

**CSCE 624**  
**Sketch Recognition Proposal**  
**SketchShop - Photo Editing with Sketch**  
**Li-Hen Chen - 928003907**

## **1 Summary**

Photo editing exists everywhere in our daily life from the advertisements to robust conference papers. We will design a website system called SketchShop which allows users to add some vivid icons on her photos by just a few simple sketches. We adopt GAN to do style transfer on the sketch from users and mask the output on the photos uploaded by users.

## **2 Motivation**

We all know about how useful photo editing technique is. Besides, photo editing has a lot of application and one of the most famous is Photoshop. For example, by using Photoshop, we can easily change the cloth for a person or make a person wear glasses which gives us a chance to have an initial idea how one looks in different clothes. This triggers an idea that when we would like to modify the photo to make someone dress differently, we can simply sketch new wearing such as a hat or glasses on one's photo and apply deep learning technique to turn the sketch into a vivid wearing.

## **3 Prior Work**

In order to accomplish the task we come up with style transfer method. Indeed, there are plenty of researches and papers focusing this area. One of the most impressive paper is [1], I still remembered vividly this paper was proposed during my sophomore year. This paper adopted a simple CNN with a loss function which combines the style and content images information to accomplish the image style transfer. To understand style transfer domain deeply, [2] provides a great review of style transfer from the traditional style transfer without neural network to state-of-the-art neural style transfer. There are two papers [3][4] which I think are feasible in our application. The first one [3] proposed a method, pix2pix using generative adversarial network (GAN) to accomplish the style transfer. However, the biggest drawback of this method is that it requires paired images to train the GAN. As a result, with the drawback of pix2pix comes [4]. The paper [4] is able to utilize unpaired images to generate style transfer images which is called CycleGan. Although these two methods are exceptional, the most serious problem is that generative adversarial network is difficult to train. Thus, we decide to deal with the

issue by adding sketch features in our discriminator.

There are two novelties in our project. As I mentioned, due to the GAN is difficult to train, the first novelty is that we want to add sketch features in our GAN discriminator to make the model converge more easily. The second novelty is our idea. As a saying goes “Technology stems from humanity”. Undoubtedly, the papers of style transfer are springing up on major conferences every year. But these papers are focusing on the algorithm and neural network, we wish we could apply these brilliant ideas to real world application which has not been seen before.

## **4 Methodology**

We finally determine to use GAN to build our SketchShop system. GANs are neural networks that generate synthetic data given certain input data. The two important models in GAN are the discriminative model and the generative model. The discriminative model operates like a normal binary classifier that's able to classify images into different categories. It determines whether an image is real and from a given dataset or is artificially generated. The discriminative model tries to predict certain classes given certain features. The generative model tries to predict features given classes. This involves determining the probability of a feature given a class. During the training process, weights and biases are adjusted through backpropagation until the discriminator learns to distinguish real images of shoes from fake images. The generator gets feedback from the discriminator and uses it to produce images that are more ‘real’. The discriminator network is a convolutional neural network that classifies the images as either fake or real. The generator produces new images through a de-convolutional neural network. We will embed the sketch features to the discriminative model to make it have higher accuracy on distinguishing real images from fake images.

Beside the model, this is a real application which means it needs a graphical user interface. We will design a web interface to let the users upload their photos and then use a blank paper layer where users can sketch interactively on this layer. Finally, the system will feed the sketch into the model, obtain a real wearing stuff and mask the image with the photo uploaded by the users. The reason we decide to use website as UI not use local apps is originated from the concept “Software as a service”. Because we wish to promote the system to everyone and a browser and an internet connection is all that's usually required to access a SaaS application, which can therefore be made available on a wide range of desktop and mobile devices.

## **5 Evaluation**

Certainly, it is impossible to evaluate the work by numbers or statistical data. The first step is we will evaluate our work visually to see whether the images we generate are reasonable or not. If the images generated by the model are reasonable enough, we will do user experience research. Each time a user completes the whole process of the system, we will collect their feedbacks as our evaluation resources.

## **6 Reference**

- [1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
- [2] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2019). Neural style transfer: A review. IEEE transactions on visualization and computer graphics.
- [3] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
- [4] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
- [5] Rubine, D. (1992, May). Combining gestures and direct manipulation. In Chi (Vol. 92, pp. 659-660).