IMAGE STYLE TRANSFER

A Research Poster on Comparing Techniques: A Closer Look at CycleGAN vs. Neural Style Transfer (NST)

NEMANJA VUJADINOVIC
FACULTY OF TEHNICAL SCIENCE
SOFTWARE ENGINEERING AND INFORMATIONAL TECHNOLOGIES

Introduction

Image style transfer is a computer vision technique used for generating new images, merging the content of one image with the style of the second.
In this project we will compare two approaches for it: Generative Adversarial Networks (GAN) and Neural Style Transfer (NST). We explore their unique approaches, highlighting how each method contributes to creating visually appealing and distinct transformations in images.

Data

For the generation of our images, landscapes from Flickr and portraits from CelebA dataset were employed. Style references for Ukiyo-e and Cubism were derived from WikiArt.

Methodology

CycleGAN, a GAN variant, was employed for unpaired image-to-image translation. Distinguishing itself from conventional GANs, CycleGAN incorporates additional components such as cycle-consistency loss, identity loss, and adversarial loss. Moreover, its generator, designed as a UNet, incorporates residual blocks, a departure from the standard GAN UNet architecture. Notably, CycleGAN is designed for bidirectional mapping, enhancing its ability to generate realistic translations in both directions.
NST employs transfer learning, utilizing the pre-trained convolutional network. Process of NST involves performing mathematical operations on the matrices derived from both the content and style images after which new image is being generated.

Training and hyperparameters

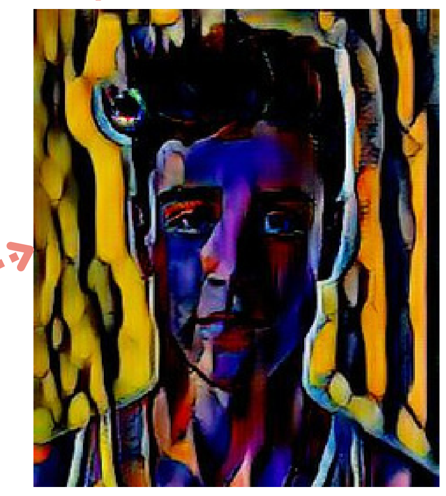
For CycleGAN, the model underwent training from scratch for 100 epochs with a learning rate of 0.0002 and batch size 1. Two separate models were saved—one specialized for translating landscape images and the other for portraits. The landscape model underwent training using 1000 landscape images and 1100 Ukiyo-e style images. The portrait model was trained on a dataset comprising 500 portrait images and 75 Cubism-style images.
In case of NST, the training involved extracting multiple layers from the VGG-19 network. The content-extracted layer was assigned a weight of 1.0, while each style-extracted layer carried an equal weight of 0.2. The loss function comprised two components: the content loss was multiplied by a hyperparameter of 5 while the style loss was multiplied by a hyperparameter of 80. These adjustments were crucial in fine-tuning the model, ensuring a harmonious transfer of both content and style during the training phase.

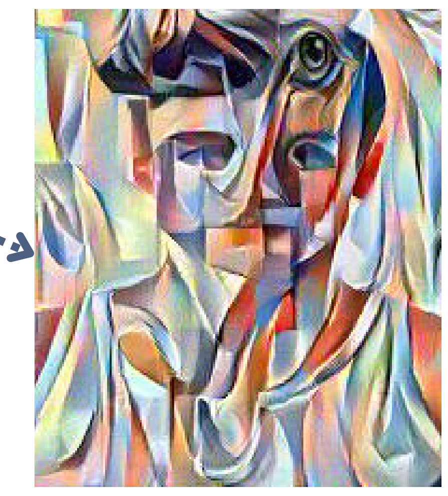
Training insights

During CycleGAN training, loss started at 7, declining to around 3 when using a learning rate of 0.0002. Conversely, with a learning rate of 0.001, the loss oscillated and stagnated around 4. However, the focus shifted from loss metrics to visual quality, prioritizing perceptual excellence over numerical values. In contrast to the conventional training, NST's approach focused on visualization over epochs.

Results and comparison

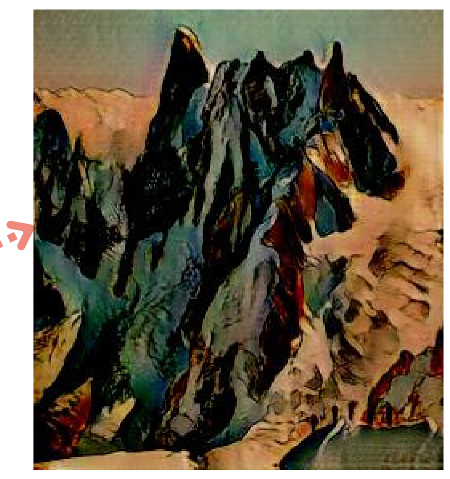
Portrait to cubism style transfer

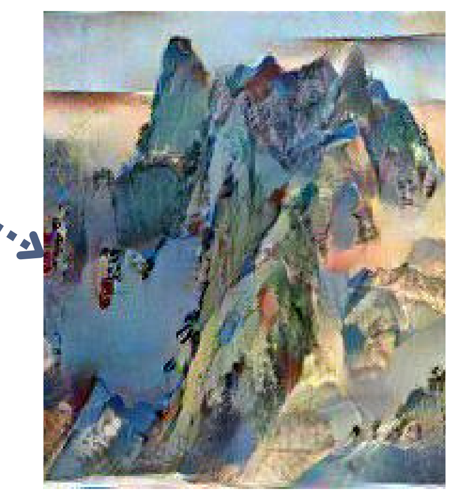





The CycleGAN model excels in preserving portrait features while applying Cubist style, maintaining integrity in eyes, nose, and body. NST, on the other hand, distorts portraits more extensively to match Cubist style, emphasizing CycleGAN's nuanced approach to style transfer. It's noteworthy that NST performs well in achieving a unique artistic interpretation, but obviously, in different way then CycleGAN does.

Landscape to Ukiyo-e style transfer



The CycleGAN model demonstrated a superior understanding of Ukiyo-e's characteristic traits, producing visually pleasing results. In contrast, NST, which relies on a single reference image, struggled to capture the nuanced and subtle features inherent in Ukiyo-e, resulting in less satisfactory outcomes. This highlights the advantage of CycleGAN, which, by training on a diverse set of images, excels in creating a cohesive style rather than relying solely on individual characteristics.

Conclusion

Both approaches, CycleGAN and NST, deliver outstanding results, with CycleGAN excelling in realism and NST bringing an artistic touch. NST requires less time compared to CycleGAN, however it performs less optimally when the distinctive style features in the reference image are not obvious. Therefore, the choice between the two depends on the desired balance between realistic and artistic outcomes, as well as time considerations and viewer preferences.