

Neural Style Transfer

Zhengyuan Li

University of Illinois Urbana-Champaign

zli138@illinois.edu

Yuxuan Liu

University of Illinois Urbana-Champaign

yuxuan38@illinois.edu

Yifan Shen

University of Illinois Urbana-Champaign

yifan26@illinois.edu

Tiansu Chen

University of Illinois Urbana-Champaign

tiansuc2@illinois.edu



Figure 1. A demo of style transfer from [7]. The top left image is the content image. The style reference image for each synthesized one is shown on the top right corner.

1. Motivation

Neural Style transfer aims to synthesize a new image with its style and content from different reference images. Recently, researchers [3] found that certain features in image segmentation neural networks correspond to the style and content semantically. Afterward, many methods [2, 5, 6] have been proposed to generate more appealing artistic images in more efficient ways. Our goal is to understand this task through experiments and try to improve on the SOTA methods.

Our work will have two goals in parallel. The first one is to explore extensions and modifications based on existing methods, which could be risky. The second part is to run the code of existing methods on novel data sets, which is more practical. In this case, our maximum goal is to design a new method to create artistic images by neural style transfer, which should be comparable to current state-of-

the-art methods both qualitatively and quantitatively. Our minimum goal is to use existing methods in a new data set and analyze its performance.

2. Relationship to Our Background

Among our four members, Zhengyuan is working on other computer vision topics. He is familiar with deep learning frameworks and general techniques. But he has not done research on this specific topic. Yifan does not have experience with computer vision regarding image, but he has some other experience like dataset distillation. Yuxuan has not previously participated in computer vision related research. For him, this is an expedition into a new field of computer science. He is familiar with the general principles of software engineering, and will be participating in the research efforts with an engineering perspective. Tiansu also has no experience in computer vision, but he worked on several NLP

research topics before and is familiar with common deep learning frameworks.

3. Resources

We plan to build our code on the publicly available code of a SOTA method [2]. We will modify different components of their method and try to improve on it. We are also going to adapt their code to novel data sets.

Conceptually, any image set that contains artistic images can be the style dataset. We plan to use a data set¹ on Kaggle as our style dataset. For the content dataset, we prefer realistic photos².

4. Reservations

We expect that adapting the code to novel datasets is a problem of time and is mainly dirty work. Also, we could utilize existing code as baselines, which should not be hard. However, since the method itself is powerful, achieving performance improvement could be hard.

The most difficult part of our project should be the hyperparameter tuning. Neural Style Transfer involves many hyperparameters, such as learning rates, regularization parameters, feature weights. These hyperparameters will impact the quality of the generated images significantly. Also, we need to avoid overfitting, which can result in images that have a distorted or unnatural appearance. Thus, finding the appropriate hyperparameters can be a challenging and time-consuming process, and it may require extensive experimentation and testing.

5. Member roles

We divide our project into four parts.

The first one is about research and literature review. Zhengyuan will focus on researching the latest developments in Neural Style Transfer, including any new techniques, architectures, or optimization algorithms that have been proposed in recent literature.

The second part is data collection and preprocessing: Yifan and Tiansu will focus on collecting and preprocessing the data for the project. This includes finding suitable style and content images, resizing and normalizing the images, and preparing the data for use in the model.

The third part is model development and training: Yifan, Zhengyuan, and Yuxuan will focus on developing and training the Neural Style Transfer model. This involves selecting an appropriate architecture, tuning hyperparameters, and running experiments to find the best model configuration.

¹<https://www.kaggle.com/datasets/thedownhill/art-images-drawings-painting-sculpture-engraving>

²<https://www.kaggle.com/datasets/tompaulat/modern-architecture-100k-small-images>

Last part is evaluation and analysis: Yuxuan and Tiansu will focus on evaluating the results of the Neural Style Transfer model and conducting a thorough analysis of the model's performance. This includes quantitative analysis of the model's output, as well as qualitative analysis of the artistic quality of the generated images. And all of our four members will write the report when our project is finished.

All of our four members will schedule regular meetings once a week, to discuss the project's progress and any issues that have come up. What's more, we will set up a shared document, where everyone can contribute their ideas, notes, and updates. All of us can see what's been done and what still needs to be completed on this document. Last, we will use Github to collaborate the coding work. This allows every member to work on their part simultaneously without interfering with each other's work.

6. Video Style Transfer

Extending single-frame style transfer to videos requires incorporating the temporal relation between frames to mitigate artifacts. A commonly used prior is optical flow. The estimation of optical flow is indifferentiable, and thus we can compute the per-pixel loss between the current frame and the wrapped previous frame. We can write the loss as in [4]:

$$L(\hat{x}^t, \hat{x}^{t-1}) = \frac{1}{D} \sum_{k=1}^D c_k (\hat{x}_k^t - f(\hat{x}_k^{t-1}))^2$$

where \hat{x}^t and \hat{x}^{t-1} are the stylized frames at adjacent time steps. f is the function that wraps the frame with a pre-computed optical flow. D is the dimension of the output and c is the per-pixel confidence of the optical flow.

Now the problem is how to use this loss. One way [8] is to directly optimize the generated result, which could lead to slow generation and high quality. Another way [1, 4] is to train a network in order to generate temporally consistent results. We assume that per-frame generation can already give satisfactory results. In this case, we can train an additional RNN to do corrections in order to synthesize temporally coherent videos.

References

- [1] Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. Coherent online video style transfer. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1105–1114, 2017. 2
- [2] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11326–11336, 2022. 1, 2

- [3] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016. [1](#)
- [4] Haozhi Huang, Hao Wang, Wenhan Luo, Lin Ma, Wenhao Jiang, Xiaolong Zhu, Zhifeng Li, and Wei Liu. Real-time neural style transfer for videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 783–791, 2017. [2](#)
- [5] 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](#)
- [6] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. *Advances in neural information processing systems*, 30, 2017. [1](#)
- [7] Yingnan Ma, Chenqiu Zhao, Xudong Li, and Anup Basu. Rast: Restorable arbitrary style transfer via multi-restoration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 331–340, 2023. [1](#)
- [8] Manuel Ruder, Alexey Dosovitskiy, and Thomas Brox. Artistic style transfer for videos. In *Pattern Recognition: 38th German Conference, GCPR 2016, Hannover, Germany, September 12-15, 2016, Proceedings 38*, pages 26–36. Springer, 2016. [2](#)