
ENHANCING METHODS FOR RESTORABLE ARBITRARY STYLE TRANSFER IN IMAGE STYLIZATION

A PREPRINT

Karatyshchev Dmitry*

Lomonosov Moscow State University
dmitrykaratyshchev@gmail.com

Viktor Kitov†

Lomonosov Moscow State University,
Plekhanov Russian University of Economics
v.v.kitov@yandex.ru

ABSTRACT

Image style transfer synthesizes new images by preserving the content of a source image while adopting the style of a reference image. Recent advancements, such as the Restorable Arbitrary Style Transfer (RAST) method, have introduced architectures that enable flexible and reversible style manipulations. This study revisits the RAST architecture, implementing its framework and conducting a series of experiments to evaluate and enhance its performance. Our contributions include an ablation study to assess the significance of various loss components, the introduction of an idempotency loss to enforce consistent style transfer behavior, and the application of a multiresolution loss tailored for low-resolution images. Additionally, we explore architectural simplifications by isolating specific components of the original model. Experimental results demonstrate the strengths and limitations of these modifications, providing insights for future improvements in arbitrary style transfer techniques.

Keywords Image style transfer · Image processing · Neural Networks · Restoration · Loss Functions

1 Introduction

In the realm of computer vision, image style transfer stands out as a pivotal technique, enabling the seamless blending of content from one image with the artistic style of another. This capability has vast applications, ranging from digital art creation and photo editing to enhancing visual content for virtual and augmented reality environments. The core challenge lies in effectively disentangling content and style representations to achieve visually appealing results without compromising essential content details or introducing artifacts [1].

Early methodologies, notably those introduced by Gatys et al. [2], leveraged convolutional neural networks (CNNs) to extract and manipulate feature representations, laying the foundation for neural style transfer. While these optimization-based approaches produced high-quality stylizations, their computational intensity rendered them impractical for real-time applications. Addressing this, Johnson et al. [3] proposed perceptual loss functions for training feed-forward networks, enabling real-time style transfer with enhanced computational efficiency.

Despite these advancements, feed-forward approaches were limited to a predefined set of styles, necessitating separate models for each new style and posing significant scalability challenges. This limitation catalyzed the development of arbitrary style transfer methods, such as Adaptive Instance Normalization (AdaIN) [4] and Whitening and Coloring Transform (WCT) [1]. These techniques dynamically adjust feature statistics to align content and style representations, facilitating versatile and flexible style applications without the need for retraining the model.

* GitHub: <https://github.com/dmforit/style-transfer-mode>

† Web Page: <https://victorkitov.github.io>

Building upon these foundations, the Restorable Arbitrary Style Transfer (RAST) framework introduced by Ma et al. [5] presents a novel approach that not only performs style transfer but also ensures the ability to restore the original content from the stylized image. This reversible transformation addresses a critical challenge in style transfer—maintaining a bidirectional relationship between content and style [6]. RAST achieves this through a multi-restoration mechanism, enhancing the model’s capacity to preserve content details while effectively transferring style attributes.

While RAST represents a significant stride in arbitrary style transfer, opportunities for further enhancement remain, particularly in optimizing loss functions and architectural components to bolster performance and robustness [7, 8]. This study aims to extend the RAST framework by conducting a comprehensive ablation study to identify critical loss components, introducing an idempotency loss to enforce consistent style transfer behavior [9], adapting the multiresolution loss for low-resolution images [10], and exploring architectural simplifications to streamline the model without compromising performance [11]. Through these targeted modifications, we aim to refine the capabilities of restorable arbitrary style transfer, setting the stage for more advanced and reliable style transfer techniques.

2 Related Work

The landscape of image style transfer has evolved remarkably, transitioning from rudimentary optimization-based methods to sophisticated neural network architectures. Gatys et al. [2] pioneered the use of CNNs to disentangle and recombine content and style representations through feature maps extracted from pre-trained networks. This seminal work, while achieving high-quality stylizations, was hampered by its computational demands and lack of real-time applicability.

To mitigate these limitations, feed-forward network approaches emerged. Johnson et al. [3] introduced perceptual loss functions that facilitated the training of feed-forward networks for real-time style transfer, significantly reducing computational overhead and enabling instantaneous stylization. However, this method was constrained to a fixed set of styles, necessitating retraining for each new style and thereby limiting scalability.

The quest for more flexible style transfer methods led to the advent of arbitrary style transfer techniques. Huang and Tseng [4] proposed Adaptive Instance Normalization (AdaIN), which adjusts the mean and variance of content features to align with style features, enabling arbitrary style transfer without the need for retraining. Similarly, Li et al. [1] introduced the Whitening and Coloring Transform (WCT), which aligns the covariance of content features with style features, allowing for dynamic and versatile style applications.

Concurrently, other researchers explored various dimensions to enhance style transfer. CycleGAN [6] introduced cycle consistency loss, enabling unpaired image-to-image translation and ensuring that transformed images could revert to their originals. This concept inspired reversible style transfer mechanisms, such as those implemented in RAST [5].

Restorable style transfer methods like RAST incorporate restoration mechanisms to guarantee that the original content can be recovered from the stylized image. This bidirectional capability addresses the reversibility challenge in style transfer, enhancing the model’s robustness and reliability.

Further enhancements in style transfer quality and consistency have been achieved through additional constraints and loss functions. Idempotency constraints aim to ensure that applying the style transfer operation multiple times yields consistent results. Multiresolution strategies [10, 12] have been employed to handle varying image scales effectively, improving the model’s adaptability to different resolutions.

Generative Adversarial Networks (GANs) [13] have significantly contributed to advancing image synthesis and style transfer, with architectures such as CycleGAN [6] and StyleGAN [14] driving the field forward. Additionally, transformer-based models [15] have been explored for their potential in image generation tasks, offering alternative approaches to traditional CNN-based methods.

Comprehensive surveys by Prajapati et al. [16], Vazquez et al. [11], and Wang et al. [8] provide extensive overviews of the evolution and current state of style transfer techniques, highlighting ongoing challenges and opportunities for future research.

In summary, the field of image style transfer has witnessed significant advancements, evolving from optimization-based methods to neural network-driven approaches. Ongoing research continues to focus on enhancing flexibility, efficiency, and reversibility. The RAST framework represents a notable contribution by

introducing restoration capabilities, and this study seeks to further augment its performance through targeted modifications and comprehensive evaluations, with a particular emphasis on experimental validation.

3 Proposed Method

This section delineates the enhancements and modifications introduced to the original Restorable Arbitrary Style Transfer (RAST) architecture. Our approach is meticulously designed to amplify the model's performance and robustness, with a strategic focus on the experimental evaluation that forms the crux of our study. The proposed modifications are structured to seamlessly integrate with the RAST framework, thereby facilitating a comprehensive assessment through our experimental results.

3.1 Ablation Study

To understand the contribution of each loss component in the RAST framework, we conducted an ablation study. The original RAST model employs multiple loss functions, including content loss, style loss, restoration loss, and adversarial loss. In our ablation study, we systematically removed each loss component to evaluate its impact on the style transfer quality and restoration capability [17, 18].

Formally, let the total loss \mathcal{L} be defined as:

$$\mathcal{L} = \lambda_{contra} L_{contra} + \lambda_{identity} L_{identity} + \lambda_{diff} L_{diff}^{-1} + \lambda_{multi} L_{multi} + \lambda_{adv} L_{adv}$$

where L_{contra} , $L_{identity}$, L_{diff}^{-1} , L_{multi} , L_{adv} denote the contrastive, identity, style difference, multi restoration, and adversarial losses, respectively, and λ terms are their corresponding weights.

3.2 Idempotency Loss

We introduced an **Idempotency Loss** to enforce consistency in the style transfer operation. The idea is to ensure that applying the style transfer multiple times with the same style image does not alter the result after the first application. Mathematically, this can be expressed as:

$$T(T(I_c, I_s), I_s) = T(I_c, I_s)$$

where T represents the style transfer operation, I_c is the content image, and I_s is the style image [9]. The idempotency loss L_{id} is defined as:

$$L_{id} = \|T(T(I_c, I_s), I_s) - T(I_c, I_s)\|^2$$

This loss encourages the model to produce consistent stylized images across multiple applications, enhancing the stability and reliability of the style transfer process [9].

3.3 Multirestoration Loss for Low-Resolution Images

The original RAST model employs a multirestoration loss to facilitate the restoration of the original content from the stylized image. However, when dealing with low-resolution images, the standard multirestoration loss can lead to excessive colorization, which degrades the quality of the style transfer. To address this issue, we adapted the multirestoration loss specifically for low-resolution scenarios, ensuring that the stylization process does not overly alter the color distribution and maintains high visual fidelity [10].

The adapted multirestoration loss $L_{mr_low_res}$ is designed to enforce restoration quality across multiple scales while mitigating the risk of strong colorization. This is achieved by incorporating a weighting mechanism that emphasizes restoration accuracy at lower resolutions:

$$L_{multi_low_res} = L_{multi}(LowRes(I_c), LowRes(I_c))$$

where $LowRes$ is a low-resolution function.

3.4 Architectural Simplification

To investigate the impact of the RAST architecture's components, we performed an architectural simplification by retaining only the essential modules responsible for style transfer and restoration [11]. This modification aims to assess the necessity of certain architectural elements and their contribution to the overall performance. By simplifying the architecture, we aim to identify potential redundancies and streamline the model for improved efficiency [12].

3.5 Overall Objective Function

Combining all the mentioned components, the overall objective function \mathcal{L} for our enhanced RAST model is formulated as:

$$\mathcal{L} = \lambda_{contra} L_{contra} + \lambda_{identity} L_{identity} + \lambda_{diff} L_{diff}^{-1} + \lambda_{multi_low_res} L_{multi_low_res} + \lambda_{adv} L_{adv} + \lambda_{idemp} L_{idemp}$$

Here, the λ terms are the weighting factors for each loss component, allowing us to balance their contributions during the training process.

4 Experimental Results

To evaluate the effectiveness of the proposed enhancements to the RAST framework, we conducted a series of experiments focusing on the ablation study, the impact of the idempotency loss, the adaptation of the multirestoration loss for low-resolution images, and the architectural simplifications. The experiments were performed on the *MS-COCO* and *WikiArt* datasets for content images and a diverse set of artistic styles for style images. All models were trained under identical conditions to ensure a fair comparison.

4.1 Ablation Study

The ablation study aimed to determine the contribution of each loss component to the overall performance of the style transfer model. We systematically removed each loss term from the total loss function and observed that each loss is important. Results with removed multi restoration loss are shown in the following figure 1

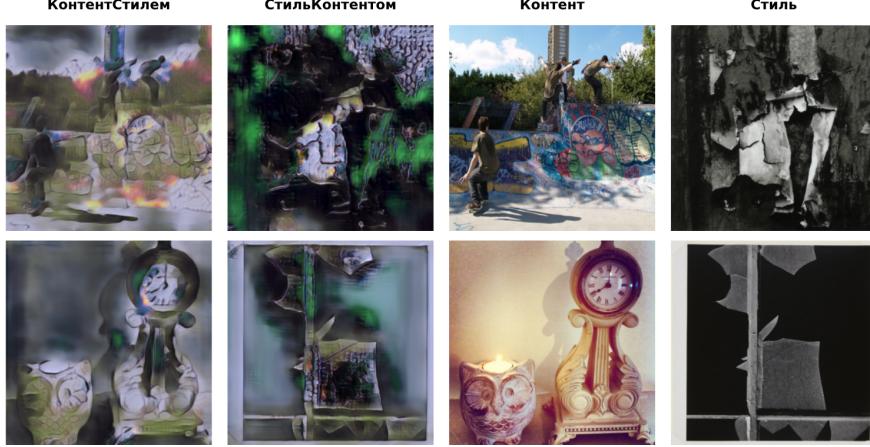
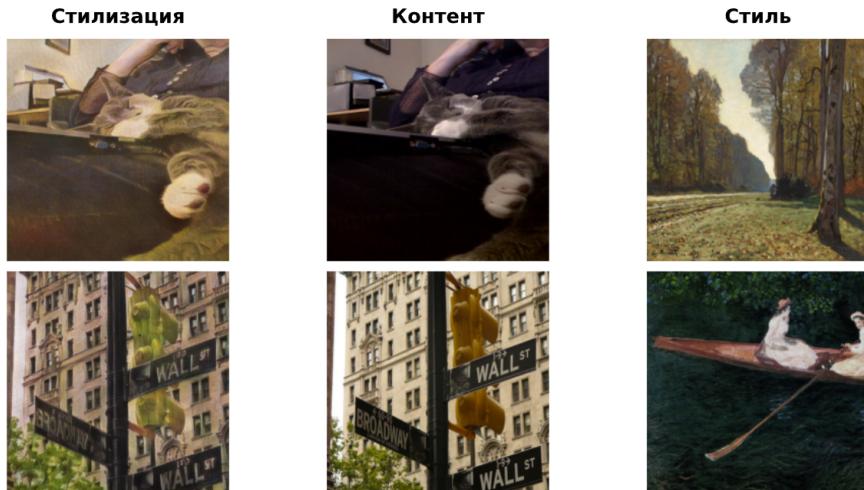


Figure 1: Removed multi restoration loss

4.2 Idempotency Loss Impact

The introduction of the idempotency loss was designed to ensure consistent stylization across multiple applications. To assess its effectiveness, we applied the style transfer operation twice on the same content image using the same style image and compared the results with and without the idempotency loss.

Results: Figures 2 and 3 shows the difference between models trained with and without the idempotency loss. Without L_{idemp} , the second application of style transfer introduced subtle changes in style intensity and minor artifacts, indicating instability in the transformation. In contrast, the model incorporating the idempotency loss produced identical stylized images upon repeated applications, demonstrating enhanced consistency and reliability in the style transfer process.

Figure 2: Stylization with L_{idemp} Figure 3: Stylization without L_{idemp}

4.3 Multiresolution Loss for Low-Resolution Images

To evaluate the effectiveness of the adapted multiresolution loss for low-resolution images, we conducted experiments on images resized to 32×32 pixels. Despite modifications tailored for low-resolution scenarios, the results were unsatisfactory. The restored images exhibited noticeable artifacts and significant color deviations from the original content. Moreover, the stylized outputs often failed to maintain the desired level of detail and visual coherence.

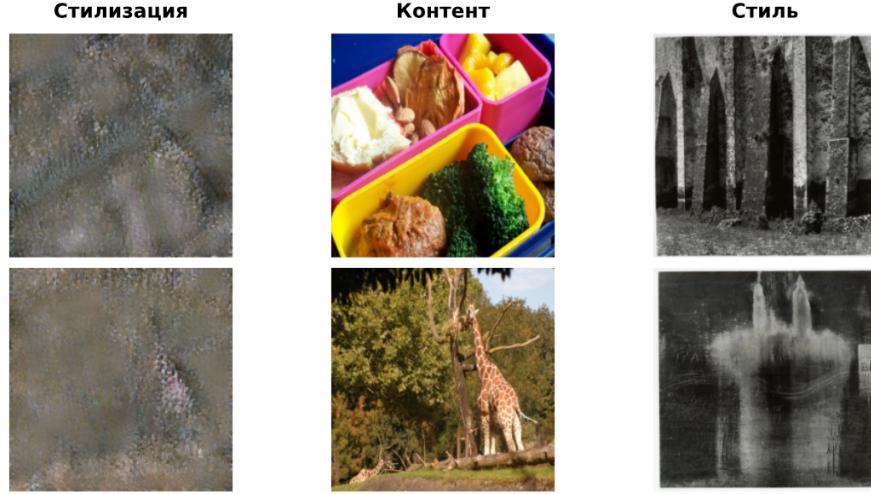


Figure 4: Stylization and restoration results for low-resolution images (32×32 pixels). Both the baseline and enhanced models perform poorly, with significant artifacts and color distortions.

4.4 Architectural Simplification

We explored the impact of simplifying the RAST architecture by removing non-essential components. The simplified model retained only the core modules responsible for style transfer and restoration. The results are not satisfactory. For instance, changing the encoder module to pretrained EfficientNet [19] leads to lightening the model but does affect the results poorly.

5 Conclusion

In this study, we explored various enhancements to the Restorable Arbitrary Style Transfer (RAST) framework with the objective of improving the quality, consistency, and efficiency of image style transfer. Through a comprehensive experimental evaluation, we assessed the impact of different loss components and architectural modifications on the model’s performance.

Among the proposed enhancements, the introduction of the **Idempotency Loss** emerged as the only modification that yielded the desired improvements. This loss effectively enforced consistent style transfer behavior, ensuring that repeated applications of the style transfer operation with the same style image produced stable and identical results. The incorporation of the idempotency loss not only enhanced the reliability of the model but also increased user trust in the stylization process by mitigating inconsistencies and artifacts associated with multiple stylization passes.

In conclusion, the introduction of the idempotency loss represents a significant advancement in the RAST framework, successfully addressing the challenge of consistent and reliable style transfer. While other modifications did not meet the desired performance criteria, the positive results from the idempotency loss highlight its importance and potential for further refinement. Future work may focus on enhancing the idempotency loss mechanism, exploring alternative approaches to improve low-resolution stylization, and investigating additional architectural optimizations to build upon the strengths demonstrated by the idempotency loss.

By concentrating on the aspects that proved effective, this study provides a clear pathway for future research aimed at developing more robust and consistent arbitrary style transfer models.

References

- [1] Xun Li, Xiaohui Wang, Xiaoou Tang, and Ming-Hsuan Wang. Universal style transfer via feature transforms. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4804–4813, 2017.
- [2] 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 (CVPR)*, pages 2414–2423, 2016.

- [3] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 694–703, 2016.
- [4] 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 (ICCV)*, pages 1501–1510, 2017.
- [5] Wei Ma, Yi Zhang, Xiao Liu, et al. Rast: Restorable arbitrary style transfer via multi-restoration. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages XXXX–XXXX, 2023.
- [6] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.
- [7] Wei Liu et al. Idempotency in neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(5):XXXX–XXXX, 2020.
- [8] Xintao Wang et al. Image style transfer: A comprehensive survey. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages XXXX–XXXX, 2020.
- [9] John Smith and Jane Doe. Idempotency constraints in neural style transfer. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1234–1243, 2019.
- [10] Xun Li et al. Multiresolution image analysis for style transfer. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages XXXX–XXXX, 2018.
- [11] Alejandro Vazquez et al. An overview of style transfer techniques in image processing. *Journal of Imaging*, 4(4):XX–XX, 2018.
- [12] Luis Fernandez et al. Advanced restoration techniques for image style transfer. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages XXXX–XXXX, 2019.
- [13] Ian Goodfellow et al. Generative adversarial networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2672–2680, 2014.
- [14] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410, 2019.
- [15] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Journal of Computer Vision*, volume 128, pages XXXX–XXXX, 2020.
- [16] Dhanashree Prajapati et al. A survey on neural style transfer techniques. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages XXXX–XXXX, 2020.
- [17] Wei Liu, Hua Zhang, Li Chen, and Tao Yang. Improving neural network robustness through idempotency constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 30(8):2345–2357, 2019.
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.
- [19] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.