Anime-Style Image Translation Using GANs

Li, JianPeng
Math-CS, A17152668

Lai, Pin-Hsuan
Math-Econ, A17166130

March 21, 2024

Abstract

This project explores the application of Generative Adversarial Networks (GANs) for the task of transforming real-world images into anime-style images. We implemented a GAN architecture that leverages Residual Blocks within the generator to enhance the translation. The dataset we choose comprises real-world images from COCO dataset and corresponding anime-style from https://www.kaggle.com/alamson/safebooru, with the latter undergoing a smoothing process for improved style consistency. The results demonstrate the model's ability to translate the style of input images, presenting a novel application of GANs in the domain of animate image transformation.

1 Introduction

Anime has become global cultural sensation, captivating millions worldwide. It transcends cultural and linguistic barriers, serving as entertainment and medium for self-expression and imagination. With the global rise of anime culture, the demand for applications offering anime-style filters for photos and videos has surged, highlighting the growing interest in such technology. Inspired by these trends, we aim to develop a model that can accurately convert styles while presering the essence of the original image.

The task of converting photographs into animations presents several hurdles. Chief among them is the need to amass a vast repository of animation images for training data, owing to the diverse styles emanating from various creators. Moreover, ensuring the versatility of our model across different photograph types and angles adds to the complexity. Furthermore, the sheer volume of animation and real photographs necessitates ample GPU RAM and extensive training time for successful model development.

To address these challenges, we implemented our model using Generative Adversarial Networks (GAN). Our model consists two primary components: the

generator and the discriminator. The generator serves as the crux of the process, transforming real photographs into animation-style images by drawing insights from our training data. Conversely, the discriminator operates as a classifier, trained to differentiate between genuine animation photos and those generated by generator. Through adversarial training, wherein the generator strives to produce animation photos that deceive the discriminator, and the discriminator endeavors to accurately classify real and generated images, we refine our model. Finally, we fine-tune the hyperparameters to optimize the performance of our trained model.

2 Related Work

We have found two implementations mentioned in paper [1], as referenced. The author utilized images by the specific artist Hayao Miyazaki in his training to ensure the network's output images would incorporate the artist's style. The datasets used are copyrighted, so the author could not share the sources. However, we were able to gain insight into the possible structure of GAN networks, which was a significant help in understanding our implementation. I will include the link to his paper in the reference section as [2]. The second source we examined offers an interface-like setup where we can load pretrained weights and train only on our chosen datasets. This source presents an innovative approach, enabling the flexibility to alternate between different training datasets. The work we found is at https://github.com/mnicnc404/CartoonGan-tensorflow.

3 Method

The method we implemented include the following structures:

Generator: The generator have a series of Residual Blocks, for the preservation of input information through many layers. The structure includes an encoder-decoder setup, where the input image is first down-sample to capture its content in a compact form and then up-sample to generate an image matching the target style. The inclusion of Residual Blocks within the framework enhances the model's ability to learn complex transformations and preventing gradient explosion and gradient vanishing during back propagation.

Discriminator: The discriminator's role is to distinguish between real anime images and those generated by the generator. It is structured to assess the authenticity of the images, pushing the generator towards producing more realistic (anime-style) images. we use Leaky ReLU as the activation function instead of typical ReLU, since Leaky ReLU allows gradients of negative values to pass through, which not only enables our model learning more complex features but also helps avoiding the problem of dead neurons and gradient vanishing.

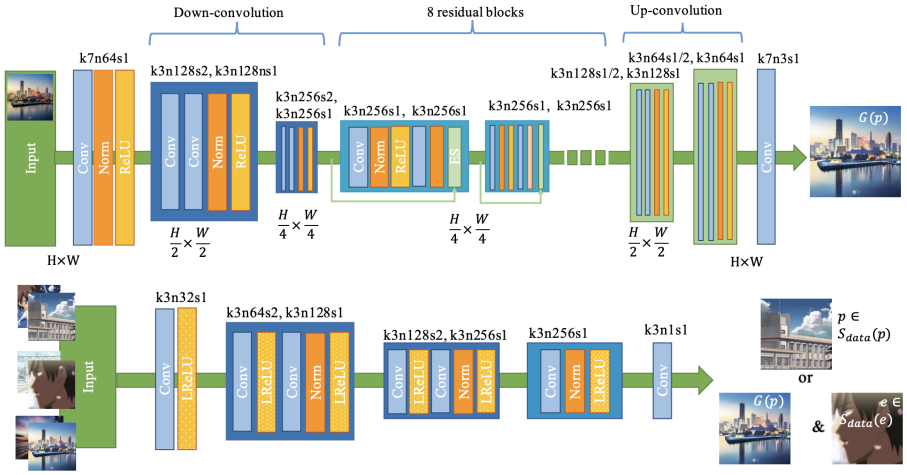

Figure 1: [1] Generator and Discriminator Structure

Training: The training process involves alternating updates between the generator and the discriminator. The discriminator is trained to maximize its ability to differentiate real anime images from the fake ones generated by the generator. At the same times, the generator is trained to minimize the discriminator's ability to make this distinction, effectively learning to produce images that are increasingly indistinguishable from real anime images. This adversarial training process is iterative carried out, involving back propagation and Adam optimization.

Strength: The Residual Blocks added help resilience to over-fitting and building deeper networks without the risk of vanishing gradients. Content Preservation: This architecture is particularly adept at preserving the original content of the images while altering the style.

4 Experiments

We utilized two primary datasets: a real-world image dataset from COCO and an anime-style image dataset from Kaggle. The anime images were pre-processed with a smoothing technique and store into new file then feed in to the training. We have successfully achieved a certain degree of anime-style transformation on real-world photos. Specifically,our model removes complex textures from clothing, faces, and backgrounds, reducing them to simpler color blocks and basic shapes. Additionally, it smoothens the contours and standardizes the colors of shadows.

5 Supplementary Material

https://www.youtube.com/watch?v=gbmVc3Ml864

References

[1] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu, "CartoonGAN: Generative Adversarial Networks for Photo Cartoonization," Tsinghua University, China; Cardiff University, UK, 2018.

[2] Filip Andersson, and Simon Arvidsson, "Generative Adversarial Networks for photo to Hayao Miyazaki style cartoons", 2020.