Face-aware Arbitrary Image Style Transfer

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

Arbitrary style transfer techniques have achieved significant results in machine learning, enabling the transformation of input images into artistic outputs while preserving content structures. Although existing methods for style transfer address the challenge of preserving facial details during stylization, they often suffer from noticeable artifacts and fail to maintain identity consistency. Recent advancements, such as RAST (Restorable Arbitrary Style Transfer), have addressed the challenge of maintaining high-fidelity restorability during style transfer by integrating multi-restoration processes. Building upon the foundational improvements introduced by RAST, our work introduces a novel framework for Face-aware Arbitrary Image Style Transfer that specifically enhances the handling of facial regions. Recognizing that existing methods often fail to preserve the details and identity of faces, we propose an architecture that incorporates adaptive facial region detection and targeted feature modulation to maintain facial consistency while applying stylistic transformations. This improvement ensures that essential facial features, such as expressions and identity markers, remain recognizable and coherent in the stylized output. Extensive experiments demonstrate that our enhanced model not only preserves global and local content better but also significantly outperforms baseline methods in perceptual quality, particularly in human-centric images. This work paves the way for more refined and identity-preserving artistic transformations, bridging the gap between arbitrary style transfer and realistic face representation.

Keywords Image style transfer · Image processing

1 Introduction

In recent years, arbitrary style transfer has emerged as a powerful tool in the realm of computer vision and machine learning, enabling the transformation of images into various artistic styles while preserving the underlying content structure. This capability has opened up numerous applications, from digital art creation to enhancing visual effects in multimedia. However, despite the impressive progress in arbitrary style transfer, one of the most challenging aspects remains the preservation of facial details and identity consistency during stylization. Faces, being highly sensitive and complex structures, often suffer from artifacts and distortions that undermine the realism and integrity of the transformed images.

The development of arbitrary style transfer has its roots in early attempts to separate image content and style, drawing inspiration from the field of neuroscience. The foundational work by Gatys et al. [1] in 2015 introduced the concept of using convolutional neural networks (CNNs) to achieve style transfer by optimizing images to match both content and style representations learned from deep layers of pre-trained networks. This breakthrough demonstrated that complex visual patterns could be transferred between images, sparking a surge of interest in the field.

---

[*]GitHub: https://github.com/dmforit/style-transfer-mode
[†]Web Page: https://victorkitov.github.io

Subsequent methods sought to address limitations such as computational inefficiency and the restriction to a limited set of styles. Techniques like Fast Neural Style Transfer, proposed by Johnson et al., used feed-forward neural networks to achieve real-time stylization [2, 3], enabling broader applications. The evolution continued with the introduction of universal style transfer approaches, allowing arbitrary styles to be applied without the need for per-style retraining. These approaches laid the groundwork for more sophisticated methods that improved adaptability and efficiency.

Recent advancements have focused on enhancing the quality and versatility of arbitrary style transfer, including methods that prioritize maintaining structural coherence and integrating multi-level feature transformations. Despite these advances, preserving facial identity during stylization remained a persistent challenge, prompting further exploration and innovation in face-aware algorithms.

Existing methodologies have attempted to address these challenges with varying degrees of success. While some approaches integrate specific constraints or additional processes to maintain facial features, they frequently encounter trade-offs that compromise either the quality of the stylization or the fidelity of the facial identity. Recent innovations, such as the Restorable Arbitrary Style Transfer (RAST) framework [4], have shown that multi-restoration techniques can enhance the overall quality of style transfer, ensuring that critical content can be restored after the stylization process. Yet, even these advanced methods fall short when applied to images where facial detail is paramount, leading to outputs that may lack the nuance necessary for identity preservation.

To bridge this gap, our work presents a novel framework specifically designed for Face-aware Arbitrary Image Style Transfer. This framework incorporates adaptive facial region detection and targeted feature modulation, enhancing the capacity to preserve facial identity and expression integrity throughout the stylization process. By prioritizing the accurate representation of facial features and identity markers, our method ensures that the stylized output remains coherent and recognizable, even in highly stylized transformations. Our extensive experiments validate that this approach not only maintains global and local content but also delivers superior perceptual quality compared to existing state-of-the-art methods, particularly in scenarios involving human-centric imagery. This advancement sets a new benchmark for identity-preserving style transfer and paves the way for more realistic and expressive artistic transformations.

2   Related work

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. See Section 2.

2.1   Headings: second level

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})} \tag{1}$$

3   Proposed method

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio,

commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

4 Experimental results

Nulla ac nisl. Nullam urna nulla, ullamcorper in, interdum sit amet, gravida ut, risus. Aenean ac enim. In luctus. Phasellus eu quam vitae turpis viverra pellentesque. Duis feugiat felis ut enim. Phasellus pharetra, sem id porttitor sodales, magna nunc aliquet nibh, nec blandit nisl mauris at pede. Suspendisse risus risus, lobortis eget, semper at, imperdiet sit amet, quam. Quisque scelerisque dapibus nibh. Nam enim. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Nunc ut metus. Ut metus justo, auctor at, ultrices eu, sagittis ut, purus. Aliquam aliquam.

5 Conclusion

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetuer quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

References

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

[2] 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.

[3] Yijun Li, Chen Fang, Jimei Yang, Xiaohui Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. *arXiv preprint arXiv:1705.08086*, 2017.

[4] 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 (WACV)*, pages 331–340, January 2023.