ART AND MACHINE LEARNING
CMU 2022 SPRING
PROJECT 2

Symbolic Art Generation With BigGAN On Small Dataset

Zhouyao Xie
Nikhil Yadala
Yifan He
Guannan Tang

DESCRIPTION

Our project intends to study symbolism in painting with the aid of machine learning techniques. We aim to generate symbolic paintings with BigGAN model fine-tuned on a small dataset (~100 images) of symbolic paintings. Our first attempt with training the model on a dataset of symbolic paintings yielded almost completely random yellow grids. We think that the complexity of our training target might exceed the capacity of our current model given the limited number of iterations (300) and small dataset size (100 images). To verify our hypothesis, we fed an abstract painting dataset (25 images), which contains simpler color schemes and compositions, into our training pipeline. After just 100 iterations, the model yielded better results compared with the previous one. This suggests that more iterations and/or more data are needed to improve our model's performance on symbolic painting generation.

CONCEPT

Symbolism is an influential movement in European literature and visual arts that began in the late nineteenth century. It initially developed as a french literary movement in the 1880s, gaining popular credence with the publication in 1886 of Jean Moréas' manifesto in Le Figaro. Symbolism aimed at the symbolic representation of absolute truth through language and metaphorical images, which required abstract thinking to engage from the creation of form to content, and it pursued spirituality, imagination, and dreams. In contrast, naturalism and realism focus on expressing reality in detail, so symbolism is largely a challenge against them. Symbolist painters used mythological and dreamlike imagery. The symbols they use are not those familiar from mainstream iconography, but strong references to personal, private, vague, and ambiguous. It is said that symbolism in painting is more of a philosophy than an actual artistic style.

When we appreciate symbolic paintings, it is not usually an intuitive experience for us to understand the meaning behind the works. Trying to understand the hidden meaning requires a certain amount of thought and speculation. However, in today's computer vision field, most research studies are more in the realm of naturalism and realism. They pursue a higher classification accuracy, greater resolution, etc. We believe that, from a certain point of view, computer-generated paintings need to be more complicated in order to be regarded as works of art, and are worthy of repeated appreciation and understanding by the audience. Exploring algorithms to generate symbolism paintings would be a worthwhile endeavor. We look forward to getting our algorithms to understand and extract features that are different from previous tasks, and generate some unexpected works.

Most state-of-the-art image generation models are deep neural models that rely on a huge image dataset to train, often on the order of tens of thousands of images. While the performance of these models are often extremely well, it leads to the question of whether the model has really learned the art style, or has it just learned to memorize the whole dataset. Besides, training a model on such gigantic datasets requires great computational resources, which is both costly and time consuming. As a result, we decided to research about alternative approaches and investigate methods to perform style-conditioned generation using a small symbolic art training dataset.

Transfer leaning, in which a model that has been pretrained on a large dataset is then transferred to a domain with sparse data annotation, natural became our research focus. A study by Noguchi and Harada called “Image Generation From Small Datasets via Batch Statistics Adaptation” [1] caught our attention during our literature research. They proposed a novel transfer learning approach in which only specific model parameters are updated while the rest of the parameters are fixed. By updating only the scale and shift parameters in the generator, they showed that they were able to significantly reduce the number of samples needed to fine tune the model. In fact, they showed that their method led to good performance by training on less than 100 images.

TECHNIQUE

Dataset

The dataset we use is from Wiki-Art: Visual Art Encyclopedia [2], which contains 15 genres of painting. We take the 3000 files under the symbolic-painting label as our dataset.

Pretrained Model

We used BigGAN-256 generator pre-trained on ImageNet as our pre-train model. BigGAN is a type of generative adversarial network that is able to generate high-fidelity, high-resolution images developed by DeepMind [3]. The BigGAN-256 model contains 55.9 million parameters and produces 256x256 pixels images. We obtained the BigGAN-256 model weights using the Chainer implementation [4].

Model Fine-tuning

We then used the batch statistics adaptation method proposed by Noguchi and Harada to fine-tune our model. We fine-tuned our model on two datasets, and the hyperparameters used are summarized in the tables below:

Dataset	Symbolic art data
Dataset size	100
Iteration	300
Batch size	5
Learning rate	0.03
Learning rate scale	0.2

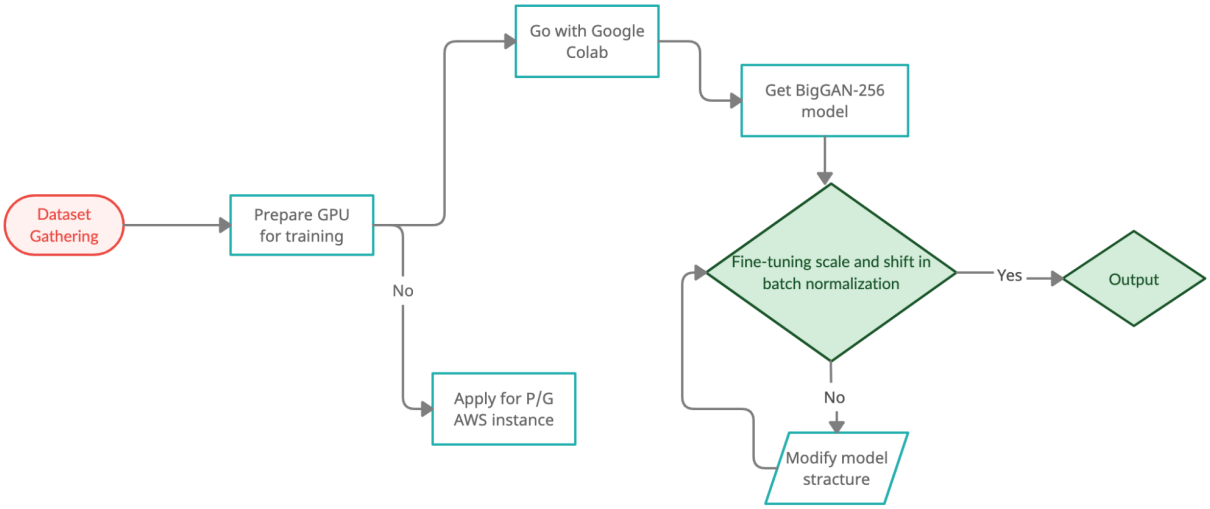
Dataset	Abstract art data
Dataset size	25
Iteration	100
Batch size	5
Learning rate	0.05
Learning rate scale	0.2

FURTHER EXPLORATIONS

Going forward, we would like to see how the generated model could be used to make AI based creative images through textual prompts. To do so, we will be using CLIP and test with various prompts to generate images (that would ideally be in the symbolic art space). CLIP is a multimodal language and vision model that is trained on several self supervised pretraining objectives to ground, align, learn whether a given piece of text is related to a given image. This is achieved by training through a contrastive loss where the training data is obtained from internet scraping of 1.28 Million images and their captions. The contrastive objective function constraints the similarity metric between positive pair (related image and caption) be higher than a negative pair (unrelated image and caption). There have been several successful demonstrations of text conditioned generative image models trained using CLIP as the evaluation metric to measure how well the generated image had aligned with the text.

We are interested in making the comparison with the original BIgGan that was trained on Image Net VS the finetuned BigGan performance w.r.t textual prompts indicating that the style and context (of symbolic art).

PROCESS

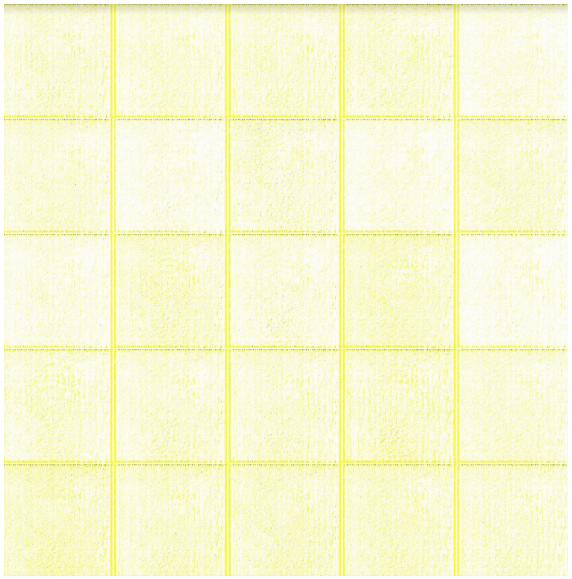


We started with looking for relevant datasets for the project. We narrowed down our scope to a dataset of paintings, datasets of various animals and symbolic art. After observing the quantity, diversity of the data, we proceeded to use the current dataset for further experimentation. We have had several issues with setting up the project on the AWS. Initially, it took ~2 weeks to get approvals for the P and G Instances with GPU availability. Moreover, the vcpu limits were not sufficient to start a VM with more than one GPU. This has resulted in very slow training, and we have been hitting various CUDA related issues with the codebase. The code base was primarily in chainer, however bigGan Model was generated from pytorch/tensorflow models. After necessary conversions, we went back to using Google colab notebook as that seemed to be the best option.

We have described our idea on using the CLIP in the above section. During the process, we also discussed the possibility of extending this approach to fine tune the model on 2 different datasets (belonging to two different domains or worlds) and generate the images from BigGan from text prompts which have the words and context of both the worlds. This way, even though the BigGAN generator might have never seen an image (during training process) in a space that is at the intersection of both the worlds, we can guide it to generate the image through CLIP, nevertheless.

We have explained why we picked our specific result in the below section in detail.

RESULT



This is one of the images the AI has generated after 200 epochs from a randomly sampled latent space. This is an interesting result, as the AI model has no understanding of the geometric concepts of straight lines or squares explicitly. Yet, it is surprising that the model has been able to generate perfectly straight lines with checkered squares. This shows that the model has learned various shapes and their alignment. BigGan generator interpolates the latent space “non linearly” and plots the image “pixel by pixel”. So, not only does the model need to have the global context (squares and straight line), but also co-ordinate with the other parts of the images (to produce coherent images of geometric shapes). This reminds us of the ancient greek’s notion of how geometric shapes correspond the perfect natural order.

This is a separate training result that we obtained by training our model with 25 images from the abstract art dataset. The training yielded seemingly rose-like human head portraits. The color composition here reminds us of the pop art from Andy Warhol. The rose red blends itself in a rather chaotic background, representing a rose tearing in chaos. We interpret this as the beauty in resilience. Moreover, as a comparison to our training result on symbolism art, our second training result confirms the point that the complexity of symbolism art exceeds the capacity of our current model at the current data size. For future training, we have to consider increasing the data size and also the training time so that our model can generate meaningful outputs.

REFLECTION

Reflecting on our experience working on this project, we think that we underestimated the time and effor it would take to set up the development environment at the start of the project. We ended up spending a much longer time debugging environment issues and trying to get the model running than actually training the model. Also, since we weren't able to get the training script running on AWS, we couldn't train as many iterations as we would like. In the next project, we will start earlier so that we could have more time to tackle with the set up issues.

CODE

For our model training and inference code, please refer to our Github repository:
https://github.com/ZhouyaoXie/small-dataset-image-generation

REFERENCE

[1] Noguchi, A., & Harada, T. (2019). Image generation from small datasets via batch statistics adaptation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 2750-2758).
[2] https://www.kaggle.com/ipythonx/wikiart-gangogh-creating-art-gan
[3] Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096 (2018).
[4] https://github.com/nogu-atsu/chainer-BIGGAN