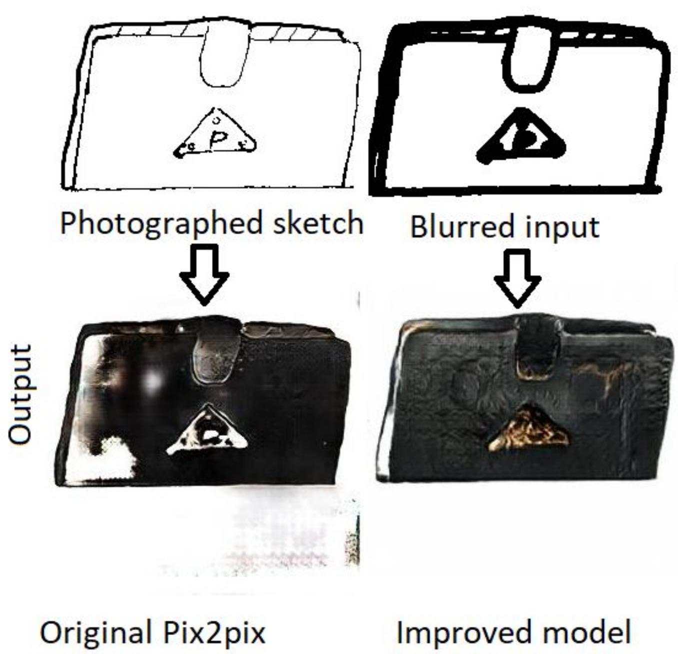
Generating Realistic Sketch-Based Images via GANs

Tianyue Zhao and Andrew Zhao

Henry M. Gunn High School

Abstract:

This paper discusses ways to improve currently existing image-generation programs with the help of conventional image processing tools. By utilizing a Generative Adversarial Network (or GAN), computers can be trained to generate images. In particular, the SBIS (Sketch-based Image Synthesis) program generates a fully colored image from a pseudo-sketch. However, there exists a large gap between computer-generated images and natural images. To bridge the gap, a combination of Gaussian blurring and binary thresholding was applied to the pseudo-sketches used for training. The resulting pseudo-sketches were used to retrain the GAN. As a result, the output of the Sketch-Based Synthesis program was substantially naturalized. One major discrepancy that was fixed was the amount of color on the sketches: many parts of the original GAN had white spots on the colored images, but every part of the image was colored in our improved version. However, our method only focused on handbags. Thus, a way forward is to apply our smoothing method to other objects such as tables and chairs and to observe if anything happens.



Generated images

Original wallet

Introduction:

Have you ever felt the tedium of image composing? If your answer is yes, you're not alone. Creating images using processing software such as Adobe Photoshop. often requires hours of highly repetitive work. The complicated user interface of image editing software worsens the situation. For example, to color an image in Adobe Photoshop, you need to first select one of the many paint brushes, then select one of nearly infinite colors. In addition, you need to constantly modify which tool and which color you are using, compounding the tediousness of the process. Thankfully there is an alternate method: Sketch-based image synthesis solves this problem by automatically generating images with a rough sketch as input. Current sketch-based image synthesis programs show potential as they can generate colored images out of rough sketches. However, these programs are limited in that they only output low-resolution and slightly unrealistic images unsuitable for real life use (Isola, 2017).

Most SBIS (Sketch-based image synthesis) programs utilize Generative Adversarial Networks (GAN for short) to generate images. While machine learning typically requires one neural network, a GAN consists of two separate neural networks. The first network, called the generator network, generates images from sample input and the other one, by the name of the discriminator network, tries to distinguish the examples from the generator network from real life examples (Goodfellow, 2014). In essence, the generator is attempting to trick the discriminator into judging that generated images are natural images For example, in a GAN that seeks to synthesize images of buildings, the generator network generates photos of buildings, and the discriminator tries to separate the synthesized images of buildings from images of buildings in real life. In a typical sketch-based image synthesis program, the "generator" network generates images with a rough sketch as input. Meanwhile, the "discriminator" tries to distinguish between the actual image that the rough sketch corresponds to and the image from the “generator” network. This unique structure bypasses the need for human feedback, which in turn allows GAN’s to generate acceptable images (Cush, 2017).

Literature Review:

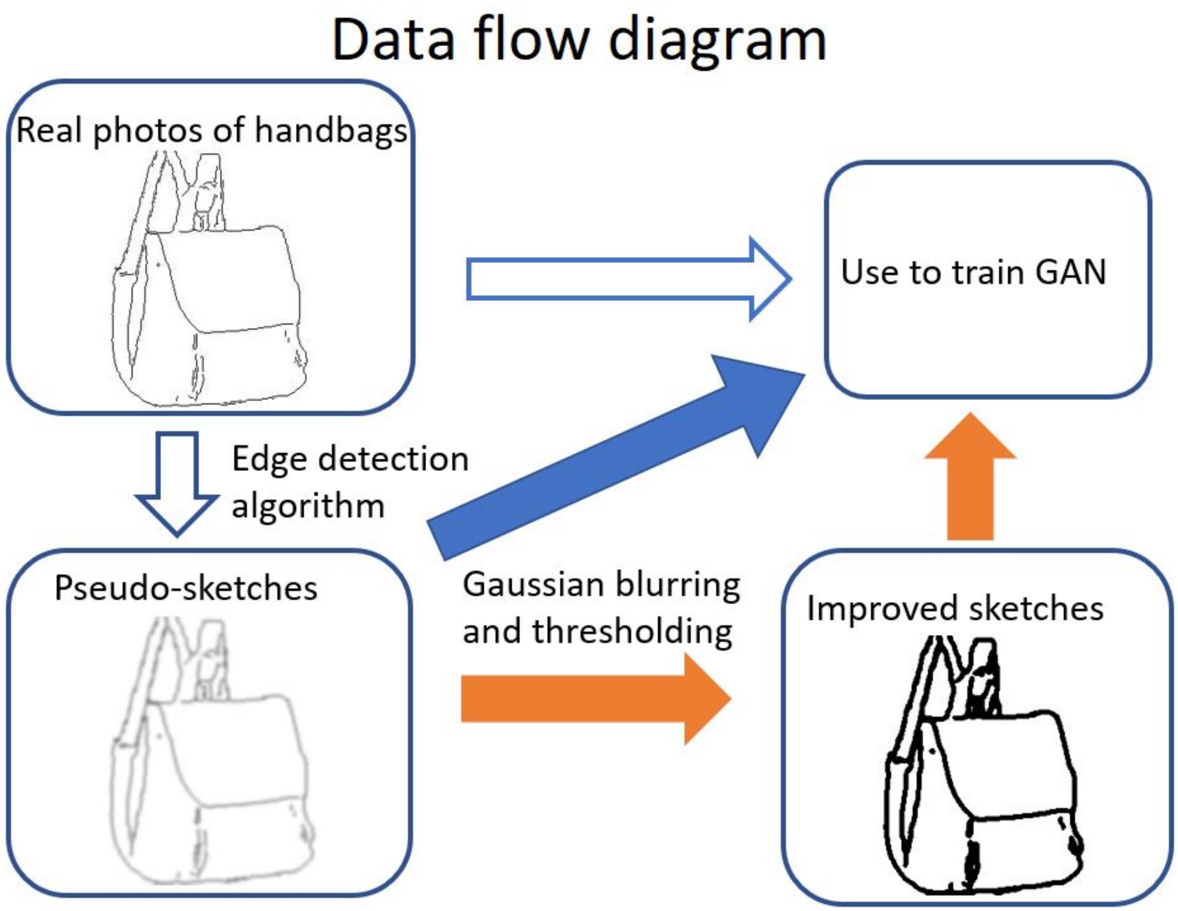
As GAN is a rather new technology, documentation on it is not as abundant as that of other machine learning algorithms. Usage of GAN’s to generate images was first proposed by Ian Goodfellow, who in his paper (2014) showed that GAN confers various benefits over other methods. This work was followed by several research papers that used GAN as a tool for various image-generation problems. For example, Ledig (2015) showed that GAN’s can be utilized to create super-high resolution images. Reed (2016) used a GAN to create a program that can generate images using imputed texts. Vondrick (2016) even showed that GAN can be used to design videos. Other developments in the image-generation industry include colorizing images using machine learning (Zhang, 2016), using GAN to simulate biological structures (Oskin, 2017) ,and creating computer-generated anime faces (Jin, 2017). These papers illustrate the potential of GAN’s when they are applied to image generation.

However, there is still a large gap between computer generated images and natural images. For example, an analysis on GAN generated images shows statistical differences between them and natural images (Zeng, 2017). Specifically, GAN generated images are not “smooth”: one can see a checkerboard like pattern if one zooms in close. By smoothing out the pixels when enhancing resolution, the second network in the experiment will hopefully correct this problem.

Additionally, training GAN’s is hard, due to mode collapse, which is when the generator tends to produce similar images without making any improvement, constantly occurring. However, a follow-up paper by

Ian Goodfellow (2016) discussed several ways to remedy this problem, including having the discriminator looking at multiple images instead of one at a time.

Research Methodologies:



Data flow diagram.

We seek to eliminate irregularities such as jagged edges in our pseudo-sketches: our goal is to mass-produce ‘smooth sketches’ that are less influenced by random noise. Our method must be fast, as we need to process 130000 pseudo-sketches. We settled on a combination of a blurring step followed by a thresholding step, as detailed below. To test the new pseudo-sketches out, we ran a GAN on a Tesla K40 GPU server, and used the code from pix2pix as our basis. Our main goal is to compare sketches from GAN’s without the improvement and GAN’s with them.

To form a control group, we first trained the GAN with the original photographs and the pseudo-sketches. The GAN utilized stochastic gradient descent and was trained on a Tesla K40 computing unit: it took 40 hours to converge. The results from this training serve as a control for our experiment.

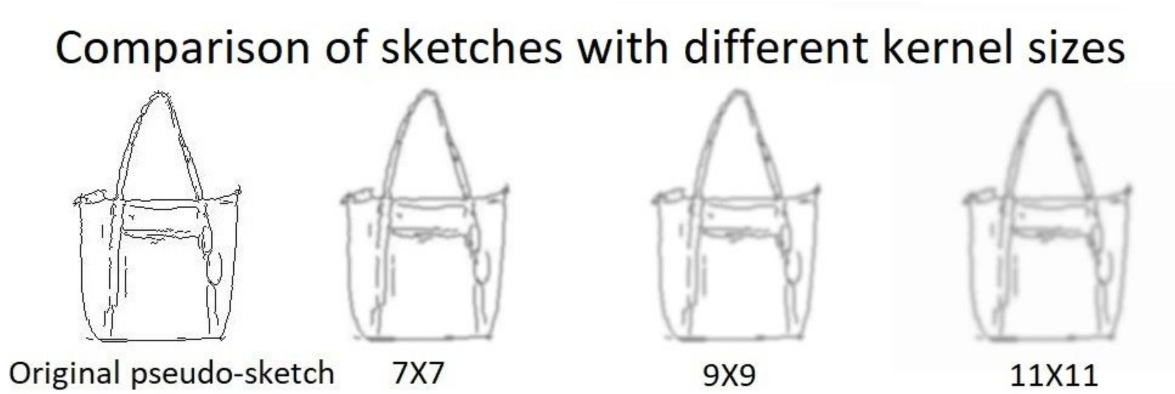


Figure X

The pseudo-sketches are first blurred by utilizing Gaussian blurring: a process where a Gaussian function is applied to the image in order to 'smooth' it. One crucial parameter in this process is the kernel size, where larger kernel sizes lead to blurrier pictures. After some experimentation, we found that a 9X9 pixel kernel is optimal. Figure X contains the blurred sketches generated for 3 kernel size selections. The 7x7 kernel failed to eliminate the unnecessary details, while the 11x11 kernel begins to blur important details in the pseudo-sketch. But the 9x9 kernel provides a middle road between the two, so it is the most optimal kernel size.

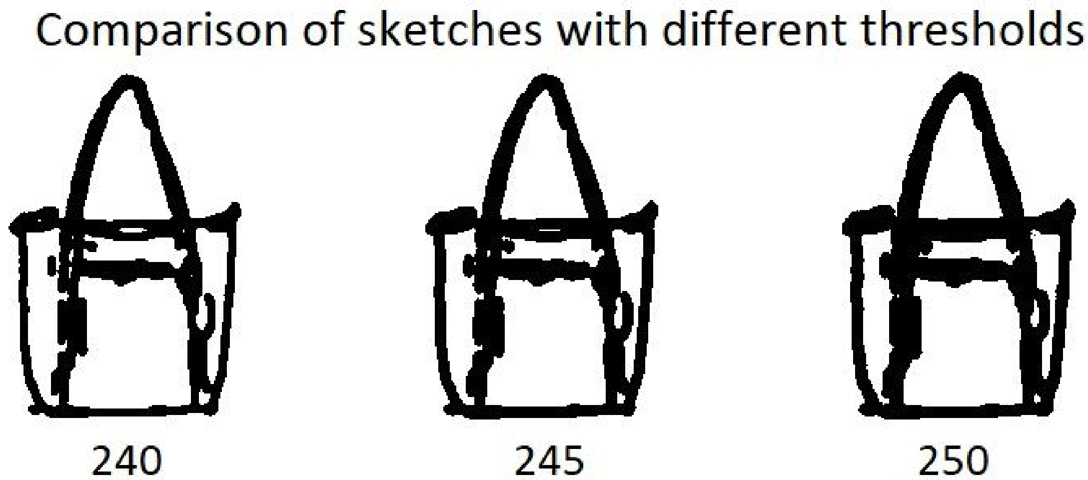


Figure Y

The resulting sketches from the blurring step are processed with thresholding. Each pixel has a color value between 0 and 255: the original pseudo-sketches had pixels with values of either 0 or 255. But Gaussian blurring can lead to pixels with values between 0 and 255, so we use thresholding to have a sketch that is black-white. Thresholding is the process of making all pixels with values greater than a threshold parameter black, and making all other pixels white. The threshold value is a crucial parameter: the higher the threshold, the more details will be erased. After some experimentation, we determined that 245 was the best. Figure Y shows that a threshold of 240 did not sufficiently erase the details, while a threshold of 245 erased too many details and resulted in excessively thick lines.

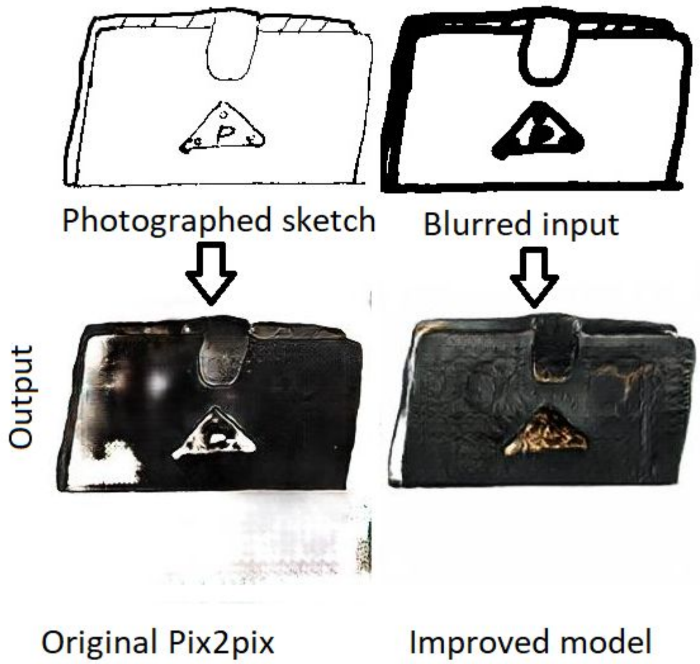
With our method, all 130000 sketches were processed in about 10 minutes. Then, we retrained the GAN with these improved images. For reference, the GAN was trained with stochastic gradient descent on a Tesla K40 computing unit, and took 40 hours to sufficiently converge.

Results and Findings:

Left: The pseudo-sketch after blurring and thresholding was applied.

Middle: The original image

Right: The original pseudo-sketch.



Left: The left side shows an image generated from the original. The right side features a image from the improved model.

Right: The original image.

Row 1: The original image

Row 2: The images from the retrained GAN,

Row 3: The images from the control GAN.

Discussion and Analysis

In general, we notice that the output of the modified SBIS (short for Sketch Based Image Synthesis) Neural Network on human-drawn sketches is substantially naturalized when compared to outputs from the original model. Especially noticeable is the disappearance of white space from the improved images. The original model left uncolored patches throughout the image, while virtually all of the improved images were completely colored. We have also improved the resilience of SBIS algorithms with respect to changing input sketches. For example, the performance of the model does not differ much between sketches drawn with thicker lines and thinner lines, while the original model only worked with sketches composed of thin lines. Overall, a comparison between the original model and the modified model shows that augmenting the SBIS Neural Network with conventional image processing methods creates highly successful results. Therefore, we have moved a step closer to making everyday use of SBIS programs more viable.

Conclusions, Implications, Next Steps

The generated images have been substantially improved by using augmented test data to train the SBIS network. However, some issues more dependent on the nature of the neural network still remain. For example, computation of the mean power spectrum for computer generated images fail to match distributions of ones generated using natural images. This result correlates with the appearance that the generated images appear to have lower definition when compared with natural images, even though they are both 256x256. In addition, many large scale images are 512x512, while current SBIS methods can only generate 256x256 images. A possible solution to this would be to run another neural network that is trained to upscale images. These suggestions can further enhance the quality of computer generated images.

We have come up with several potential next steps to research; they are listed below.

1. Generalized SBIS program

Right now, our program mainly focuses on handbags. We can see if our results hold if we replicate our experiment using different 'themes' of test data, such as if we used animal pictures.

2. Color-hints

The author can currently only control the shape, not the color, of the generated image. We can work towards enabling color-control by placing blots of the desired color in parts of the sketch.

3.Additional augmentation

We have shown that augmenting neural networks with conventional image processing algorithms can result in great improvements. More improvements could be made if we explored other potential augmentation methods.

References:

Aylien. (2016, August 24). An introduction to generative adversarial networks (with code in TensorFlow). Retrieved from http://blog.aylien.com/introduction-generative-adversarial-networks-code-tensorflow/

Georgia Institute of Technology. (n.d.). Sketch-based image synthesis. Retrieved October 29, 2017, from https://www.cc.gatech.edu/~hays/7476/projects/Cusuh/

Goodfellow, I., Salimans, T., Zaremba, W., Cheung, V., Radford, A., Chen, X., (2016) Improved Techniques for

training GANs. Retrieved from https://arxiv.org/pdf/1606.03498.pdf

Isola, P., Zhu, J., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. Retrieved from https://arxiv.org/abs/1611.07004

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. Retrieved from http://papers.nips.cc/paper/5423-generative-adversarial-nets

Güçlütürk, Y., Güçlü, U., Van Lier, R., & Van Gerven, M. A. (2016). Convolutional sketch inversion. Retrieved from https://arxiv.org/abs/1606.03073

Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H., & Fang, Z. (2017). Towards the automatic anime characters creation with generative adversarial networks. Retrieved from https://arxiv.org/abs/1708.05509

Karpathy, A. (n.d.). Generative adversarial networks. Retrieved from https://cs.stanford.edu/people/karpathy/gan/

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., … Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. Retrieved from https://arxiv.org/abs/1609.04802

Liu, Y., Qin, Z., Luo, Z., & Wang, H. (2017). Auto-painter: Cartoon image generation from sketch by using generative adversarial networks. Retrieved from https://arxiv.org/abs/1705.01908

Osokin, A., Chessel, A., Carazo Salas, R. E., & Vaggi, F. (2017). GANs for biological image synthesis. Retrieved from https://arxiv.org/abs/1708.04692

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. Retrieved from https://arxiv.org/abs/1605.05396

Sangkloy, P., Lu, J., Fang, C., Yu, F., & Hays, J. (2016). Scribbler: Controlling deep image synthesis with sketch and color. Retrieved from https://arxiv.org/abs/1612.00835

Vondrick, C., Pirsiavash, H., & Torralba, A. (2016). Generative videos with scene dynamics. Retrieved from http://papers.nips.cc/paper/6194-generating-videos-with-scene-dynamics

Zeng, Y., Lu, H., & Borji, A. (n.d.). Statistics of deep generated images. Retrieved from https://arxiv.org/abs/1708.02688

Zhang, R., Isola, P., & Efros, A. (2016). Colorful image colorization. Retrieved from https://arxiv.org/abs/1603.08511

PLEASE HIGHLIGHT WHERE YOU THINK YOU BELONG ON THE RUBRIC!

Milestone #6 - The Final Paper Rubric (100 pts)

Required components:
(5 pts each/20 pts total)

☐ Submitted on Schoology

☐ Draft submitted on/before due date

☐ Citations are in APA or MLA parenthetical format

☐ All sections are complete (Title, Intro, Methods, Results, Conclusion, Implications, and Citations)

If any above boxes are unchecked, re-submit by: ________________________

Content Area	Performance Levels (80 pts)		
Title (5 points)	Research question is clearly and effectively conveyed by the title of the poster. Appropriate title and all contributors listed 5		Research question is vague and/or confusing as conveyed by the title of the poster 1
Abstract (10 points)	The abstract is a succinct, single-paragraph summary of your paper's purpose, main points, methods, findings, and conclusions. Length is 150-250 words 10	The abstract is not 150-250 words 8	The abstract does not address purpose, main points, methods, findings, and conclusions. 6
Introduction (10 points)	Rationale/justification for the study is clear; Relevant previous work is thoroughly reviewed; Gap in knowledge/exploration identified; References are cited appropriately; includes graphic that is effective 10	References are not cited appropriately and/or does not include graphic that is effective 8	Rationale/justification for the study is not clear, relevant previous work is not thoroughly reviewed and/or gap in knowledge/exploration is not identified. 6
Methodologies (10 points)	Clear description of inquiry approach, methods and procedure; includes graphic (i.e. timeline or execution plan) that is effective 10	Does not include graphic that is effective 8	Description of inquiry approach, methods and procedure are not clear; no graphic included 6
Data, Results and Findings (10 points)	Findings are presented clearly and accurately; Figures/tables/graphs used appropriately and effectively to present data 10	Figures/tables/graphs are not used appropriately and/or effectively to present data 8	Findings are not presented clearly and/or accurately 6
Discussion, Analysis, Evaluation (10 points)	Clearly addresses how findings relate to research question; Conclusions are sufficiently supported by results 10	Attempts to address how findings relate to research question but connection is vague or unclear 8	Does not address how findings relate to research question; conclusions are not supported by results 6
Conclusions, Implications, Next Steps (10 points)	Results are placed into broader framework; Importance of findings is addressed; future directions are discussed 10	Importance of findings is not addressed OR future directions are not discussed 8	Importance of findings is not addressed; future directions are not discussed 6
Apply Conventions Grammar & Mechanics	Word choice and syntax enhances communication through variety, emphasis, and precision. No grammatical, spelling, and punctuation errors.	Word choice and syntax adheres to established conventions of grammar, punctuation, spelling, usage and mechanics. Some errors, but they do not interfere with the author's meaning.	Use of grammar, punctuation, spelling, style and mechanics convey the student's ideas; however, errors interfere with communication and/or credibility.

(10 points)	10	8	6