Art Style Replication

By: Colin Pascual, Sophia Li, Grace Yang, Andrew Mao

Introduction

The paper's objective is to use neural representations to separate and recombine the content and style components of any arbitrary images, which provides a neural algorithm for creating artistic images. We chose this paper because the idea of being able to learn the "style" of a piece of artwork was really fascinating, and we wanted to explore the DL structures that enable this learning.

Original	Monet	Cezanne	Ukiyoe	Bob Ross	Van Gogh	Chinese - UNet Pix-Pix	Monet - UNet Pix-Pix

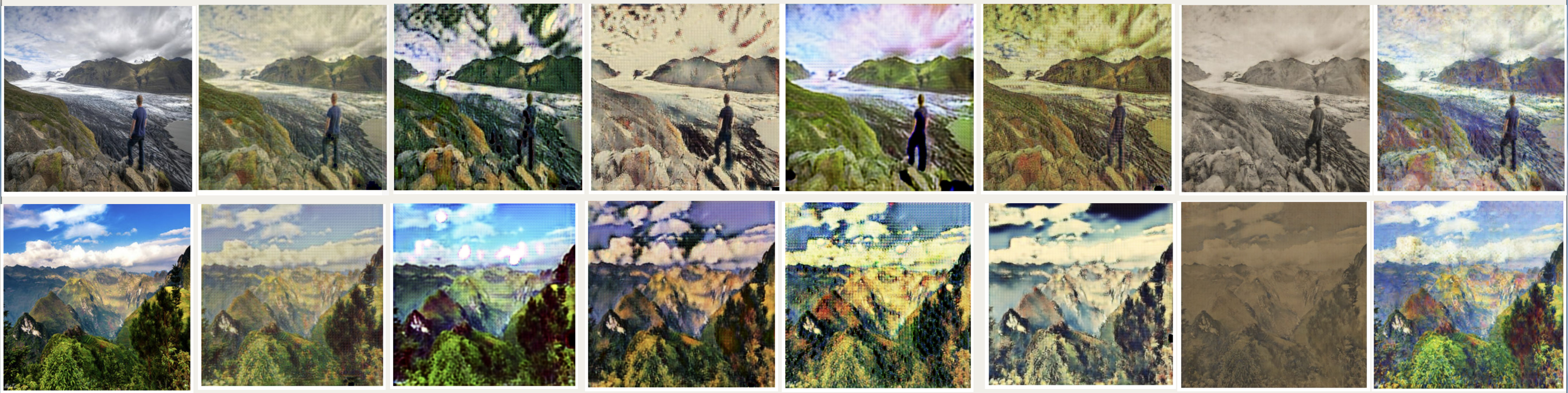

Fig. 2: A sample of our results showing the input content images being transferred to the artistic styles of Monet, Van Gogh, Cezanne, Ukiyo-e, Bob Ross, and traditional Chinese art.

Methodology

Dataset

One of the advantages of CycleGAN is that it can be trained on unpaired image datasets for style transfer. Thus, to train our GAN in a certain art style, we used a photos dataset and found datasets of art styles on Kaggle. Our photos dataset consisted of 2956 images, and art style datasets ranged from 250 to 3000 images. We preprocessed all data by normalizing to (-0.5, 0.5) and random cropping to avoid overfitting, as well as resizing to (256, 256).

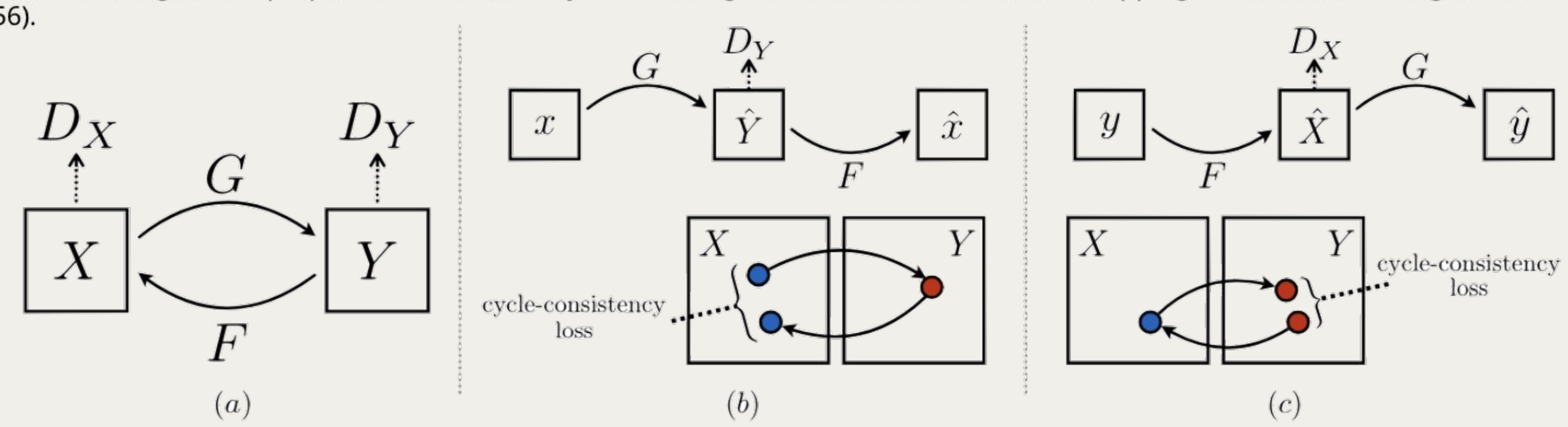

Fig. 2: Model & Loss Representations

Model

We first implemented style transfer by reimplementing neural style transfer from *A Neural Algorithm of Artistic Style*, using the VGGNet CNN architecture to capture the content of the images. However, this model simply used an algorithm that minimized content and style loss on the resulting images, and didn't require any training. Thus, we built a second model, reimplementing CycleGAN to generate images in a certain style using adversarial learning. The architecture uses two generator and discriminators, with the goal being to learn a mapping between the two chosen domains (ex: photos -> Monet art). The two generators have identical architectures: three downsampling convolutional layers, nine residual blocks, followed by two more transposed convolutional layers for upsampling. The discriminator follows the PatchGAN architecture, running a series of convolutional layers across the generated image with leaky_relu activations in a patch-like manner. We set alpha = 0.2 for our activations as guided in the paper.

We also experimented with reimplementing the Pix2Pix U-NET architecture as a generator which uses downsampling and upsampling convolutional layers along with skip-connections.

Training

The CycleGAN architecture loss function consists of three components: adversarial loss, cycle loss, and identity loss. For adversarial loss, we experimented with both binary cross entropy loss and least-squares loss to optimize generated output from the discriminators and the discriminators from generated outputs, qualitatively finding that BCE loss produced much better results wit h the U-NET generator while least-squared yielded better results for the resnet generator. Cycle loss and identity loss consists of the difference between the result from X to Y and back to X and X, and vice versa, while identity loss (to preserve color composition) is simply the difference between the mapping from X to X and Y to Y. We used the Adam optimizer with a learning rate = 0.0002 and beta_1 = 0.5 for all discriminators and generators, training for 100 epochs with a batch size of 4.

Discussion

We created a successful model that takes an image and an artist as input and returns a stylized version of that image based in the inputted artist's art style. As illustrated above, we were able to reimplement the original paper using additional datasets.

U-Net vs Resnet generator and BCE loss vs least-squared adversarial loss

- Although the paper uses the resnet generator with least-squared loss, we noticed that using a U-Net generator with binary cross entropy loss and the same hyperparameters produced sharper, more clear outputs. We also experimented with combining U-Net and least squared loss and resnet and binary cross entropy loss and noticed that these combinations did not perform as well.

Poor documentation of the original model

- While there was documentation on the architecture used, many functions were unfamiliar and took us quite long to understand completely

Mode Collapse

- For certain images, the results had white or black spots, which we determined to be caused by mode collapse of the GAN

Gridding Effect

- We also noticed that for some outputs there was a uniform blurred gridding/cell effect throughout the image. Researchers often attribute to this to the transpose convolutional layers used in upsampling.

Future Work

- **Model architecture:**
  - Our current model successfully transfers styles in terms of colors and textures.
  - We'd like to test future iterations on changes in geometric transformations (ex. geometric abstraction, surrealism).
  - Extend our model architecture to transfer art styles in animated images and video footage by applying our current process across multiple frames.
- **Dataset:**
  - Aim to train on larger datasets consisting of a more diverse spectrum of image content and styles; in doing so, we would also require more computational resources for preprocessing.
  - Instead of separating "content" and "style" inputs, we could explore further ways to adapt our model to utilise combined datasets.

Source

Original Paper: Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017.