From Drawings to Dreams: Generating Abstract Image Scenes From Pix2Pix Model

Ran Xu
UC Santa Cruz
Santa Cruz, United States
rxu3@ucsc.edu

Ella Dagan
UC Santa Cruz
Santa Cruz, United States
ella@ucsc.edu

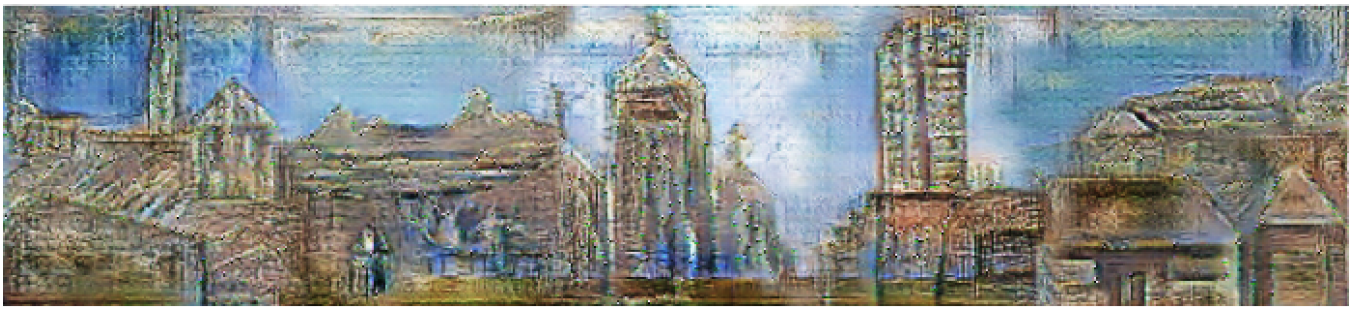

Figure 1: Impressionist-like artwork made from multiple generated images. The images were generated from recall drawings of different scenes, made by multiple people.

ABSTRACT

In this paper we describe our process and preliminary results of a Research-through-Design project in which we created abstract Impressionist-like artwork based on scene drawings as the input.

KEYWORDS

datasets, GAN, pix2pix, drawings, image generator

ACM Reference Format:
Ran Xu and Ella Dagan. 2019. From Drawings to Dreams: Generating Abstract Image Scenes From Pix2Pix Model. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

What if we could generate scene images that feel like a scene from a dream? Inspired by Hertzmann [6] inquiry into questioning if computers can create art, we wanted to explore the space of making art by "collaborating" with a machine model. We followed a Research-through-Design (RtD) approach [5, 8]. Gaver describes it as "a generative discipline," that is "able to create multiple new worlds rather than describing a single existing one." [5] We were interested in the space of image generation that would create results, unlike real-world images.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 BACKGROUND

Initially, we were inspired by a dataset of drawings from research that utilized images of scene categories to measure the recall capability of participants based on their drawings. To evaluate the recall drawing of the image scenes [4] used Amazon Mechanical Turk (AMT) by recruiting thousands of blind scorers to assess the "drawings diagnosticity, number of objects, extraneous objects, spatial detail, and object size within the drawings." [4] The main results of this work were that "Drawings from memory reveal the object and spatial information maintained within visual recall memory after a delay," and that "drawings from memory contain meaningful and quantifiable knowledge." [4]

The drawings from memory represented an accurate spatial map of the entire image, and we wondered if we could train a model based scene images to create spatially viable images generated from simple drawings that we would consider as new and interesting scenes. This made us think about human recall ability vs. machine recall ability and we wondered what we could generate from the recall drawings

dataset if we managed to create a machine model for those drawings and their scene images.

Figure 2: Example drawings from original recall study and the image scene as "ground truth".

3 METHOD

The memory recall study in [4] generated a dataset of 2682 scene drawings, approx. 30 drawings per image scene. We reached out to the authors of [4] to ask for the original scene images they used in their study. Thankfully we received access to that dataset and began hand matching drawings to their scene images, to create a "ground truth" per each drawing. Unfortunately, the original images of the drawings from the recall study were not formatted as squares so we had to reformat all (see Figure 2).

We utilized pix2pix model and the code presented in [7] to train a model based on the dataset we created from the drawing made in [4]. After around 7 hours of training to the dataset the results were interesting to a degree but not promising. The reason behind is that our dataset has varies of categories of scenes, as well as the dataset is not large enough. The model we used seems to work well when there are clear edges matching to the ground truth images, and with at least 400 images (drawings per one ground truth image). See initial results of images generated when testing in Figure 3.

We realized our training dataset was too small and since the model works best with clear edges we decided to try something else. We found another dataset [2] of scenes that had 24,335 images (see Figure 4).

4 RESULTS

Since the model words good with detecting edges and we did not have hand drawings, we used the canny edge detection

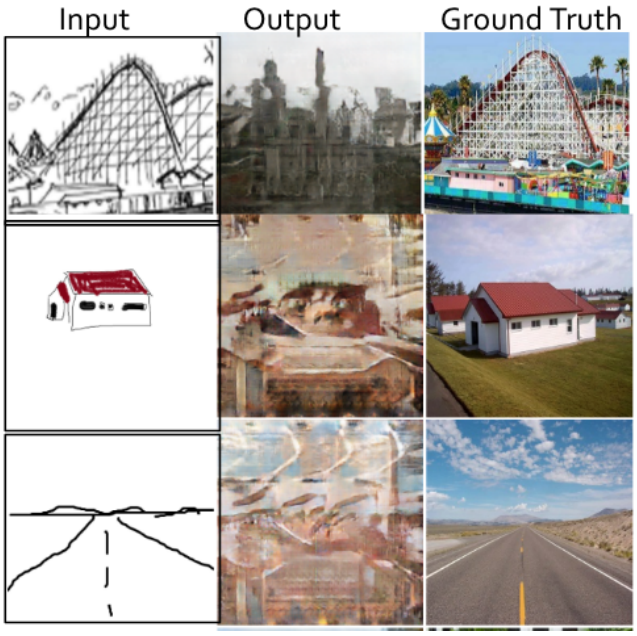

Figure 3: Testing results of initial training

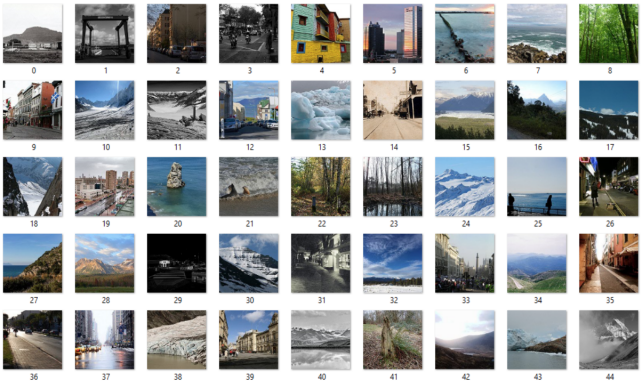

Figure 4: The new and larger dataset of scenes we trained on.

to get the line edges of the images and use it as "drawings" training data matching to the ground truth image scenes and then trained over 100 hours, up to epoch 150 (see Figure 5).

Figure 5: A sample of the training data we created by applying the canny edge detection.

We were curious about what results we will get as our dataset was highly variable. We tested by using a recall drawing matched to their ground truth scene image from our initial training dataset. We realized that the image was lacking depth, so we ran a comparison with the image edges as an input and the results were more realistic.

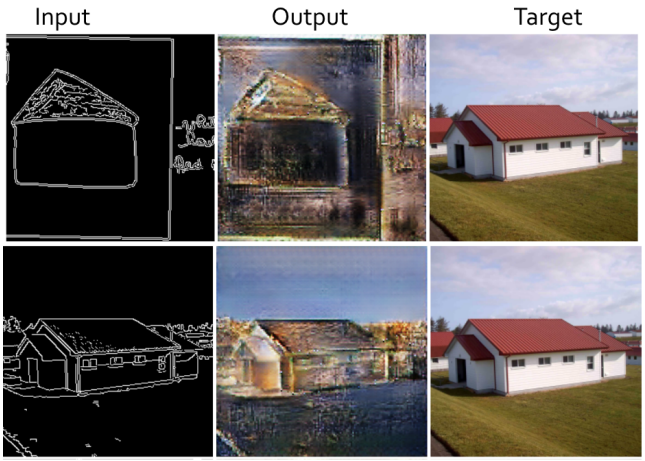

Figure 6: Top: recall drawing as input. Bottom: edge from scene image (target) as input.

The results had a lot of noise and were far from realistic, however, we could see potential in some of the images generated. Some looked as if could have been made by impressionist painters (see figure 7 8).

Figure 7: An example of the results for the different recall drawings of the same image scene of a lighthouse.

Figure 8: An example of the results for the different recall drawings of the same image scene of a tower.

We thought it could be interesting to generate a few of these "new scenes" as a base for a larger artwork piece that would combine the resulting scenes from the drawings. We created a joint artwork between multiple people's recall drawings of a house image scene, the machine model and us, interpreting and reworking the results (see figure 10).

We also tried another for combining the generated image scenes. Instead of combining the results from the recall drawings of the same image scene (e.g. an image of a house), we

Figure 9: An example of the results for the different recall drawings of the same image scene of a house.

Figure 10: An example of six results for the different recall drawings of the same image scene of a house collaged to create a new scene.

combined (collaged) results from various image scene recall drawings. The recall drawings, in this case, were of different categories of scene images such as a playground, sky-rise building, farm, mountain landscape etc. We appreciated the result of this step (see figure 1).

Figure 11: Different scene images.

Figure 12: The results of different scene images generated from various recall drawings.

From the image above, the colors in the result are blended, and the objects cannot be distinguished in the output. At this time, we are not completely pleased with our results since there is more noise than we aimed for. Similar with the project of from NVIDIA [3], they turns doodles into stunning, photorealistic landscapes. create new scenes, we would like to take a different approach and train on specific item category separately. We want to work towards creating a new scene by generating a few separated items from drawings, and then collage to create a new artwork. The idea is similar to what we presented above, only the training method is different. In order to get more realistic results from our training, we need to find the appropriate specific categories to train on, "part of scenes," such as trees, rivers, types of building, etc. Then the person who creates the drawing would be directed to draw from memory per each of the categories or

would be able to choose a few categories out of many. The model would generate separate images and will then collage these together by adding masks for each of them to create the new scene artwork.

We began testing this concept by using specific categories for training data, like categories of flowers (sunflower, daisy, dandelion, rose) from [1] datasets. Other categories we have trained on include fruit, trees, and river, etc. We are using this as a test on a relatively small number of images (up to 1000 images per category) to check results before moving forward on this concept. Testing results based on edges as input yield pretty good results. See Figure

Figure 13: Results from line edges of daisy images.

Use the different models, we would like to combine them into one scene. First example we have tried is showed in Figure 14. We found a drawing online (see Figure 14 left image), which contains houses, trees, and river. After we get the edges of the image and get the result from different networks we trained based on tree dataset (see Figure 14 middle image) and river (see Figure 14 right image).

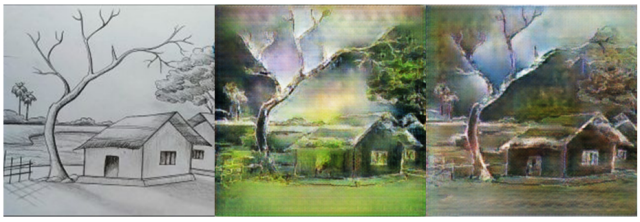

Figure 14: Different output of the same drawing from different nextwork.

In order to combine two images, we created the masks for different objects, see Figure 15 (left image). And we multiply the mask with the results in Figure 14, and get the results of Figure 14 (middle and right images). And finally we add the two images to get the result in Figure 16, where the edges of the tree are preserved and filled in more green color.

Figure 15: Mask and multiplication results.

Figure 16: Result.

Second example we want to combine different component into one image, shown in Figure 17 left column. We got images for tree and ocean and we get result from different network for trees and ocean, and we got the result as in Figure 17 right column. For combining the ocean and tree outputs, the mask we generate are shown in Figure 18 top two images, and by doing some multiplication and addition operations to the mask and the outputs, the final result is shown in Figure 18 right bottom image.

5 CONCLUSION AND FUTURE WORK

From the project, we found that pix2pix model works well for training on individual images, instead of for multiple categories in the training dataset. And the amount of images needed for training is small. If we want to generate some

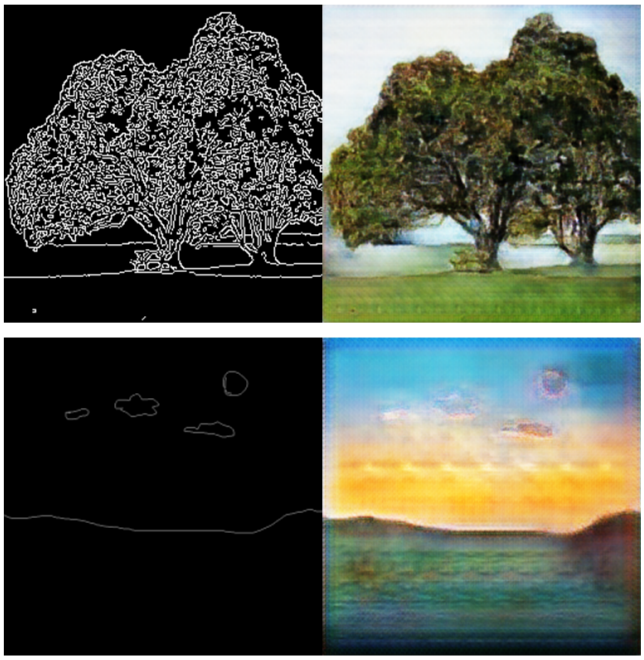

Figure 17: Ocean and tree generator output.

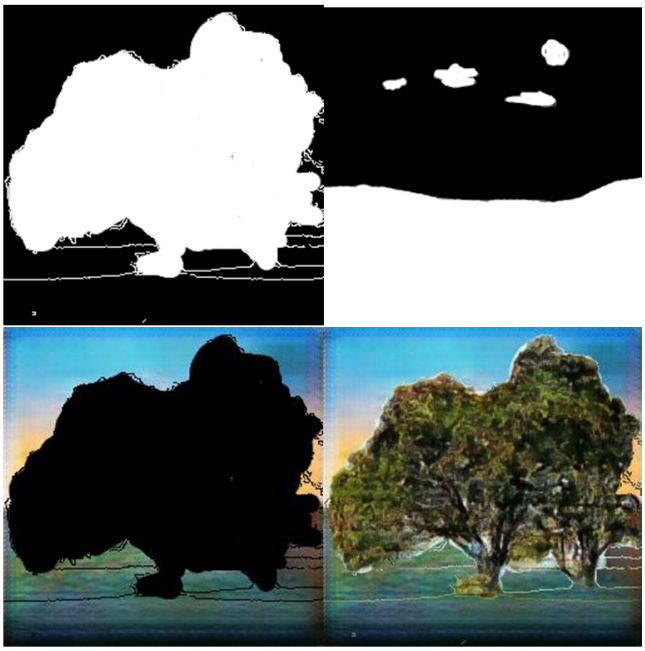

Figure 18: Applying mask and combining images.

complex scene with multiple objects in the scenes, we will need to have different network for different object. When combining several images into one, we need to create a mask for different object, and then add different images to one. Now the process of creating a mask and combing the images is manually done. In order to create scene image while people are drawing, dynamically creating the mask will be needed. Additionally, developing an user interface like what Nvidia did for their project will be necessary for making the whole process dynamically. So we can have a drawing board as well as some buttons that link to a network that have pretrained. Then the mask will be generated automatically while people are drawing. And then by combining all the images after the masks applied, user will get a scene based on different objects. On the other hand, we would like to explore other machine learning model to test whether they will have better result.

REFERENCES

[1] 2018. https://kaggle.com/alxmamaev/flowers-recognition.

[2] 2019. https://kaggle.com/nitishabharathi/scene-classification.

[3] 2019. https://blogs.nvidia.com/blog/2019/03/18/gaugan-photorealistic-landscapes-nvidia-research/?ncid=so-you-n1-78256.

[4] Wilma A Bainbridge, Elizabeth H Hall, and Chris I Baker. 2019. Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature communications* 10, 1 (2019), 5.

[5] William Gaver. 2012. What should we expect from research through design?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 937–946.

[6] Aaron Hertzmann. 2018. Can Computers Create Art?. In *Arts*, Vol. 7. Multidisciplinary Digital Publishing Institute, 18.

[7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.

[8] John Zimmerman, Erik Stolterman, and Jodi Forlizzi. 2010. An analysis and critique of Research through Design: towards a formalization of a research approach. In *proceedings of the 8th ACM conference on designing interactive systems*. ACM, 310–319.