhering to a similar distribution.

Input: Dataset B

Output: Dataset C (31,875 images) (size = 128 x 180)



Figure 2. Outputs of the VAE

III. Generating New Images

We trained an implementation of DCGAN on Dataset C to generate new images.

The generator of the DCGAN takes a noise vector as input (100 dimensional noise) and uses de-convolutional layers to upsample the noise to output a 128x128 grayscale image. The discriminator is continuously fed with these generated images as well as images from Dataset C and as the discriminator learns to distinguish between the generated and the real images, the generator learns to produce more realistic images.

Training specifications:

- input size = 128 x 180
- epochs = 20
- learning rate = 0.002
- batch size = 32
- output size = 128 x 128

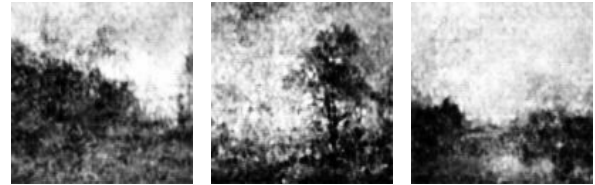


Figure 3. Outputs of the DCGAN

These new images can be considered as the first set of *generated* images in the current process.

IV. Colourizing the Images

Theoretically, DCGAN should generate images with the same distribution of objects in them with respect to Dataset C. Now, we want to add colour to these grayscale images, which corresponds with the colours found in Dataset A (subsumes Dataset B). We use CycleGAN for this style transfer process.

CycleGAN uses two adversarial networks, i.e. a pair of generator - discriminator networks. The model takes an input image from domain A which is fed to the first generator Gen_{AB} , which transforms the given input to an image in domain B. This new generated image is then fed to another generator Gen_{BA} , which converts it back into an

image from domain A. Meanwhile, the discriminators are trying to classify images in their respective domains A & B.

For our purposes, the training sets for CycleGAN were unpaired black & white, and coloured images from Dataset A.

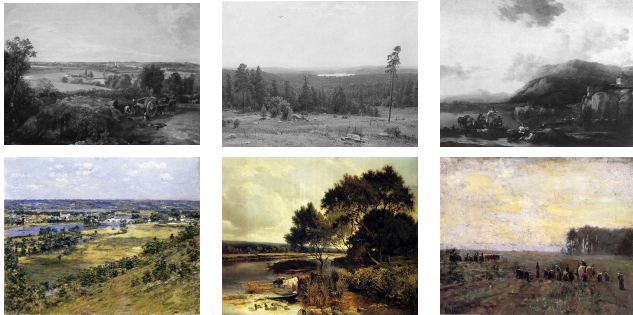


Figure 4. Training Set for Colourization

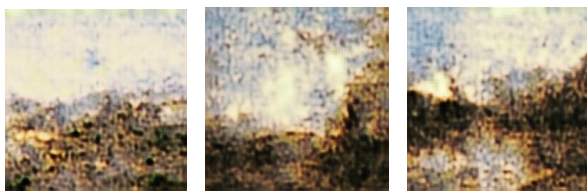


Figure 5. Outputs of CycleGAN colourization

V. Add Realistic Texture to these Images

Now that we have an *outline* of the features present in the image, we can apply textures from Dataset A to these images. Since these textures are more domain specific, compared to the process of colourization, we use paired image-to-image translation for stronger, domain specific feature learning. For this task, we use the pix2pix network.

The architecture of pix2pix is similar to that of CycleGAN, comprising of two adversarial networks, trying to generate images in both directions. The training sets for pix2pix were low resolution and high resolution images from Dataset A.

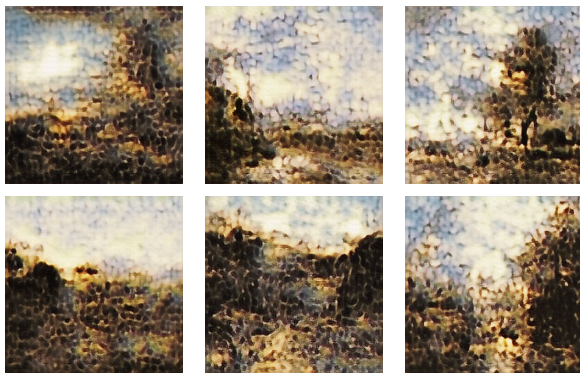


Figure 6. Outputs of pix2pix texturization

5. RESULTS OBTAINED

Our expectation from the generated landscape art was to produce landscapes that had a clear distinction between the sky and the earth and have realistic looking shrubbery and vegetation distributions. The following are the results produced by three stages of the pipeline: generation, colourization and texturization.



Figure 7. Result showing good shrubbery and vegetation and a clear distinction between the land and the sky

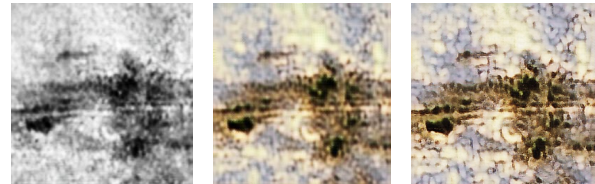


Figure 8. Result showing good representations of sky and water (with reflection) separated by shrubbery



Figure 9. Result showing good representation of the sky and clouds

The above results show that the landscape paintings generated do have a structure and a clear distinction between the land and the sky. However, the final output obtained from pix2pix has poor resolution and has lost a lot of sharpness compared to the original input. This could be because the images produced by the VAE (our curated dataset) themselves have lost some information as compared to the original images.

Also note that in our case, there is not much of a difference between the output of the CycleGan (colorized images) and the output of pix2pix. This could be due to the fact that points of noise in the images generated by the CycleGAN are probably attributed to incomplete forms when being processed by pix2pix.

CONCLUSION

We can conclude that the generated paintings have low resolution and lack sharpness due to the initial loss of

information when we generated images through the VAE. However, the pipeline proposed in this paper shows a lot of promise in generating paintings and generates realistic paintings despite having an initial dataset of only 78 images.

REFERENCES

- [1] I. J. Goodfellow, J. Pouget-Abadie et al, "Generative Adversarial Networks," NIPS '14, Vol 2, MIT Press, 2014, pp. 2672-2680.
- [2] D. P. Kingma, M. Welling, "Auto-Encoding Variational Bayes," ICLR '14, 2013.
- [3] Y. LeCun and C. Cortes. "MNIST handwritten digit database." (2010).
- [4] A. Radford, L. Metz, S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," ICLR '16, 2015.
- [5] A. Odena, C. Olah, J. Shlens, "Conditional Image Synthesis With Auxiliary Classifier GANs," ICML '17, Vol 70, 2017, pp. 2642-2651.
- [6] J. Deng, W. Dong, R. Socher , L. Li et al, "Imagenet: A large-scale hierarchical image database," 2009.
- [7] L. A. Gatys, A. S. Ecker, M. Bethge, "A Neural Algorithm of Artistic Style" 2015.
- [8] J. Y. Zhu, T. Park, et al, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," ICCV '17, 2017, pp. 2223-2232.