CSCI 1430 Final Project Report: Behind AI Artist: Neural Style Transfer

Team name: Lingyu Ma, Dawei Si, Yichen Chai, Cancan Huang
Brown University

Abstract

Ever since the ground-breaking work by Gatys et al [1], the neural style transfer (NST) has become the mainstay in the style transfer field, having attracted wide attention from both academia and industry. A variety of approaches has been proposed to improve and extend the performance of the original work. In our final project, we studied the arbitrary style transfer model, AdaIN [2], proposed by Huang et al., which is capable to transfer any image to arbitrary new styles using one single trained model while achieving speed comparable to the fastest existing approaches. We have implemented AdaIN using Tensorflow 2.0, which is publicly available at: https://github.com/cancan233/AdaIN_CS1430 and explored its application by constructing an online demo on GCP, of which source code hosted at https://github.com/lingyuma1996/CS1430_final_project. Another model Magenta [3] is also included in the demo.

1. Introduction

Several important technological advances achieved in the last few years supported the rising interest in AI Art. Fig. 1 points out several important technologies milestones leading us to current AI Art level. Among those developments, one of the most iconic AI inventions that triggered the rapid use and development of AI technologies for art was Neural Style Transfer (NST) [4]. To automatically transfer an artistic style, the first and most import issue is how to model and extract *style* from an image [5]. The ground-breaking work by Gatys et al. [1] firstly demonstrated the power of *Convolutional Neural Networks (CNNs)* to reproduce target painting styles on chosen images without losing the content information. Their experimental results found that a CNN model is capable of extracting *content* information as well as *style* information from given photos. Based upon this finding, they reformulate the style transfer problem into an optimization problem in which the output image is iteratively optimized to match the desired CNN feature distributions involving content representation of the photograph and the style representation of the artworks.

Significant efforts has been devoted to either improve or extend the original NST algorithm in both academia and industry, leading to many successful real-world applications [6, 7]. In this work, we firstly compared performances of several representative NST algorithms [1, 8, 3, 2] using the same content with different style images. After the comparison, we chose to implement the last model, AdaIN ourselves in Tensorflow 2.0, which is capable of tranfering arbitrary new styles in real-time at the speed similar to the fastest feed-forward approaches [8]. We also use Vue to construct an online demo using our model and Magenta [3] for people to test.

2. Related Work

Optimization-based method

In the work by Gatys et al. [1], an output image with random noise is used as a starting point for optimization. With a pretrained VGG16 as backbone whose parameters are frozen, it calculates the content loss with the content image as well as style loss (gram matrix differences) with the style image from feature maps and uses back-propagation to update the output image. However, it requires to be trained for each content-style image pair and the optimization process can take up a lot of time to converge for a satisfying result.

Feed-forward approach

Johnson et al. [8] proposed a feed-forward approach established on the optimization-based method, which employs a feed-forward net to fit the output image instead of optimizing from random noise. After training the content image can be easily stylized with a single forward pass, with 1000 times faster than Gatys et al. [1]. However, the feed forward network can only be used for one specific style after training and it needs to be retrained for different styles.

Instance Normalization.

Ulyanov et al. [9] proposed to use instance normalization instead of batch normalization in the feed-forward net, which prevents instance-specific mean and covariance shift simplifying the learning process as well as rendering qualitative

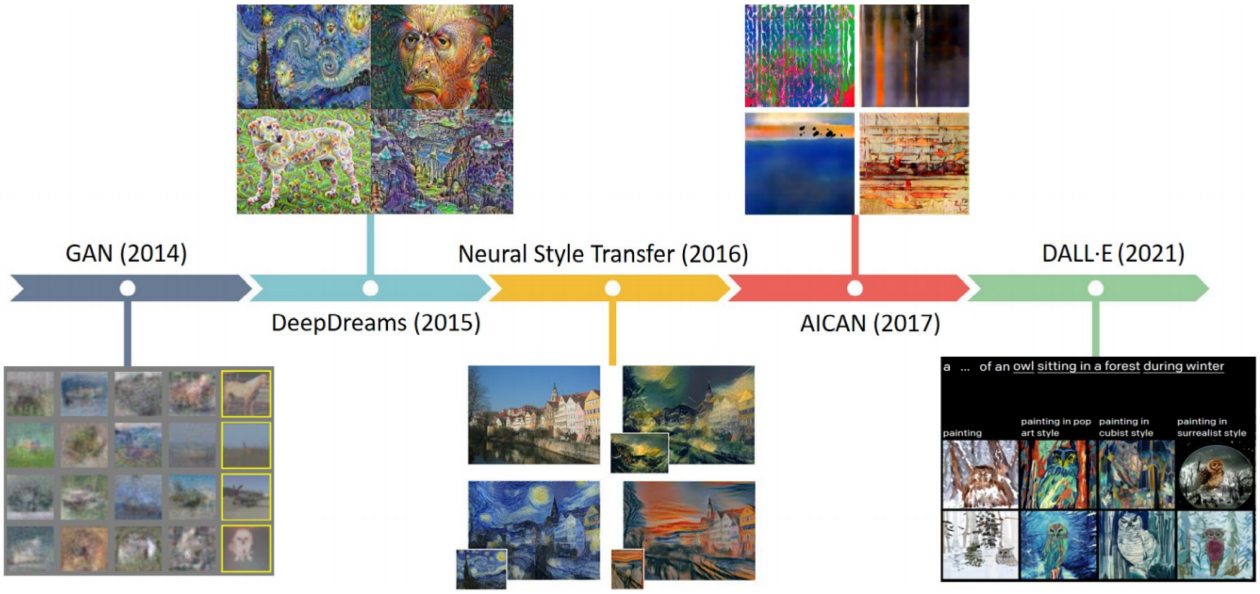

Figure 1. Illustration of several important milestones lying on the growth path of AI Artist[4]

improvements on the results.

Dumoulin et al. [3] proposed an approach to use different parameters for different style images in instance normalization while keeping the rest of the feed-forward net parameters the same. These parameters are learned during the training step. In this way, multiple styles can be learned within a single model.

Statistics of feature maps

Li et al. [10] did a theoretical analysis on neural style transfer methods. They found that the widely used gram matrix style loss can be reformulated as the Maximum Mean Discrepancy (MMD) of the two feature map distributions with a polynomial kernel. They also proposed that simpler statistics like mean and variance can also be used as the style loss function for style transfer.

3. Method

3.1. Adaptive Instance Normalization

Built upon the previous development of neural style transfer methods, Huang et al. [2] proposed an adaptive instance normalization method, which is basically normalizing the feature maps of the content image with the statistics (mean&variance) from the feature maps of the style image. The elegant point for it is using the style image statistic to replace the style parameters supposed to be learned in training step in [3]. In this way, the model can be used for any style transfer once it is trained.

$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y) \tag{1}$$

where x is content input and y is style input. AdaIN has no learnable affine parameters. Instead, it adaptively computes the affine parameters from the style input.

3.2. Model Architecture

The full architecture of AdaIN model is shown in Fig. 2. After preprocessing procedure, the content image and style image are passed through fixed pre-trained normalized VGG19 encoder f, and the AdaIN layer to get the target feature maps $t = \mathrm{AdaIN}(f(c), f(s))$, and then a decoder is trained to map it to the stylized image $T(c, s) = g(t)$.

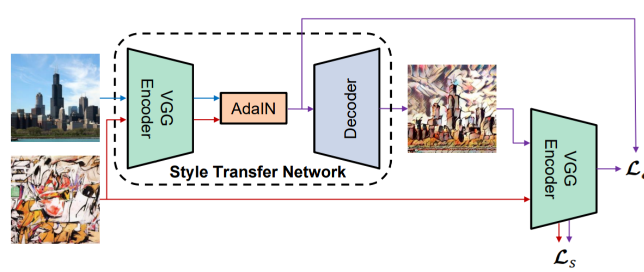

Figure 2. AdaIN model architecture

The loss of AdaIN model is as follows.

$$
\begin{aligned}
\mathcal{L} &= \mathcal{L}_c + \lambda \mathcal{L}_s & (2)\\
\mathcal{L}_c &= \|f(g(t)) - t\|_2 & (3)\\
\mathcal{L}_s &= \sum_{i=1}^{L} \|\mu(\phi_i(g(t))) - \mu(\phi_i(s))\|_2 & (4)\\
&\quad + \sum_{i=1}^{L} \|\sigma(\phi_i(g(t))) - \sigma(\phi_i(s))\|_2 & (5)
\end{aligned}
$$

where each ϕ_i denotes a layer in VGG-19. They also proposed that one can control the content-style trade-off at test time by setting

$$T(c, s, \alpha) = g((1-\alpha)f(c) + \alpha \mathrm{AdaIN}(f(c), f(s))) \quad (6)$$

3.3. Online Demo

We have implemented a website application that integrated Magenta (one network with multiple styles by Dumoulin et al.[3]) and our own trained AdaIN models. For the Magenta model, there are 42 different artistic styles for users' choices, and it allows users to upload their content images. For the AdaIN model, users can customize both content and style images to generate the newly stylized image by using our trained neural network (Fig. 3).

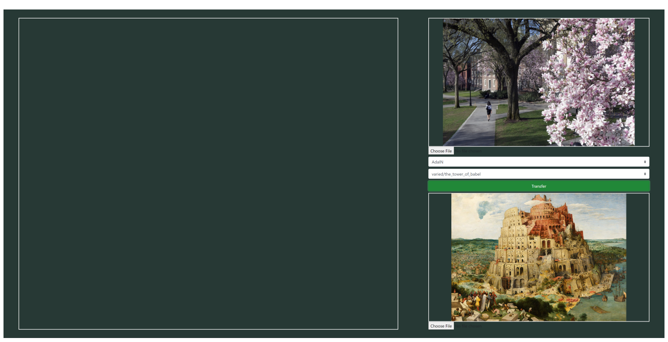

Figure 3. Screenshot for online demo application

4. Results

4.1. Comparison with existed methods

For test purpose, we choose Starry Night by Vincent Van Gogh, The Scream by Edvard Munch and Head of Clown by Georges Rouault as our style images. An beautiful image taken at the main green acts as our content image. We select the content image with the reason that its left side has quite complicated informations, such as depth, different objects, and its right side has a large color block, which we think can be a good test for our models. Fig. 4 shows the stylized images using different models respectively.

4.2. Technical Discussion

We use MSCOCO[11] as content images, and [12] datasets as style images to train the AdaIN model same as in the original paper. All model trainings were conducted using GPUs on Oscar, supercomputer of Brown University. And online demo was hosted on Google Cloud Platform (GCP) with educational credit from CS1430. For the final training results, please see Fig. 11.

Firstly we tried to directly resize the images in training data to 256×256 without maintaining their aspect ratios. The model does learn to stylize images, however, it produces images with significant checkerboard-like artifacts as shown below (especially conspicuous for small images). Using images resized and cropped to the target size while maintaining the aspect ratios solved the problem and produces good results. The inconsistency in the model when training on differently distorted images may account for the artifacts, thus maintaining aspect ratio of training data is vital for style transfer.

Figure 5. Stylized images with distorted training data (left), and training data that maintain aspect ratio (right). Style: Head of a Clown

We've also analyzed the effects of the normalization of VGG net as firstly mentioned in Gatys' paper[1], and used by most of the style transfer works. It turns out that when training on a single 256×256 image pair for hundreds of iterations, both original VGG19 and normalized VGG19 can produce similar results for the small image, but normalized VGG19 gives much better results for the original-size image. Thus the normalized VGG net is indeed more suitable for neural style transfer.

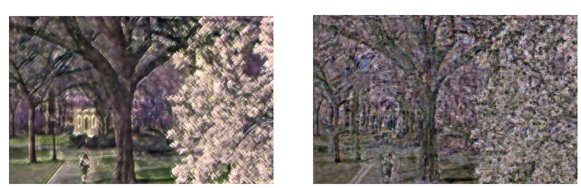

Figure 6. Stylized images for the original image after training on a single 256×256 image pair for 300 iterations, with normalized VGG19 (left), and orginal VGG19 (right). Style: Head of a Clown

We tried to use zero padding for images at first, and it

Figure 4. Comparison results for four representative models. The results are generated using the code provided by the original author.

turns out that the output image is well stylized but there are some artifacts at the borders, as the figure below shows. In contrast, model trained with reflection padding produces much more natural borders. We also find that it can still work well even if we train the model with zero padding and use reflection padding to stylize images afterwards at test time.

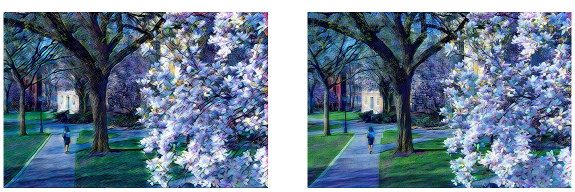

Figure 7. Comparison of the results with zero-padding (left) and reflection-padding (right). Style: The Starry Night

We also test the influence of $lambda$ in the output shown in Fig. 8. The parameter λ is used for deciding the weight of style loss in final loss for the decoder. As shown in Fig. 8, with larger λ value, the texture in the output image is more apparent. We could see there are deeper lines in image with $\lambda = 10$ than image with $\lambda = 2$.

We also notice that it's not quite appropriate to regard α as the parameter for trading off content and style as indicated in the original paper. As we can see from the figures below, instead of reconstructing the original image, setting $\alpha = 0$ actually results in a stylized image of the original image with its own style. Thus strictly speaking, the parameter α is trading off the style of the content image and the style of the style image.

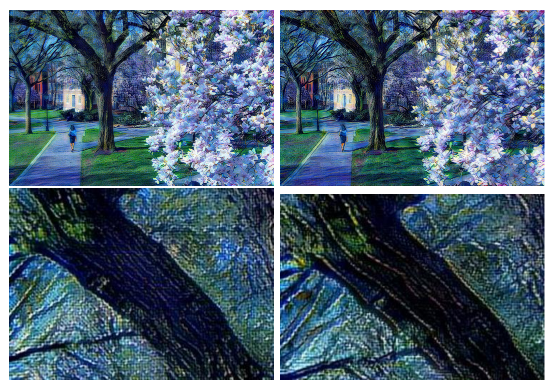

Figure 8. Comparison of output images generated by models trained with $\lambda = 2$ (left) and $\lambda = 10$(right). The bottom two images are zoomed-in images of the top two images respectively.

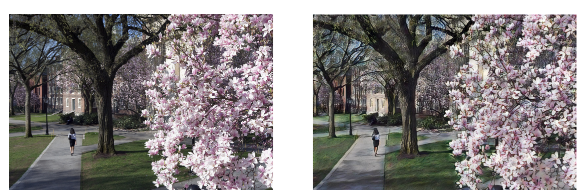

Figure 9. Comparison of original image and stylized image with trained model ($\lambda = 10$), α set to 0

Lastly, we also find that the so-called content loss in the paper is not well-defined, since t has gone through the

AdaIN layer, while $f(g(t))$ does not. This implicates that the content loss will be influenced by the style image. As we can see from the image below, setting $\lambda = 0$ does not recover the original image, but results in a not-well-stylized image.

Figure 10. Stylized image with trained model ($\lambda = 0$). Style: The Starry Night

4.3. Societal Discussion

1. Describe the socio-historical context of your project to identify three broad societal factors that could affect your data, goal, and/or hypothesis. These factors might include current or historical policies, events, social conditions, and larger societal systems. Cite at least one outside source.

   Our project aims to understand and elaborate neural style transfer algorithms, which is vital for current AI Art technologies. Thus, we would like to discuss the socio-historical context for AI artists from the following three aspects.

   - *Authorship*: AI artwork sells for $432,500 as Christie's becomes the first auction house to offer a work of art created by an algorithm [13], more than 40 times the value estimated before the auction. The works' production involved the use of Generative Adversarial Networks (GANs) and it turned out that the model has largely relied on code written by Robbie Barrat, a open source developer, who did not receive credit for the work, nor any remuneration from the sale [14]. This raise an important concept: who is the author of an artwork that is generated by a computer program? Most of the recent examples of sold AI artworks indicate that currently the authorship rights are attributed to the artist who produced the artwork using AI techniques, regardless of the narrative surrounding the creation process, e.g. the fact that the artwork was labelled as being made by an AI.
   - *Copyright*: Part of the training data for AI Art generation could include copyrighted images, the final output would in that case involve someone else's artistic contribution. This could of course be hardly noticeable in the final work, but still require an acknowledgement from an ethical perspective. In the meanwhile, not all artists are keen to expose all steps of their creative process in creating AI artworks as specific choices constitute the basis for originality and uniqueness. Yanisky-Ravid et al. [15] argue that confronting the challenges of the autonomous and automated content production calls for a reassessment of the meaning of originality.
   - *Ethical Issues*. Ethical concers were also raised in previous discussed Christie case [13] as the model being the "authors" of a work generated using code written by a third party. For good purposes, AI Art has also been used for reconstructing lost artwork, by applying neural style transfer to x-radiographs of artwork with secondary interior artwork beneath a primary exterior [16]. In this way, it, as an artistic tool, broadens creative insight and widen the landscape of inspirational ingenuity.

2. Who are the major stakeholders in this project? What is your relationship to these stakeholders?

   The stakeholders related to this project can be divided into following groups:

   - *Artists* who draw the original masterpieces. We have to use their works as style images to train the neural network for style transfer.
   - *Photo owners*. We use their photos as content images in our training step.
   - *People* shown in photos or portraits. We use MS-COCO dataset in our training step and it might include people faces.

   To all groups described above, our relationship is that we use the materials generated by artists or photo owners and recreate new instances.

3. Research or journalism on your broader project topic may have already been conducted. What was the societal impact of existing research? Discuss the implication of this research on your project

   As discussed in previous question, the artwork created by AI Artiest has revealed many issues regarding the questions of authorship and copyright, as well as raised general discussions on the ethical considerations that have to be taken into account during production, promotion and sale of an AI artwork. It also brings relevant changes to the contemporary art market as there is a rising trend to shift traditional art enterprises to corresponding online versions [4].

4. How could an individual or particular community's civil rights or civil liberties (such as privacy) be affected by your project?

   Our project doesn't include any factor that will affect an individual or particular community's civil rights. The only concern here is that the training data might include people faces which could lead to privacy leakage.

5. If you are using data, what kind of biases might this data contain? Do any of these represent underlying historical or societal biases? How can this bias be mitigated? Consider the following questions to help you:

   - Were the systems and processes used to collect the data biased against any groups?
   - Is the data being used in a manner agreed to by the individuals who provided the data?

   For our style training data, it is from the competition, Painter by Numbers. As described on the website, many of the images in the dataset were obtained from wikiart.org. Additional painting were provided by artists. There might be art style bias in the collection process.

   As all dataset were obtained from public dataset for the same purpose, we think it is used in a manner agreed to by the individuals who provided the data.

5. Conclusion

Over the past several years, NST has continued to become an inspiring research area, motivated by both scientific challenges and industrial demands. Our project has compared several representative NST algorithms in history. And we also discuss the performance of AdaIN in detail from several aspects. It shows promising applications in the future research.

Despite the great progress in recent years, the area of NST is far from a mature state. Currently, the first stage of NST is to refine and optimise recent NST algorithm, aiming to perfectly imitate varieties of styles while reduce failure cases and improve stylised quality on a wider variety of style and content images. It also requires deriving more extensions from general NST algorithms. The next stage, a further trend of NST is to not just imitate human-created art with NST techniques, but rather to create a new form of AI-created art under the guidance of underlying aesthetic principles such as DALL·E [6].

References

[1] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015. 1, 3, 7

[2] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017. 1, 2, 7

[3] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016. 1, 2, 3, 7

[4] Eva Cetinic and James She. Understanding and creating art with ai: Review and outlook. *arXiv preprint arXiv:2102.09109*, 2021. 1, 2, 5

[5] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019. 1

[6] OpenAI. Dall·e: Creating images from text, Feb 2021. 1, 6

[7] Bhautik Joshi, Kristen Stewart, and David Shapiro. Bringing impressionism to life with neural style transfer in come swim. In *Proceedings of the ACM SIGGRAPH digital production symposium*, pages 1–5, 2017. 1

[8] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 1, 7

[9] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 1

[10] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. Demystifying neural style transfer. *arXiv preprint arXiv:1701.01036*, 2017. 2

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 3

[12] Kiri Nichol. Painter by numbers, wikiart. 2016. 3

[13] Person. Is artificial intelligence set to become art's next medium?: Christie's, Dec 2018. 5

[14] Jon McCormack, Toby Gifford, and Patrick Hutchings. Autonomy, authenticity, authorship and intention in computer generated art. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*, pages 35–50. Springer, 2019. 5

[15] Shlomit Yanisky-Ravid and Luis Antonio Velez-Hernandez. Copyrightability of artworks produced by creative robots and originality: the formality-objective model. *Minn. JL Sci. & Tech.*, 19:1, 2018. 5

[16] Anthony Bourached and George Cann. Raiders of the lost art. *arXiv preprint arXiv:1909.05677*, 2019. 5

Team contributions

6. Acknowledgements

This work was conducted using computational resources and services at the Center for Computation and Visualization, Brown University.

Figure 11. Results obtained using the AdaIN implemented ourselves. The content image is the same as the one used in model comparsion part.

Appendix

Fig. 11 contains several style transfer results using the AdaIN model we implemented ourselves. The style images are shown in the top left corner while the content image is the same as the one used in Section 4.1.

Please describe in one paragraph per team member what each of you contributed to the project.

Lingyu Ma As for the model comparison part, I've produced the stylized images of the Magenta model [3]. Dawei and I both focused on the implementation of the web application integrated the Magenta and AdaIN models, and I implemented the backend of the application. I've also helped to prepare the presentation slides as well as demo video. As for the paperwork, I partially contributed to the progress report and final report.

Dawei Si For the model comparison part, I've produced the stylized images based on optimization-based method [1]. Lingyu and I both focused on the implementation of the web application integrated the Magenta and AdaIN models. I have designed the frontend of the application as well as its backend structure. I have also deployed it on a GCP machine to be used for other people. For the paperwork, I partilly contributed to the progress report.

Yichen Chai As for the model comparison part, I've produced the stylized images of the AdaIN model [2]. For the implementation part, I've implemented the AdaIN model together with Cancan and I mainly focused on the model construction, training steps; I've also partially helped to prepare the presentation slides. For the paperwork, I contributed to the main of progress report and final report.

Cancan Huang For the model comparison part, I produced the test images of fast-forward net [8]. For the implementation part, I implemented the AdaIN model together with Yichen Chai and I mainly focused on the preprocess step and model training step. I've also helped to prepare the presentation slides. For the paperwork, I contributed to the main of progress report, final report and 2-min presentation.