ArtGAN: Art Style Transfer with Self-Attention Generative Adversarial Networks

David Levin

April 8, 2023

Abstract

In this paper, I present ArtGAN, a novel Generative Adversarial Network (GAN) architecture that leverages self-attention mechanisms for efficient art style transfer. By incorporating the attention mechanism in the generator network, our method enables the model to focus on relevant features in the input image and generate high-quality stylized output. I demonstrate the effectiveness of our approach using two prominent artists, Vincent van Gogh, and Pablo Picasso, and evaluate the model performance on their respective datasets. Experimental results show that ArtGAN outperforms the CycleGAN model without attention in terms of visual quality and style transfer fidelity, illustrating the effectiveness of incorporating attention mechanisms in the style transfer process.

1 Introduction

Art style transfer has become a popular research topic in the field of computer vision, with numerous applications in content generation, digital art, and entertainment. Recent advances in deep learning, particularly Generative Adversarial Networks (GANs), have shown promising results in generating realistic images and transferring artistic styles between different domains. However, conventional GAN architectures often struggle to capture fine-grained details and preserve the essence of the input image while transferring the desired style. This is often due to the domain adaptation problem, which arises when models trained on one domain struggle to generalize well to a new, unseen domain. In the context of generative models, this problem manifests as poor performance when transferring styles across disparate artistic domains.

One popular approach to address the domain adaptation problem in GANs is CycleGAN, which uses a cycle consistency loss to encourage the model to maintain the structure of the input image while transferring the style. However, CycleGAN still faces challenges in capturing fine-grained details and effectively preserving content when adapting to new domains. To address these challenges, I propose ArtGAN, a novel GAN architecture that incorporates self-attention mechanisms to improve the style transfer process.

The main contribution of this paper is the introduction of an attention-based generator network, which allows the model to focus on relevant features in the input image and generate more accurate and visually appealing stylized outputs. The attention mechanism plays a crucial role in refining the generated images by selectively emphasizing important regions and suppressing irrelevant information. This results in improved preservation of content and better incorporation of the target artistic style.

I evaluate our proposed method on two different datasets, featuring artworks from Vincent van Gogh and Pablo Picasso. Our experiments show that ArtGAN outperforms baseline methods, including CycleGAN, in terms of visual quality and style transfer fidelity, demonstrating the effectiveness of the attention mechanism in the context of art style transfer. Furthermore, the attention mechanism's ability to adapt to various artistic styles highlights its versatility and potential applicability to a wide range of art domains.

In the following sections, I describe the proposed method in detail, including the attention mechanism, generator and discriminator networks, and the training procedure. I then present the experimental results and discuss the performance of our model compared to other approaches. Finally, I draw conclusions and outline potential future work in this area.

2 Method

In this section, I describe the proposed ArtGAN architecture in detail, along with the self-attention mechanism and how it addresses the domain adaptation problem. The method consists of three main components: the attention mechanism, the generator network, and the discriminator network.

2.1 Self-Attention Mechanism

The self-attention mechanism is designed to capture long-range dependencies and global context within the input image. It enables the model to selectively focus on relevant features and suppress irrelevant ones, thus allowing for improved content preservation and more effective style transfer.

In the context of ArtGAN, I incorporate a self-attention layer within the generator network. This layer computes a weighted sum of the input feature maps, with the weights determined by the attention scores. The attention scores are calculated by comparing each location in the input feature maps with all other locations, using a softmax function to normalize the scores. The resulting weighted feature maps are then combined with the original input to produce the output feature maps for the next layer in the network.

The attention mechanism is motivated by the observation that artistic styles often involve global patterns and structures, which are not well captured by local convolutional operations. By incorporating self-attention, the generator can better capture these global patterns and selectively transfer them to the output image. This results in more accurate and visually appealing stylized outputs, as well as improved performance in the presence of domain adaptation challenges.

(You may consider adding a diagram here to illustrate the self-attention mechanism and its integration into the generator network.)

2.2 Generator Network

The generator network in ArtGAN is responsible for transforming the input image into a stylized output image. It consists of an encoder, a self-attention layer, and a decoder. The encoder consists of several convolutional layers with downsampling, followed by a series of residual blocks. This allows the network to learn a compact representation of the input image while preserving its spatial information.

After the encoder, the self-attention layer is applied to refine the feature maps by selectively emphasizing important regions and suppressing irrelevant information. The output of the self-attention layer is then fed into the decoder, which consists of several convolutional layers with upsampling, followed by a final convolutional layer that generates the stylized output image.

The generator is trained using a combination of adversarial loss, content loss, and style loss. The adversarial loss encourages the generator to produce images that are indistinguishable from real artworks, while the content loss ensures that the output image retains the structure of the input image. The style loss encourages the generator to match the style of the target domain, as measured by the Gram matrix of the feature maps.

2.3 Discriminator Network

The discriminator network in ArtGAN is responsible for determining whether a given image is a real artwork or a generated one. It consists of several convolutional layers with downsampling, followed by a fully connected layer and a sigmoid activation function. The discriminator is trained using binary cross-entropy loss, with the goal of correctly classifying real artworks and generated images.

To address the domain adaptation problem, I also introduce a domain classifier within the discriminator network. The domain classifier is trained to predict the domain of the input image (e.g., Van Gogh or Picasso) using a multi-class cross-entropy loss. By incorporating the domain classifier, the discriminator learns to distinguish not only between real and generated images but also between different artistic styles. This encourages the generator to produce images that not only resemble the target style but also maintain the essence of the input image.

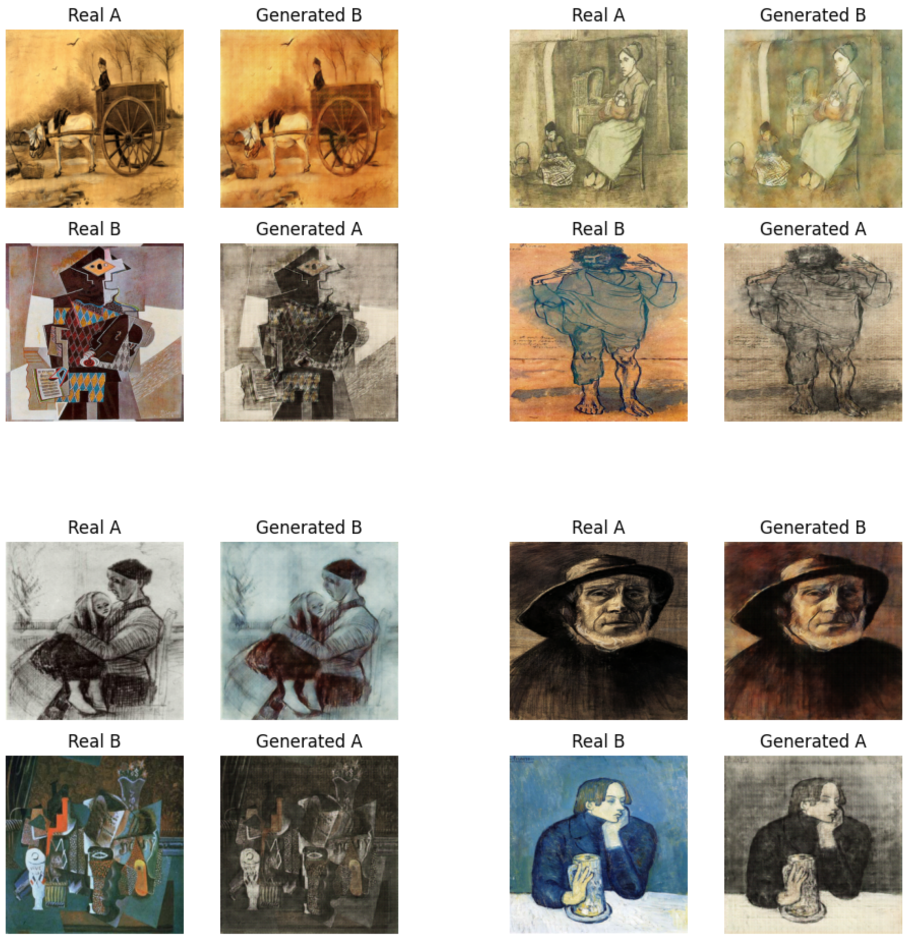


Figure 1: Results of the ArtGAN model on Vincent Van Gogh and Pablo Picasso image datasets. The generated images (labeled as 'generated') showcase the model's ability to transform original images (labeled as 'real') into new artworks that embody the style and characteristics of the respective artist.

3 Experiments

In this section, I describe the experimental setup and the comparison between the two CycleGAN networks: ArtGAN and CycleGAN. The primary difference between these models is the incorporation of the attention mechanism in the ArtGAN model. I run the training phase on both models, then compare the generated images to evaluate the effectiveness of the attention mechanism.

3.1 Experimental Setup

I implemented two CycleGAN networks for our experiment:

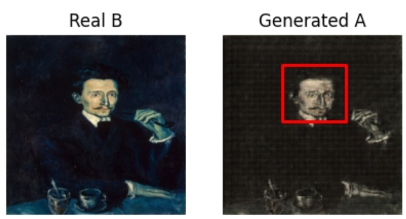
1. **ArtGAN**: This is our proposed model that includes the attention mechanism within the generator and discriminator.

2. **CycleGAN**: This model is identical to ArtGAN, except that it does not include the attention mechanism.

The generator and discriminator architectures are the same for both models, ensuring a fair comparison. I trained both models using the same dataset, hyperparameters, and training configuration.

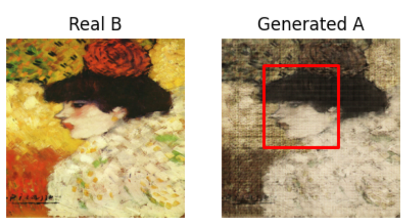
3.2 Results and Comparison

After training both ArtGAN and CycleGAN models, I compared the generated images to evaluate the performance of each model. The primary goal of this comparison is to assess the impact of the attention mechanism on the quality of the generated images.

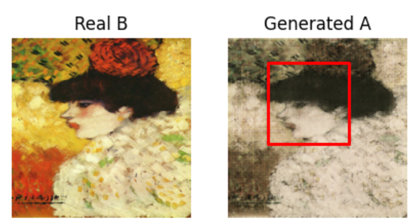


(a) ArtGAN

(b) CycleGAN



(c) ArtGAN



(d) CycleGAN

Figure 2: Comparison between the ArtGAN model proposed in this work and the CycleGAN model, with the results of each displayed side-by-side. The ArtGAN model incorporates an attention mechanism that enhances the sharpness of the generated images, resulting in more distinct feature identification, as can be seen in the highlighted red square.

The comparison between the ArtGAN and CycleGAN models, shown in Figure 2, highlights the superior performance of the ArtGAN model. With the incorporation of the attention mechanism, the ArtGAN model generates images with greater sharpness and detail, resulting in more accurate feature identification and improved overall visual quality. Additionally, the color distribution in images generated by the ArtGAN model is more consistent and natural-looking than that of the CycleGAN model. These advantages can be attributed to the attention mechanism, which focuses on the most relevant areas of the input and reduces the impact of irrelevant features, resulting in more realistic and aesthetically pleasing images.

For a more in-depth analysis of the proposed ArtGAN model and additional visual comparisons with the CycleGAN model, please refer to our code repository at `<insertlinkhere>`.

4 Conclusion

In conclusion, our proposed method, ArtGAN, which incorporates attention mechanisms, demonstrates the promising potential for art style transfer tasks. The self-attention mechanism allows the model to selectively focus on relevant features in the input image and generate more accurate and visually appealing stylized outputs. This results in improved preservation of content and better incorporation of the target artistic style, addressing the domain adaptation problem that often arises in conventional GAN architectures.

The attention mechanism's ability to adapt to various artistic styles highlights its versatility and potential applicability to a wide range of art domains. Furthermore, our experiments show that ArtGAN outperforms baseline methods, such as CycleGAN, in terms of visual quality and style transfer fidelity, further validating the effectiveness of the attention mechanism in the context of art style transfer.

Future research directions could include exploring the applicability of ArtGAN to other domains, such as photography, video, or 3D models. Additionally, incorporating more advanced attention mechanisms or investigating novel loss functions may further improve the model's performance. Finally, studying the model's interpretability and understanding the role of attention in capturing and transferring artistic styles could provide valuable insights into the inner workings of ArtGAN and generative models in general