**UNDERGRADUATE RESEARCH OPPORTUNITIES PROGRAMME (UROP)**

**Anything Style Transfer**

Yeoh Jun Ting

*Department of Electrical Engineering, College of Design and Engineering National University of Singapore*

*Semester 1,Academic Year 2023/2024*

**Abstract**

Style transfer aims to render the style of a given image for style reference to another given image for content reference and has been widely adopted in artistic generation and image editing. Existing approaches either apply the holistic style of the style image in a global manner or migrate local colors and textures of the style image to the content counterparts in a pre-defined way. In either case, only one result can be generated for a specific pair of content and style images, which therefore lacks flexibility and is hard to satisfy different users with different preferences. In this project, we aim at a novel strategy termed Any-to-Any Style Transfer to address this drawback, which enables users to interactively select styles of regions in the style image and apply them to the prescribed content regions. In this way, beyond one fixed output, personalizable style transfer is achieved through human-computer interaction. At the heart of our approach lies in (1) a region segmentation module based on Segment Anything, which supports region selection with only some clicks or drawing on images and thus takes user inputs conveniently and flexibly; and (2) an attention fusion module, which converts inputs from users to controlling signals for the style transfer model.

This project also aims to explore the use of multi-style transfer for images, with the goal to apply the ‘best’ style for each of the segmented part of the image. The generated image contains a different style for each segmented part which can be useful for automatically selecting the best styles for an image.

**Preface and Acknowledgments**

Computer Vision is a vast and growing field, and Neural style transfer is an active area that has gathered attention and made significant progress in recent times. To tell the truth, I had zero experience in Computer Vision algorithms before partaking in this UROP project. I had to learn some of the basics such as Deep Learning, ConvNet and the Pytorch library online before I could understand the research papers. Interestingly, I was also taking a relevant module called Image Processing where my advisor Xinchao Wang was the lecturer for.

Despite the difficulty and my inexperience, my advisors were willing to give me a chance to explore this project. I would like to take this occasion to thank my advisors Xinchao Wang and Songhua Liu for the opportunity and exposure to such a wonderful field. Their guidance has made me better appreciate research work and their insights have significantly enhanced my understanding of Neural Style Transfer. Without them, this project would not have been possible.

**Table of contents**

[1 Introduction 5](#_Toc151440249)

[2 Literature Review 6](#_Toc151440250)

[3 Methodology and methods 7](#_Toc151440251)

[4 Findings and data 8](#_Toc151440252)

[5 Data analysis and discussion 11](#_Toc151440253)

[6 Conclusion 13](#_Toc151440254)

[7 References 14](#_Toc151440255)

# 1 Introduction

A picture, especially one that is hand drawn by a human, usually has artistic style that is discernable to a human being. The works of famous artists such as Vicent van Gogh and Pablo Picasso have unique styles that are difficult to replicate by other artists. However, with the help of Deep Learning and Convolutional Neural Networks, the transfer of style from one image to another has been made possible [1]. The algorithm works by accepting a content image and a Style image, which are both fed into the Convolutional Neural Network. Another widely used method starts with a noise image instead of a content image (for a total of three images) and works towards an output image.

A collage of different art pieces

Description automatically generated

Figure 1: This figure was adapted from Gatys et al., 2015 [1]. The original photograph depicting the Neckarfront in Tubingen, Germany, is shown in ¨ A (Photo: Andreas Praefcke). The painting that provided the style for the respective generated image is shown in the bottom left corner of each panel. B The Shipwreck of the Minotaur by J.M.W. Turner, 1805. C The Starry Night by Vincent van Gogh, 1889. D Der Schrei by Edvard Munch, 1893.

With the use of gradient descent, the algorithm aims to minimize both style cost and content cost so that the output image best represents the content and style of the images. Ideally the generated image should also be pleasant to look at. Thus, the algorithm called Neural Style Transfer was born.

The advancement of this algorithm in the Computer Vision field has gathered attention from researchers to develop newer and more advanced techniques for several applications. Some examples include H. Huang *et al*., 2017 [2] which tackles Neural Style Transfer for videos.

The other important piece of algorithm that is used for this project is the segment-anything [3] model. This model allows the user to segment an image into different regions, which means that pixels that represent the same object will be given the same segment class. Segment Anything allows for the implementation of region-based style transfer and multiple style transfer to segments, which are areas which will be explored in this project.

# 2 Literature Review

Besides the crucial papers by Gatys et al., 2015 [1] which introduced Neural Style transfer and A. Kirillov et al., 2023 [3] which introduced the SAM model, we explore research works that are related to the project’s goals.

Similar works have been written that align with the project aims. There are two relevant papers that were written recently. S. Liu et al ,2023 [4] introduce the Any-to-Any style transfer which allows users to customize the results of style transfer through selecting regions for both the content and Style images. The backbone algorithms are the SAM, which segments the image into regions and AdaAttN [5] which is an attention-based style transfer baseline. The use of the attention map makes the algorithm focus on the style areas selected by users and allows the use of a non-2D segmented style region.

Another relevant work by K. Psychogyios et al. ,2023 [6] was published amid working on this project and proposes a similar problem statement. The paper discusses a novel loss function which smoothens the boundaries of segments, which improves the visual appeal especially around boundary pixels.

# 3 Methodology and methods

Neural Style Transfer has one limitation that requires some processing for region-based neural style transfer. Both the input content and style images must be of the same dimension and implicitly the pixels must be represented by a 2D matrix. This means that if segmentation is done by the Segment Anything model, the segmented region has a different dimension and a non-2D shape (Figure 2).

A drawing of a landscape with a blue sky and white lines

Description automatically generated

Figure 2: Segmented region that has a non-2D shape

We explore more rudimentary methods and report their results. The bounding box of the segmented region is first found by locating the upper left and lower right pixels before implementing three methods – shading black method, shading average method and running Neural Style Transfer on an averaged style image.

For the second part of this project which involves multi-style neural transfer, by injecting multiple style images for various regions, we find the ‘best’ style with the least style loss and apply the style transfer to the region. In theory, the image should have multiple styles that best suit each of the segments. Another method to determine the ‘best’ style requires human intervention and requires cycling through each of the styles so that a human can determine and decide on the result. A future extension to boundary smoothing can be applied to further enhance the smoothness of the local style transfer.

# 4 Findings and data

Rudimentary methods were used to fill up the non-segmented space for the bounding box style image, and the image is used as the style image. The content image used is a white noise image and neural style transfer is performed using three methods – shading black (non-segment pixels have an intensity value of zero), shading average method (non- segment pixels take the value of the average of segment pixels), and shading white noise method (non-segment pixels take value of noise). The aim is to use the output image as the new style image for other images.

From Figure 3 it is clear that none of the methods are useful and there are permanent smudges for the unsegmented regions. Although some of the style from segmented the image has leaked to the unsegmented region, it has not blended in well and the output image is not a smooth image. This is because of the way content and style losses are defined where the losses are computed globally, which means that conceptually the existence of these pixels will not allow the algorithm to converge to a clean output image.A collage of images of a starry night

Description automatically generated

Figure 3: Results of region-based style transfer for three methods

The second part of this project looks at finding the best style to apply for a segment of the image. The segment is found by prompting the user to click on a point on the image before using SAM to generate multiple masks. The best mask is used as the segment that will be used for the rest of the algorithm.

After choosing three styles, we run Neural Style Transfer for 1000 iterations for each of the styles to ensure that the algorithm is close to convergence and the output image is highly adapted to the style image. With the information of the style and content loss, the best style is picked by virtue of the style with the minimum style and content loss.

A collage of a dog and a dog

Description automatically generated

Figure 4: Finding Segment and its bounding box using SAM after user’s input click

A collage of images of a dog

Description automatically generated

Figure 5: Output image and losses after running neural style transfer for each of the styles

Thereafter, using some image processing and manipulation, the resulting styled image is pasted onto the original image. In this case, the third style was the best and chosen to be the style that was applied to the dog. The environment which was not part of the segment is unchanged.

A puppy sitting in the grass

Description automatically generatedA dog sitting in the grass

Description automatically generated

Figure 6: Initial Image vs Final Output Image using best style

# 5 Analysis and Discussion

As mentioned above, filling up the unsegmented area of a boundary box style image to obtain a 2D style image does not work. However, one other method may be researched in the future.

The first idea is to use a constrained boundary box which excludes unsegmented pixels. This means that the box will only contain pixels of the segment, which would resolve the problem of not having a 2D style image is resolved. However, the major drawback to the approach is the loss of detail from cropped out pixels. These pixels may contain crucial information for the overall style of the segmented region, which would reduce the effectiveness of region-based neural style transfer.

A drawing of a landscape with a blue sky and white lines

Description automatically generated

Figure 7: Constrained bounding box (green). Shaded pixels outside of green box but inside red box are not considered as part of Style Image

Another method that is very effective and already in use is the Any to Any style [4] model, which allows the user to feed the segment mask too so that only the segmented pixels are considered for the Neural Style transfer.

For the part on multiple style images, future improvements would include boundary smoothening and further testing on multiple segments using the auto segment function from SAM.

Also, the selection of best style by looking at the loss may not be accurate as a lower loss may indicate that the output image is very similar to the input content image. This would mean that not much style transfer has occurred. A better metric for choosing the best style transfer may be useful.

# 6 Conclusion

Style transfer is an interesting technique that transforms mere pictures into artworks. The development of Neural Style transfer for various applications results in meaningful results. With the combination of other technologies such as SAM as showcased in this report, the combinations of use cases are limitless.

# 7 References

[1] L. A. Gatys, A. S. Ecker and M. Bethge, "Image Style Transfer Using Convolutional Neural Networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2414-2423, doi: 10.1109/CVPR.2016.265.

[2] H. Huang et al., "Real-Time Neural Style Transfer for Videos," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 7044-7052, doi: 10.1109/CVPR.2017.745.

[3] Kirillov, Alexander & Mintun, Eric & Ravi, Nikhila & Mao, Hanzi & Rolland, Chloe & Gustafson, Laura & Xiao, Tete & Whitehead, Spencer & Berg, Alexander & Lo, Wan-Yen & Dollár, Piotr & Girshick, Ross., “Segment anything,” arXiv.org, Apr. 05, 2023. https://arxiv.org/abs/2304.02643

[4] S. Liu, J. Ye, and X. Wang, “Any-to-any style transfer: Making Picasso and da vinci collaborate,” arXiv.org, https://arxiv.org/abs/2304.09728 (accessed Nov. 20, 2023).

[5] S. Liu, et al., "AdaAttN: Revisit Attention Mechanism in Arbitrary Neural Style Transfer," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021 pp. 6629-6638.  
doi: 10.1109/ICCV48922.2021.00658

[6] K. Psychogyios, H. C. Leligou, F. Melissari, S. Bourou, Z. Anastasakis and T. Zahariadis, "SAMStyler: Enhancing Visual Creativity With Neural Style Transfer and Segment Anything Model (SAM)," in IEEE Access, vol. 11, pp. 100256-100267, 2023, doi: 10.1109/ACCESS.2023.3315235.