CS2470 DL Day: Smart Ink

12/12/2022

Tengfei Jiang, Yuanfeng Li, Zhangyi Shen

Scan this code to see our output video

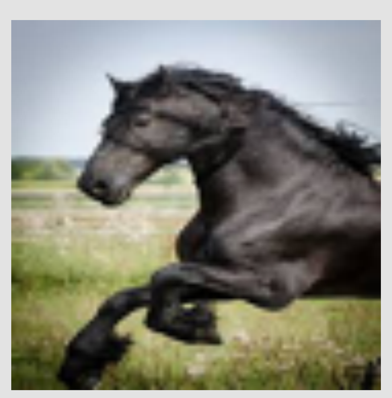
Introduction

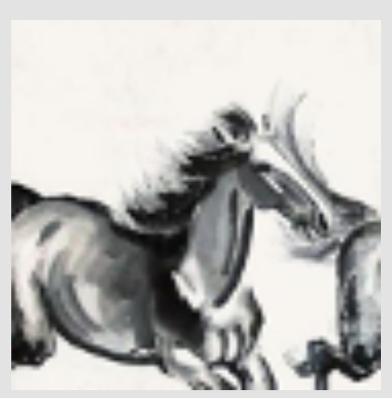
For the purpose of converting real images into other styles, such as cartoons, neural style transfer has been extensively employed. We observed that most of them concentrate on western styles but we opted to explore the Chinese ink style, which is difficult because it resembles freehand brushwork. In our project, we applied GAN and additional constraints to make neural style transfer compatible with Chinese painting techniques like blank-leaving, brushstroke, or ink washing.

Preprocessing

Our horse dataset consists of 718 horses and 912 images of paintings by the famous horse painting expert Xu Beihong.

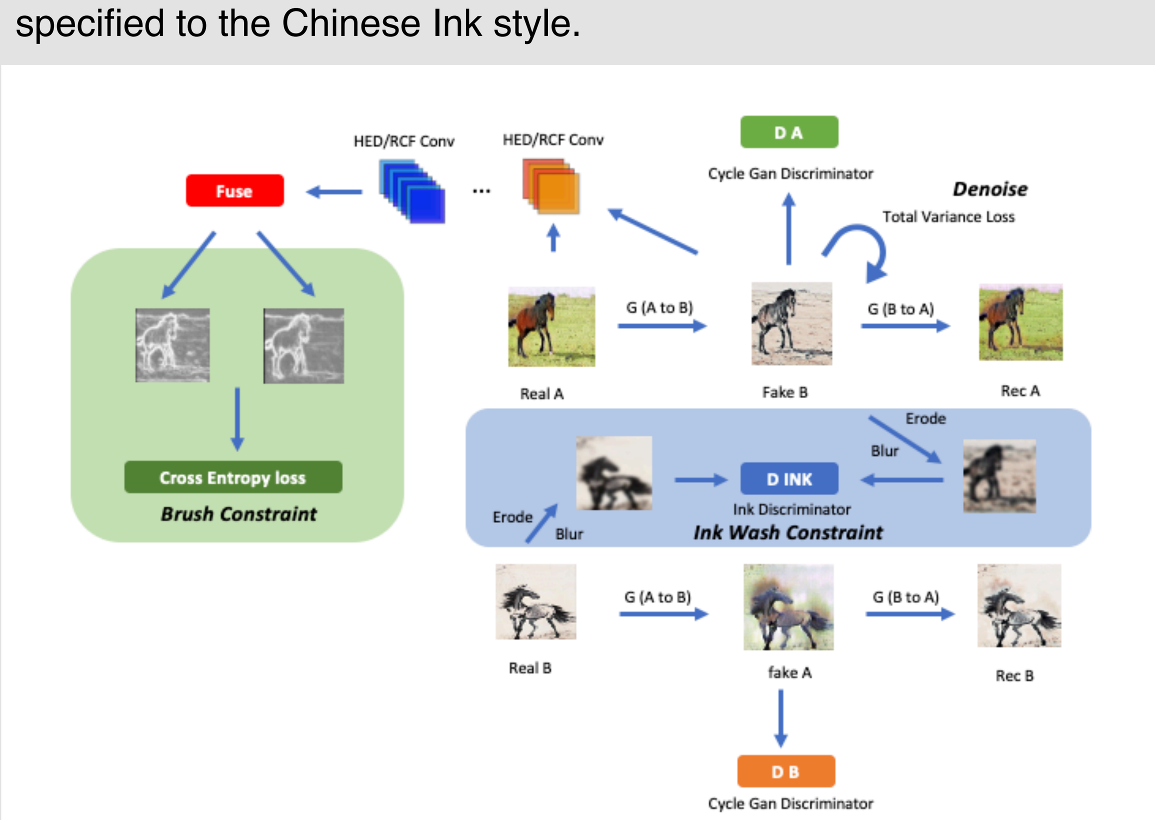
For data preprocessing, all the images will be reshaped to a squared size of (256, 256). During the training, we realized some images in the datasets, such as a fake horse from a video game, yield comparatively poorer results. We customized the datasets by removing “fake” real images. We applied data augmentation techniques, such as flipping the images or altering the contrast ratio of the images, to the dataset because there aren't many authentic paintings and qualified images.



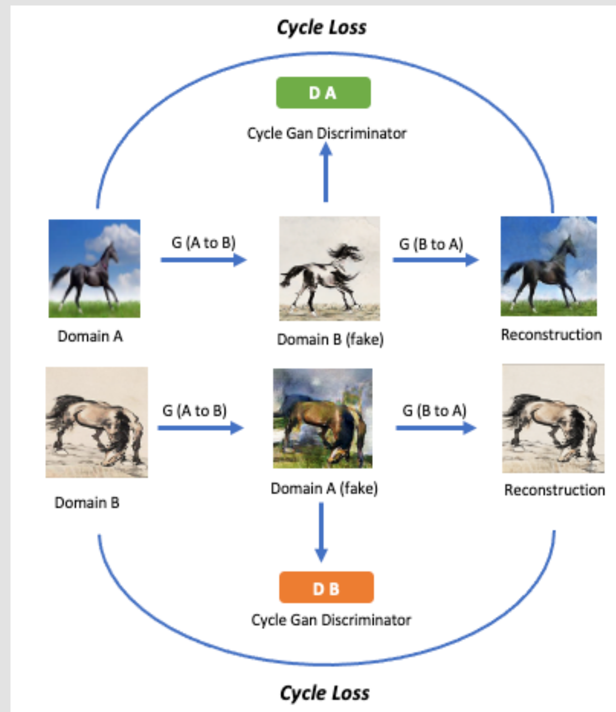


Model Architecture

We adopted CycleGAN as the basic model structure and added several constraints and refined the loss metric to make the model specified to the Chinese Ink style.



Cycle GAN



CycleGAN is a specified GAN structure that introduces a new inverse reconstruction mapping F: Y → X based on G: X → Y, and adds a consistency loss to the loss metric. By reconstruction, Cycle GAN can ensure that the generator only changes the style, and keep the content. Therefore, we do not need to input paired images for Cycle GAN. Instead, we just need to input some images specified to its style domain, without caring about its content.

CycleGAN is extremely useful for our project since we can't find sufficient paired data consisting of real images and corresponding masterpieces of Chinese Ink Paintings.

Constraint– Brush Stroke

One important characteristic of Chinese Ink Painting is that the lines of the horse’s body and texture should be very strong and bold. Therefore, for the generator, the generated image must be able to ensure that all body contours and textures are effectively retained. To force our model to focus on contours and textures, we added a new constraint to the generator loss.

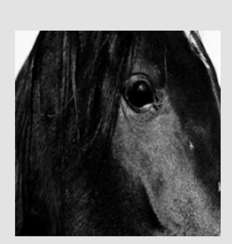
First, we will use a holistically nested edge detector(HED) to get the edge maps EA and EB of the real image (real A) and the generated painting (fake B). Second, we will add the cross entropy loss between EA and EB to the generator loss. With this constraint, we can ensure that the generator will retain the texture information well and transform these textures to mimic the "brush stroke" effect.

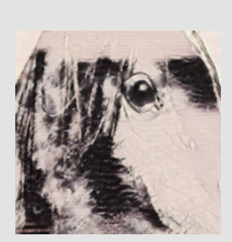
We also implemented the Richer Convolutional Features Edge Detector(RCF). This is a more refined edge detector compared to HED.

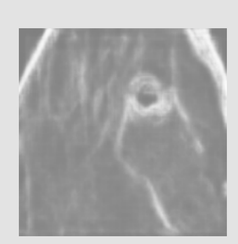
Real

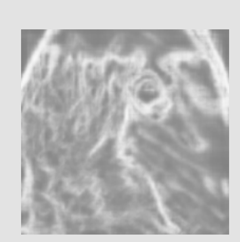
RCF

HED

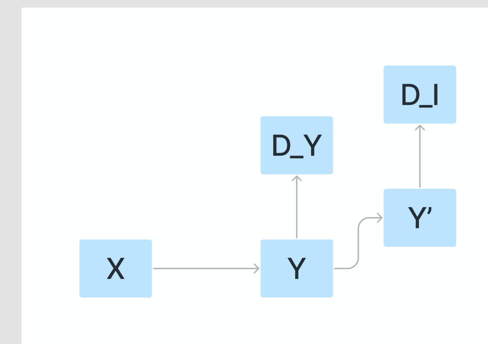








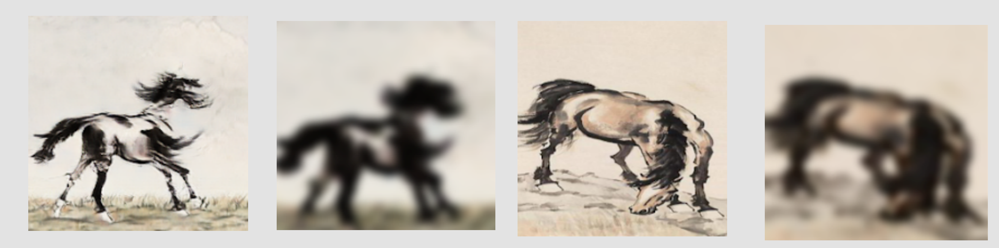
Constraints and Improvement



Chinese ink drawings are often black and white, thus the generated image's color distribution should be comparable to that of the original. As a result, we added a discriminator called D_I as a constraint to the CycleGAN model.

After the fake image was created, rather than immediately sending it to the original D Y to see if it was similar enough to a real ink drawing to trick the discriminator, we applied max-pooling and a gaussian filter to it. The blurred image will then be compared to the real image to see if they have comparable color distributions with the new discriminator D_I. Additionally, the generated fake image will be placed through the cycle and discriminator once again to see if it acquires the character of an ink painting. Therefore, we can enhance the quality of the output image by ensuring it still preserve the color distribution of the real image after transforming to ink-style paintings.

Example of the blurred image



Evaluation and Losses

To evaluate the quality of output images, we applied an evaluation model based on the paper NIMA: Neural Image Assessment which offers a method based on image classifier architectures. The CNN-based evaluation model aims to produce prediction with a correlation with not only noise, blur, or compression, but also aesthetic assessment of images. Along with the human aesthetic rating, the combined score has been used to improve our model in aesthetics.

To improve, we also applied SSIM loss in place of CycleGAN's original MSE loss function. The structural similarity (SSIM) index compares two images based on structural similarity, contrast similarity, and luminance similarity information to determine how similar they are. CycleGAN can achieve convergence more quickly and use fewer training iterations by using SSIM, which modifies pixel values from three dimensions and has a faster convergence speed than MSE loss in image reconstruction.

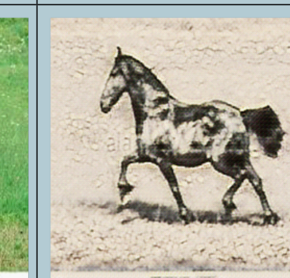
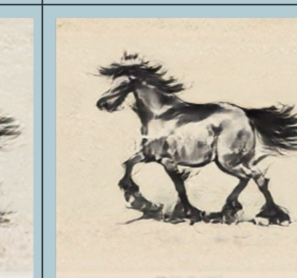
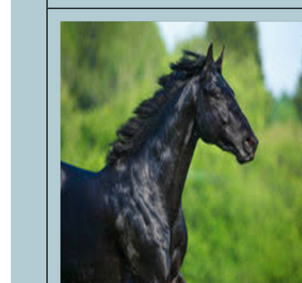
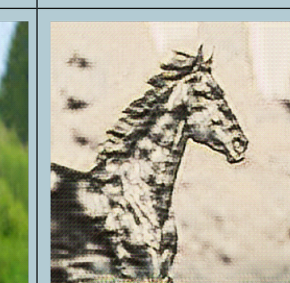
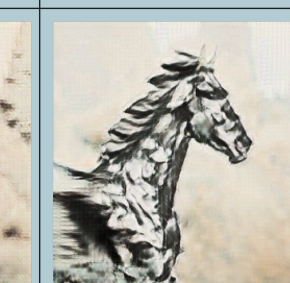
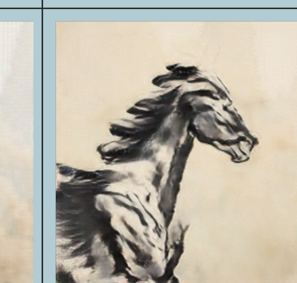
$$SSIM(x, y) = l(x, y) * c(x, y) * s(x, y)$$

We also observed that there were some images with noise. Inspired by neural image style transfer, we also added a total variance loss in generator noise to denoise the image. It has the advantage of removing noise while preserving the boundary information of the image.

$$TVL(x) = \sum i, j \left| x_{i+1,j} - x_{i,j} \right| + \left| x_{i,j+1} - x_{i,j} \right|$$

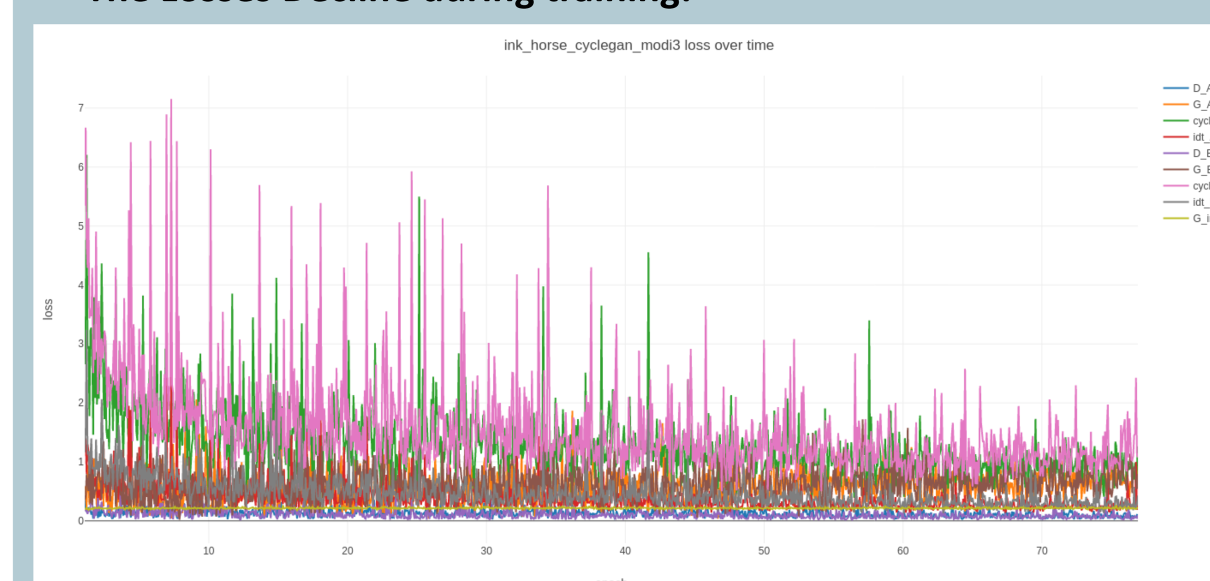
Results and Discussion

The ink style paintings which transferred from real images :

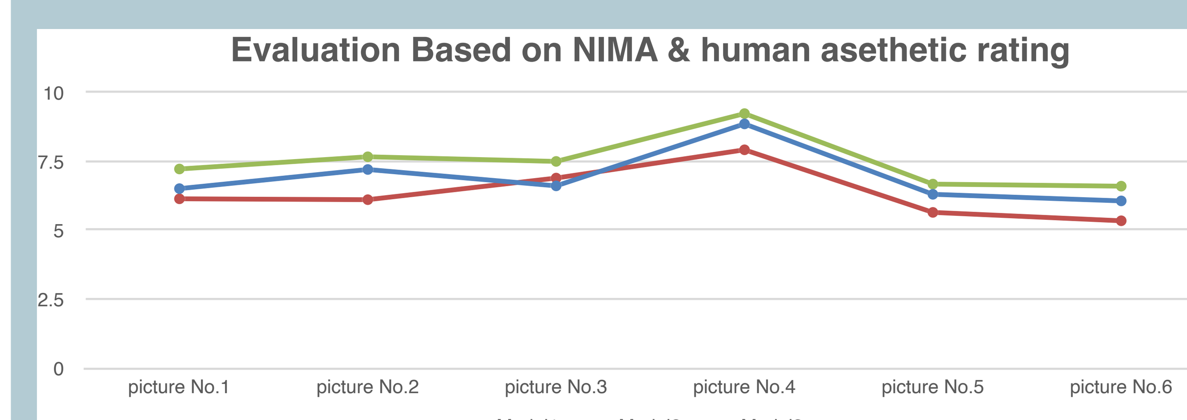
Real Image	After 5 epoch	After 20 epoch	After 60 epoch
			
			

The “fake” horses generated by the model are improving in their similarity to real Chinese Ink paintings during epochs.

The Losses Decline during training:



The Aesthetic Evaluation:



Acknowledgements

- Similarity functions. CycleGAN_ssim. (n.d.). Retrieved December 10, 2022, from https://tandon-a.github.io/CycleGAN_ssim/
- Unpaired image-to-image translation using cycle-consistent adversarial networks. CycleGAN Project Page. (n.d.). Retrieved December 10, 2022, from https://junyanz.github.io/CycleGAN/
- He, B., Gao, F., Ma, D., Shi, B., & Duan, L.-Y. (2018). Chipgan. Proceedings of the 26th ACM International Conference on Multimedia. https://doi.org/10.1145/3240508.3240655
- Aigagror. (n.d.). AIGAGROR/ML-aesthetics-NIMA: A Pytorch Nima implementation using DenseNet. Nima is a research endeavor that rates the aesthetic quality of images. GitHub. Retrieved December 10, 2022, from https://github.com/aigagror/ML-Aesthetics-NIMA