Complete Sketches Using Deep Learning

Xiaolong Li, Shruti Phadke (Equal Contribution)

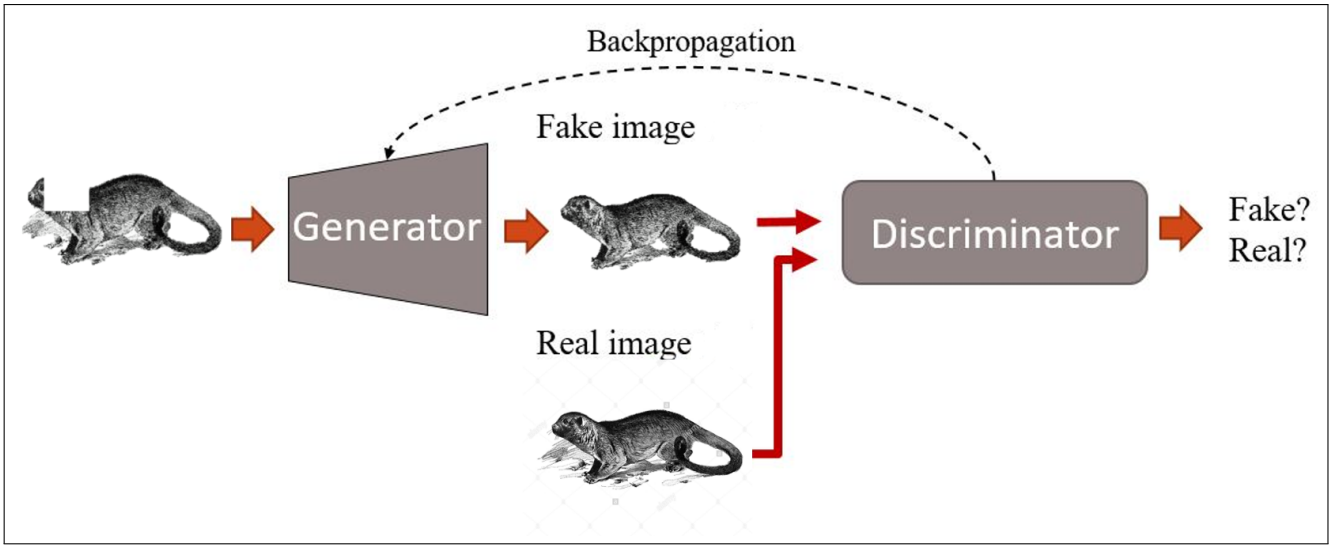

Figure 1. Simple GAN Illustration for Sketch Completion.

Abstract

In this paper we address the problem of sketch completion using Generative Adversarial Network (GAN). The problem of sketch completion is approached using pixel to pixel translation. Two types of datasets are compared for pixel to pixel translation, one with rough hand-drawn sketches and other with professionally drawn sketches. Finally, we try to utilize sketch category information as a supervision in pixel to pixel translation network for generating better sketch completion results

1. Introduction

Recent advances in deep learning have enabled problem solving in all aspects of technology. Making art, detecting sounds, creating text have become commonly explored applications of deep learning. One such application is image completion, specifically sketch completion. With increasing demand for mangas and web-toons, graphic artists are heavily depending on the technology. This paper attempts to explore ways to build easy-sketching application which can complete a sketch.

The problem of sketch completion in itself is vague and has many aspects like visual appeal, accuracy, processing time etcetera. For the scope of this paper we are focusing on the visual appeal and pixel to pixel completion of the sketch.

The input for our algorithm is an incomplete sketch, i.e, a sketch with a fixed portion occluded and the main goal is to produce a realistic completed sketch.

We strongly believe that the application like this will help the artists focus on the story content and eliminate sketching repetitive structures like background.

Even though, the idea of image completion has been explored in the deep learning context, no specific deep learning algorithm exists for sketch completion problem.

Pencil sketches are different for normal content rich images. The sketch completion problem should focus on completing the linings and shading texture without a color component. For this goal a highly specific neural network has to be trained. As mention above since no such architecture currently exists to solve this problem, in this paper we attempt to train a pixel to pixel network to deal with sketch completion application

In this paper we propose a category based sketch completion. The intuition behind this, a category of an object generally tells a lot of information regarding the texture, shape and peculiarities in an image. For example, ideally, a sketch of a sun is more likely to be completed with spikes on outside than inside with a category information. Using category information, we hypothesize that the sketch synthesized by an adversarial network will be more spatially and semantically realistic than without the category information.

We make the following four contributions.

- We created a occluded sketch dataset with category labels for hand-drawn sketches
- We created a occluded sketch dataset with category labels for professionally drawn sketches
- We trained pixel to pixel network to fit both above mention datasets and provide qualitative analysis of the result
- We modified pixel to pixel algorithm to incorporate category information as supervision with qualitative result analysis

2. Related Work

Even though, the idea of image completion has been explored in the deep learning context, no specific deep learning algorithm exists for the sketch completion problem. Thus the related work survey can be presented in following domains with respect to our main idea and design choices

Image Inpainting Inpaiting refers to completing missing portion in an image. In context of sketches, inpaiting can have large number of applications in sketch restoration and editing. Almost all of the recent efforts in image completion are based on Generative Adversarial Networks (GANs). GANs have two units, generator and discriminator. Generator attempts to generate realistic images and discriminator tries to classify image from real to fake. This process goes on as a zero sum game in which both generator and discriminator are expected to grow stronger as a result of the competition. The most recent work in image inpaiting comes from Yeh *et al*. [9] in which corrupted images are completed using a Deep Convolutional Generative Adversarial Network (DCGAN) using perceptual loss and contextual loss functions. In content aware image filling, Sauer *et al*. [8] analyze use of various neural networks for completing partially occluded images and prove that GANs give the best performance.

Other approaches include texture synthesis are presented in, Jetchev *et al*. [5], Efros *et al*. [1], Gatys *et al*. [3] with the core idea of texture synthesis using GANs.

The problem of sketch completion can also be viewed as image to image translation problem with input as patch removed image and the output as the complete image. Isola *et al*. [4] in their work attempt to achieve image translation with pixel to pixel training in different image domains. This multi modal approach covers grayscale to color, labels to image, sketch to image conversions. For testing the two datasets in out work, we use this pixel to pixel network and train it to fit our dataset.

Image Style Transfer According to the style transfer concept the image can have two components namely, content and style. The task of applying style of one image to the content of the other image is called image style transfer. For example, applying the feel and style of Van Gogh's Starry Night to any contemporary photograph can be called as image style transfer.

Auxiliary Classifier Image Synthesis A category of an object generally tells a lot of information regarding the texture, shape and peculiarities in an image. Embedding this category information in the image synthesis problem is discussed by Odena *et al*. [6] in the recent work. They propose a novel Auxiliary Classifier Generative Adversarial Network (ACGAN) which includes category information in image synthesis. We have tried to combine the technical aspect of this algorithm with that of the pixel to pixel network to solve the sketch completion problem using auxiliary classifier.

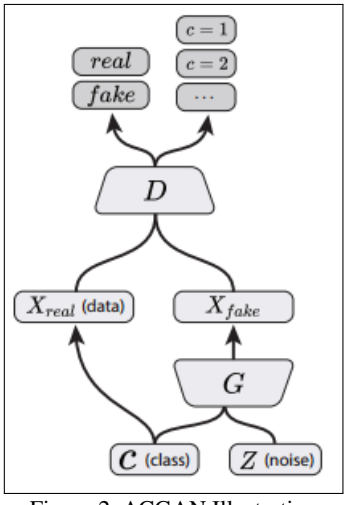

Figure 2. ACGAN Illustration

3. Overview

Generative adversarial nets were introduced as a novel way to train a generative model. Our neural network are based on the pix2pix cGANs Isola *et al*. [4], it treats the sketch piece as a special image query to get the completed sketch, the completed sketch will show clearly as an animal, airplane or other objects. A transformed U-net Ronneberge *et al*. [7], is chosen as the generator, it takes both category label and sketch piece as input to synthesize the fake image. The label is used as auxiliary information to guide generator make up sketches towards the primary shape of sketches in that category. Also, the discriminator should not only estimates the probability that a sample came from the training data rather than G, but also need to tell the probability class distribution of the synthesized image.

During the training, the generator is supposed to generator image that could fake the discriminator, and the discriminator will be trained with the ground truth image pair and fake image pair combined by the synthesized sketch and sketch query at the same time. Every generated sample has a corresponding class label c, the generator uses random noise z, input query sketch piece X, together with class label c to generate image $Y_{fake} = G(c,z,X)$, the discriminator gives both a probability distribution over sources under the condition of input sketch sample, and a probability over class labels, $P(S|Y,X)$, $P(C|Y,X) = D(Y,X)$, L_S stands for the log likelihood of the correct source,and L_C represents the log likelihood of the correct class.

$$L_S = \mathbb{E}[log\mathbb{P}(S = real|Y_{real},X)] + \mathbb{E}[log\mathbb{P}(S = fake|Y_{fake},X)] \tag{1}$$

$$L_C = \mathbb{E}[log\mathbb{P}(C = c|Y_{real},X)] + \mathbb{E}[log\mathbb{P}(C = c|Y_{fake},X)] \tag{2}$$

In our experiment, we also adopted the L_1 norm factor and add it to the generator loss,

$$L_{L1} = \mathbb{E}[\|Y_{real} - Y_{fake}\|_1]. \tag{3}$$

Discriminator is trained to maximize $L_S + L_C - L_{L1}$, and generator is trained to maximize $L_C - L_S$.

4. Method

4.1. Network Architecture

A encoder-decoder architecture comes with both convolution and deconvolution layers, which is suitable to function as the generator. After the last layer in the decoder, a convolution is applied to map to the number of output channels 3 in general,followed by a Tanh function. All ReLUs in the encoder are leaky, with slope 0.2, while ReLUs in the decoder are not leaky. The same strategy is applied for decoder. For decoder, after the last layer, a convolution is applied to map to a 1 dimensional output. Classes are embedded by simply adding one linear layer with 8 channels attached to the input of U-net structure.

4.2. Optimization Methods

During the training, we used the visdom tool developed by Facebook to track the training process, which will be able to display on-going input image, synthesized image and real image, the loss for L_1 and GAN will also be shown here. Different learning rate and training epochs are also tried to explore the best training set-up. To get the discriminator produce class information, a cross-entropy function is applied to the linear layer after softmax operation.

5. Experimental Results

5.1. Datasets

Two datasets were used to evaluate the pixel to pixel translation for sketch completion problem. The first dataset is a hand drawn sketch dataset, publicly available with the work of Eitx *et al*. [2]. This dataset was modified for the sketch completion as follows. Out of 250 original categories, 100 categories were randomly selected with 80 images each. Each image was white-patched in the input set out of five patch locations to maintain variation in occlusion. The training set contained 6500 images , the validation set contained 1500 images and the test set contained 1500 images. The test set and validation set has equal number of images from every category.

The second dataset used was collected by us using python web query. The same hundred categories were used as before with the same number of images, train, test and validation split and distribution. This dataset as opposed to the previous one, has finely drawn sketches with texture and fine lining. The judgment of fineness of a sketch had no numerical reasoning. This sketches were selected on pure visual appeal of a professionally drawn sketch.

5.2. Results

Please refer to Figure 5, Figure 6 and Figure 7.

5.3. Implementation details

For both datasets, we applied the same training strategy as suggested by pix2pix cGANs. The sketch query image and complete sketch images are aligned together in the same image. The learning rate is set as 0.0002, with 200 epochs for the whole datasets. In each epoch of training, the iteration number depends on the data pairs, about 1000 iterations for the first simple sketch dataset, and 3420 iterations for the second sketch dataset.

For comparison, the second dataset is also trained with the original pix2pix network as a baseline. All the parameters remain the same to better compare with each other.

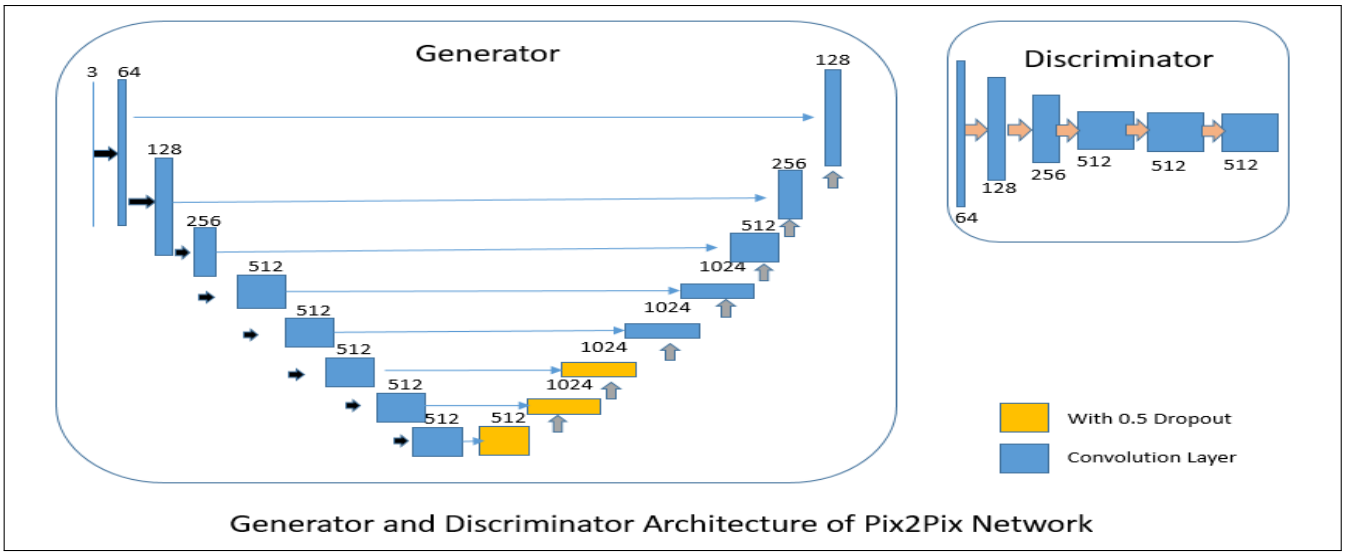

Figure 3. Pixel 2 Pixel Network Architecture

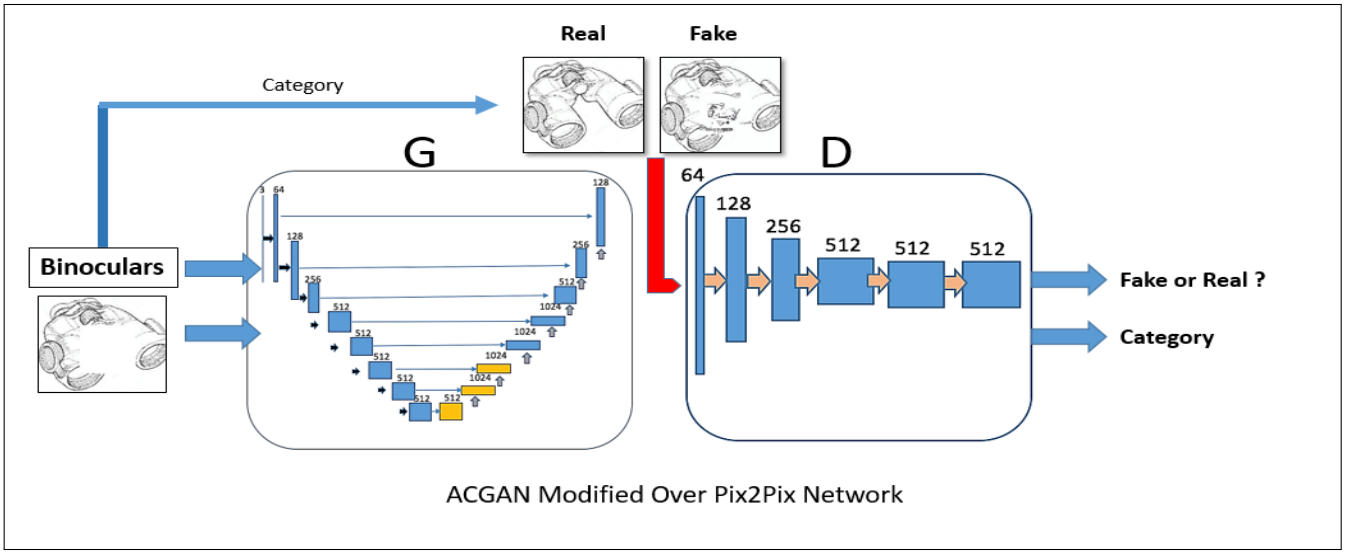

Figure 4. Auxiliary Classifier GAN Modified Over Pix2Pix Network

5.4. Qualitative evaluation and Ablation Study

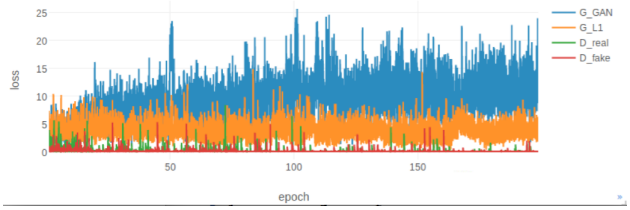

Figure 8. Loss change during training in 200 epochs with Pix2Pix on good sketch dataset

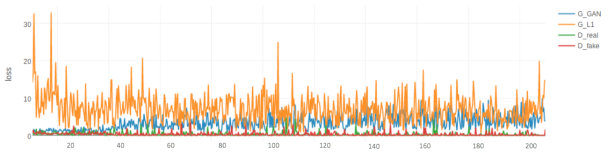

Figure 9. Loss change during training in 200 epochs with our modified ACGAN

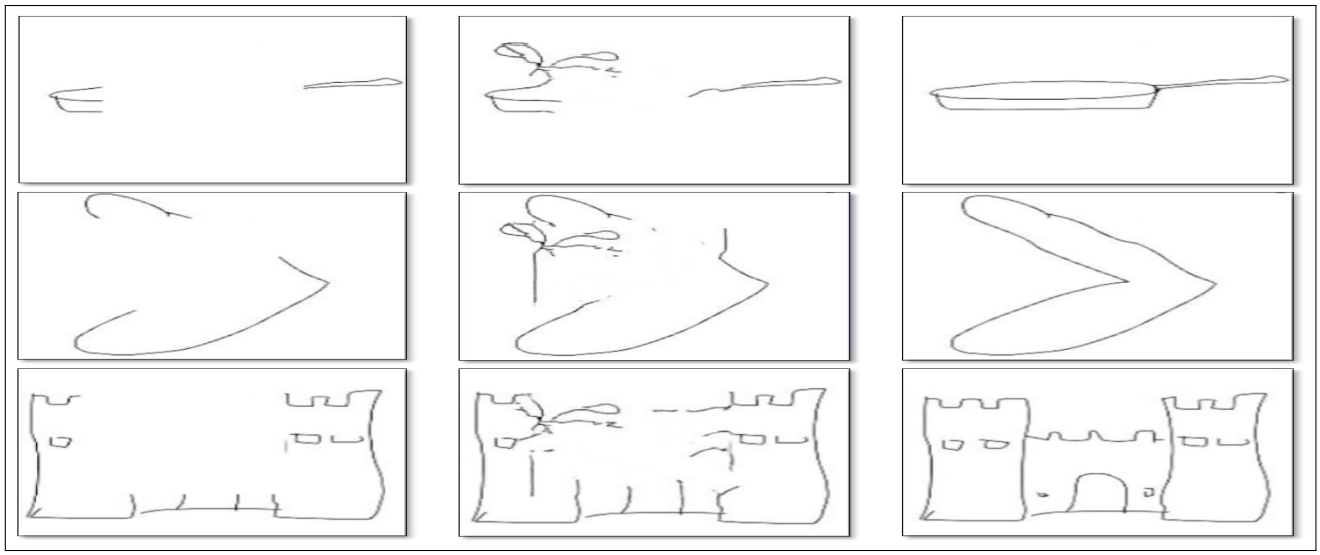

Figure 5. Results of a pixel to pixel training network for the hand drawn sketch dataset

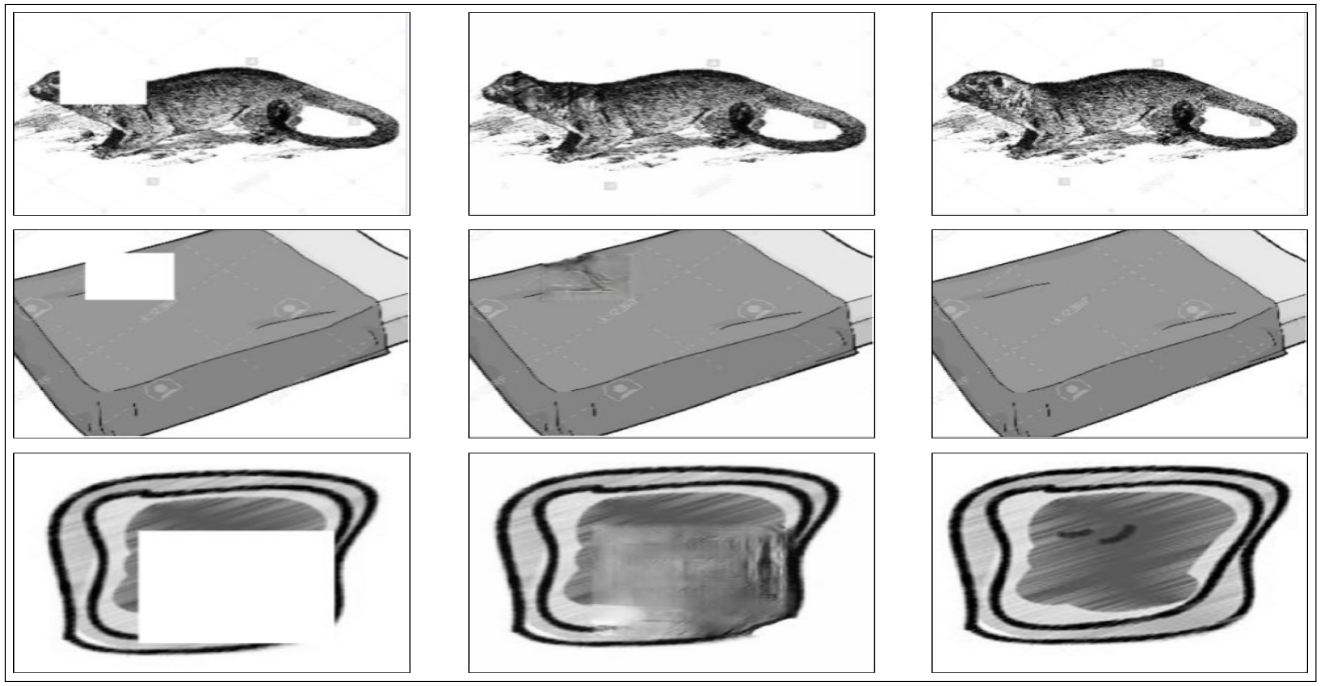

Figure 6. Results of a pixel to pixel training network for the professionally drawn sketch dataset

From the figures of losses for hand-drawn sketches, good sketches and good sketches with auxiliary classifier, following inferences can be made.

- The overall generator loss in the modified CGAN is much lesser compared to that of the simple Pix2Pix proving the hypothesis that using category information leads to better sketch completion.
- Average discriminator loss can be attributed to the fact that only a small part of the original image was varied in the fake image, making it hard to differentiate between a real and a fake image.
- Average L1 loss with Pix2Pix is lower than that of the modified ACGAN. Even though, it goes against the hypothesis of this paper, it can be explained with the images in two datasets. A hand drawn dataset has single line drawings, meaning even after taking a patch out, there isn't much variation in terms of L1 loss. As opposed to this, the good sketch dataset has textures,

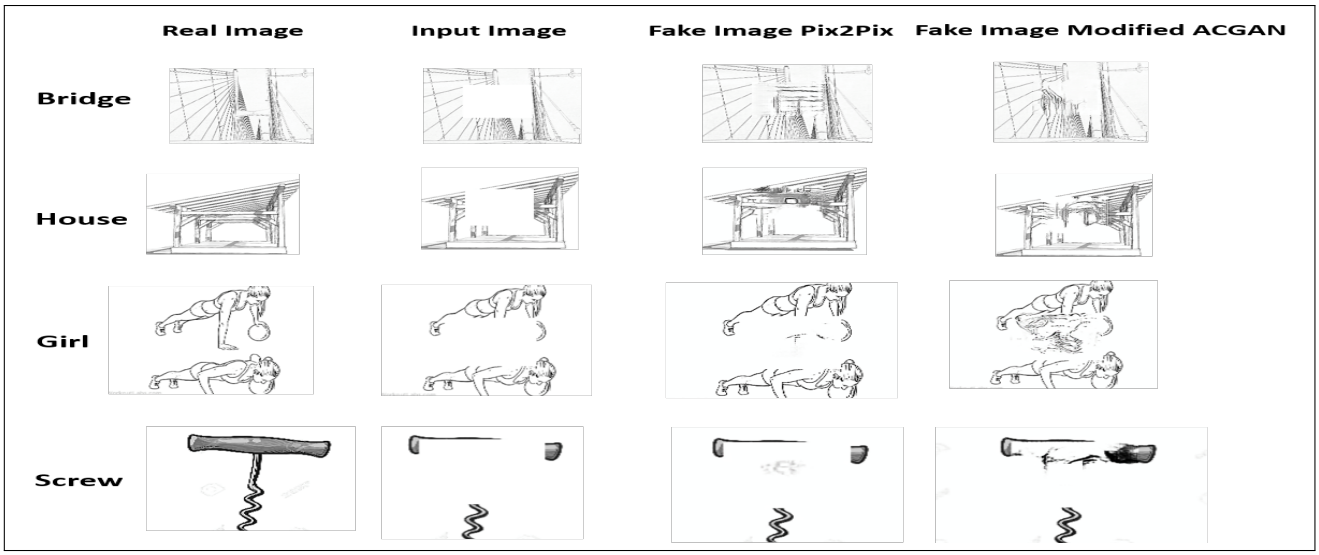

Figure 7. Results of a pixel to pixel training network for the professionally drawn sketch dataset

shading and rich sketch detail making the same occluded patch more susceptible to the L1 loss.

- Even though the overall generator loss for the modified ACGAN is lesser than Pix2Pix network, it it still large in absolute. Using a patch based L1 loss calculation where L1 loss will be calculated only for the occluded part can give better insight into the real effect for the category information.

5.5. Failure modes

Figure 5, Figure 6 and Figure 7 can be considered for the analyzing the failure modes of the method.

- Figure 5 indicated that Pix2Pix us really inefficient in completing line drawings. It includes a specific structure for every image learned from the previous examples
- Figure 6 indicated that the Pix2Pix algorithm is better that filling up patterns and textures than completing lines.
- Overall comparative display in Figure 7 indicated that modified ACGAN is filling up more line details in the Girl, House and Screw images but it still misses the important details.

6. Conclusions

In this work, we introduce our intuition on using auxiliary label information to expand the application of pix2pix cGANs. We have tested the performance of cGANs on both simple sketch completion and sophisticated sketch completion, which shows that cGANs could also be useful to complete the sketch completion task. Given good sketches, GANs could synthesize sketches with showing more details instead of random lines. We also demonstrate that adding label information, we could improve its performance to some degree. When the loss function become complex compared to the traditional GANs, it could be much easier to have oscillation patterns of loss during the training process.

Although results in our experiments suggest that auxiliary classifier could enhance the performance of cGANs, we still need to optimize the performance of our current network model and adjust the architecture, then better performance would be achieved.

References

[1] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346. ACM, 2001. 2

[2] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):44:1–44:10, 2012. 3

[3] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 262–270, 2015. 2

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arxiv*, 2016. 2, 3

[5] N. Jetchev, U. Bergmann, and R. Vollgraf. Texture synthesis with spatial generative adversarial networks. *arXiv preprint arXiv:1611.08207*, 2016. 2

[6] A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier gans. *arXiv preprint arXiv:1610.09585*, 2016. 2

[7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. 3

[8] C. Sauer, R. Kaplan, and A. Lin. Neural fill: Content aware image fill with generative adversarial neural networks. 2

[9] R. Yeh, C. Chen, T. Yian Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic Image Inpainting with Perceptual and Contextual Losses. *ArXiv e-prints*, July 2016. 2