Approaches for Local Artistic Control of Mobile Neural Style Transfer

Max Reimann
Hasso Plattner Institute for Digital Engineering, University of Potsdam

Mandy Klingbeil
Digital Masterpieces GmbH

Sebastian Pasewaldt
Digital Masterpieces GmbH

Amir Semmo
Hasso Plattner Institute for Digital Engineering, University of Potsdam

Jürgen Döllner
Hasso Plattner Institute for Digital Engineering, University of Potsdam

Matthias Trapp
Hasso Plattner Institute for Digital Engineering, University of Potsdam

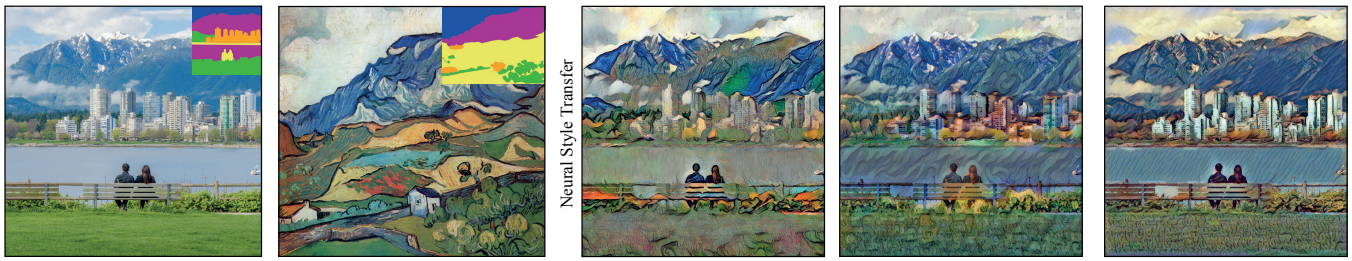

Figure 1: Implemented approaches for location-based filtering of NSTs. Content image © karamysh, used with permission[1].

ABSTRACT

This work presents enhancements to state-of-the-art adaptive neural style transfer techniques, thereby providing a generalized user interface with creativity tool support for lower-level local control to facilitate the demanding interactive editing on mobile devices. The approaches are implemented in a mobile app that is designed for orchestration of three neural style transfer techniques using iterative, multi-style generative and adaptive neural networks that can be locally controlled by on-screen painting metaphors to perform location-based filtering and direct the composition. Based on first user tests, we conclude with insights, showing different levels of satisfaction for the implemented techniques and user interaction design, pointing out directions for future research.

CCS CONCEPTS

• **Computing methodologies** → **Non-photorealistic rendering**; **Image processing**;

KEYWORDS

neural style transfer, mobile devices, local artistic control

ACM Reference format:
Max Reimann, Mandy Klingbeil, Sebastian Pasewaldt, Amir Semmo, Jürgen Döllner, and Matthias Trapp. 2018. Approaches for Local Artistic Control of Mobile Neural Style Transfer In *Proceedings of The Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering, Victoria, BC, Canada, August 17–19, 2018 (Expressive '18)*, 2 pages.
DOI: 10.1145/3229147.3229188

Expressive '18, August 17–19, 2018, Victoria, BC, Canada

1 BACKGROUND

Mobile expressive rendering had a relevant share in the development of a new user group that discovered the creation of art to be fascinating: *mobile artists*; a group of serious hobbyists that are eager to adapt powerful and flexible tools that enable new ways of expressing their creativity. To learn more about their requirements, secondary market research, task analysis, user surveys and interviews have been conducted to help design an interactive prototype that fosters creativity, increases productivity and establishes a high user satisfaction when using effects based on a neural style transfer (NST). Important requirements involve: an easy exploration of tools and effects, versatile effect presets, the ability to combine styles and define own styles, and fine-grained artistic control of a style transfer. However, current mobile apps are limited in the number of provided styles and typically lack local control over the style transfer, making their usage unattractive to artists [Semmo et al. 2017]. In this work, we present an iOS app that implements and enhances style transfer technologies to allow for local creative control that facilitates an interactive, artistic image editing (Figure 1). Our app targets mobile artists with basic image editing know-how by using established on-screen painting metaphors for the local definition of sub-styles and the successive application to content images.

[1] http://www.shutterstock.com/image-photo/sea-walk-kitsilano-beach-park-downtown-158669972

2 MOBILE NEURAL STYLE TRANSFER APP

Our mobile app implements and extends three different NST techniques, each providing a trade-off between flexibility, picture quality and run-time. We adapt the iterative NST of Gatys et al. [2016], which defines style transfer as an optimization problem of finding a stylized target image, whose content is similar to a content image and whose style is similar to a style image in terms of summary statistics over activations of a deep neural network. This approach is extended using using semantic masks for style loss calculations similar to [Luan et al. 2017], which is implemented as a web service as its computational complexity does not allow on-device processing. To enable near interactive performance mobile NST apps typically use single-style feed-forward networks [Johnson et al. 2016] which are trained offline and applied on-device in a single forward pass. By contrast, we propose to use multi-style networks such as the *MSG*-Net [Zhang and Dana 2017] to enable local control, particularly because they smoothly blend styles and reduce training/inference times against using multiple single-style networks. At this, we extend the *MSG*-Net training procedure using a loss term based on semantic masks to learn a decoder that is aware of spatially varying styles. Although *MSG*-Nets can learn a large amount of styles, the computational burden for training new styles is still high and therefore does not permit end-users to add custom styles. To cope with this limitation, we implement a third technique that enables a transfer of arbitrary, new styles by using an encoder-decoder network containing an adaptive instance normalization layer (*adaIn* [Huang and Belongie 2017]). Similar to the *MSG*-Net, we enable local control by merging different styles in feature space after performing *adaIn*.

Compared to the global approaches, we are able to increase the NST quality through plausible semantic matching but with a reduced performance—slower by factor 2-5, depending on the number of sub-styles—, thus hindering an interactive editing. To this end, we propose interactive modifications with a deferred transfer using multi-style feed-forward and adaptive NSTs, which are also useable as visual aids for the server-sided iterative NST approach to synthesize high-quality artistic renditions offline.

2.1 User Interface and Implementation

The first prototype was based on design choices to achieve a high learnability of the system and the reduction of error sources. We decided to focus the usability design on the definition of sub-style brushes, since this approach addresses the most complex challenges that need to be solved to clearly communicate functionalities to end-users. These include: (a) how to define sub-style brushes as a user, (b) how to visualize pre-defined and user-defined sub-style brushes of a style image and (c) how to apply different sub-style brushes to a content image. The app provides style images with pre-defined sub-style brushes for all three techniques, visualized through color mappings. Users can draw masks to define new brushes and then draw content masks to spatially direct the style application (Figure 2). In addition, we propose a two-tiered approach to ensure an interactive editing and a better imagination of the stylized output: first, users blend global stylizations in image space using interactive painting metaphors to preview the stylization; second, they can perform an on-demand stylization with the described feature space merging using either *MSG* or *adaIn* networks.

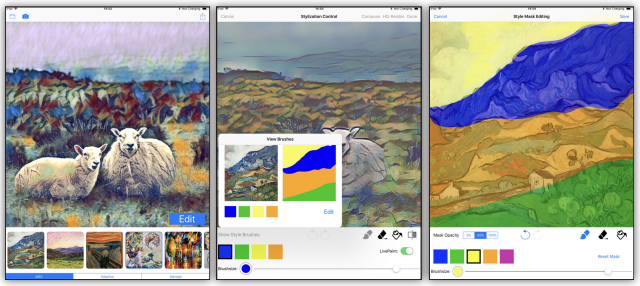

Figure 2: App screens for selecting a NST technique and locally applying (sub-) styles to a content image. Content image © Rick Barrett on Unsplash.com, used with permission.

The iterative and feed-forward NSTs are implemented using PyTorch and CoreML for iOS respectively. We implemented custom *CoreML* layers for switching style layers (*MSG*) and adaptive networks using *Metal*. The performance depends on the image resolution and number of masks: processing an 720^2 image with two masks takes 1.0 second (*adaIn*) and 1.5 seconds (*MSG*) on an iPad Pro 10.5", and 3 minutes for the iterative NST (GTX-1080TI GPU).

2.2 Results

Initial results indicate that the interactive editing approaches lead to more plausible texture and color transfers, in particular for the iterative and *MSG* methods. The adaptive approach generally lacks to transfer complex structures, however a clear separation between the different sub-styles could be achieved using local control.

A first usability study was conducted with six participants who where asked to tune a globally stylized result. All participants were convinced that the local controls improved the overall style quality, but found that the color mappings between masks and styles caused confusion as many associated them with a plain image colorization. The majority also liked the interactive preview more than purely drawing content masks, but frequently switched between both modes to detect uncovered image regions, indicating that this interaction metaphor can be further improved. Generally, the majority of users wished for more semantically meaningful symbols to represent the styles during definition and application.

ACKNOWLEDGMENTS

This research was funded by the Federal Ministry of Education and Research (BMBF), Germany, for the AVA project 01IS15041.

REFERENCES

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. In *Proc. CVPR*. IEEE Computer Society, Los Alamitos, 2414–2423. https://doi.org/10.1109/CVPR.2016.265

Xun Huang and Serge Belongie. 2017. *Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization.* arXiv.org report 1703.06868. https://arxiv.org/abs/1703.06868

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Proc. ECCV*. Springer International, Cham, Switzerland, 694–711. https://doi.org/10.1007/978-3-319-46475-6_43

Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. 2017. *Deep Photo Style Transfer.* CoRR abs/1703.07511. http://arxiv.org/abs/1703.07511

Amir Semmo, Tobias Isenberg, and Jürgen Döllner. 2017. Neural Style Transfer: A Paradigm Shift for Image-based Artistic Rendering?. In *Proc. NPAR*. ACM, New York, 5:1–5:13. https://doi.org/10.1145/3092919.3092920

Hang Zhang and Kristin Dana. 2017. *Multi-style Generative Network for Real-time Transfer.* arXiv.org report 1703.06953. https://arxiv.org/abs/1703.06953