Cartoon style transfer based on DPST

Chen Zehua, 31520231154285, xinyuan class
Li Weicong, 31520231154297, xinyuan class
Nie Wenjie, 31520231154307, xinyuan class
Mo Songyin, 31520231154306, xinyuan class
Wang Hengbo, 30920231154362, xinyuan class

Abstract

Style transfer technology brings a unique perspective to the field of image processing, allowing the synthesis of attractive images that blend different image styles while retaining the content characteristics of the original image. The original intention of this research is to use style transfer technology to achieve the synthesis of cartoon style and real-life pictures, thereby creating visually attractive scenery photos while retaining the scenery characteristics of the original pictures. In the exploration of this area, previous research has focused on style transfer experiments on realistic landscape image datasets. This study focuses on converting the data set to a cartoon atlas to observe the applicability of the model under different datasets. This attempt not only expanded the application scope, but also explored the performance of the style transfer model on the cartoon image datasets. We retrain the model on the cartoon dataset and observe its generalizability. Next, the results were compared with a model specifically designed for cartoon style transfer. Through these experiments, we evaluate the model's performance on new datasets and explore its ability to adapt across different styles of images. This research is of great significance for promoting the application of style transfer technology in the field of image synthesis. The experimental methods and results of this study are inspiring for future image processing and style transfer research, and provide useful reference and guidance for exploring a wider range of applications.

Introduction

The realm of style transfer in photography has opened new avenues for creative image synthesis by seamlessly blending styles from various images, resulting in visually appealing compositions while retaining the essence of the original scenes. This transformative technique has sparked immense interest, particularly in the field of computer vision and artificial intelligence, for its potential in reshaping visual narratives. Our proposed research venture embarks on a compelling exploration in the domain of photographic style transfer, aiming to leverage the innovative fusion of styles from diverse datasets to enhance the aesthetic appeal of landscape photographs.

Motivation

The motivation behind our research endeavor stems from the profound impact that style transfer techniques can have on the visual aesthetics of photographs. By borrowing styles from disparate sources, such as cartoons, we envision the creation of captivating landscape images that transcend traditional boundaries. The allure lies in the ability to synthesize visually striking scenes while preserving the unique characteristics of the original photographs. This juxtaposition of realism and artistic interpretation opens doors to novel forms of visual storytelling, making our exploration not only technologically intriguing but also artistically inspiring.

Innovation

Building upon the foundations laid by previous research, our innovative approach diverges from conventional studies by delving into unexplored territory. While existing literature primarily focuses on style transfer experiments within the realm of real-world scenery datasets, our research takes a bold step forward. We propose a paradigm shift by transposing the experimentation onto a dataset comprising cartoons and animated illustrations. This pioneering shift allows us to observe how well-established style transfer models perform when confronted with the distinct challenges posed by cartoonish imagery. By venturing into this uncharted domain, we aim to uncover the intricacies of translating the essence of cartoons into real-world landscapes, offering unique insights into the intersection of reality and artistic interpretation.

In summary, our proposed research not only aligns with the current trends in computational photography but also pushes the boundaries of creativity by exploring uncharted territories within the realm of style transfer. By harnessing the power of artificial intelligence to blend realism with artistic styles, we aspire to contribute significantly to the evolving landscape of visual storytelling and redefine the possibilities within the realm of digital imagery.

Related Work

Global style transfer algorithms process an image by applying a spatially-invariant transfer function. These methods are effective and can handle simple styles like global color shifts

(e.g., sepia) and tone curves (e.g., high or low contrast). For instance, Reinhard et al.(Reinhard et al. 2001) match the means and standard deviations between the input and reference style image after converting them into a decorrelated color space.Pitié et al.(Pitie, Kokaram, and Dahyot 2005) describe an algorithm to transfer the full 3D color histogram using a series of 1D histograms. As we shall see in the result section, these methods are limited in their ability to match sophisticated styles.

Local style transfer algorithms based on spatial color mappings are more expressive and can handle a broad class of applications such as time-of-day hallucination(Shih et al. 2013; Gatys, Ecker, and Bethge 2016a), transfer of artistic edits (Bae, Paris, and Durand 2006a; Shih et al. 2014; Sunkavalli et al. 2010), weather and season change(Gardner et al. 2016; Laffont et al. 2014), and painterly stylization(Li and Wand 2016; Selim, Elgharib, and Doyle 2016; Hertzmann et al. 2001). Our work is most directly related to the line of work initiated by Gatys et al. that employs the feature maps of discriminatively trained deep convolutional neural networks such as VGG-19(Simonyan and Zisserman 2014) to achieve groundbreaking performance for painterly style transfer(Li and Wand 2016). The main difference with these techniques is that our work aims for photorealistic transfer, which, as we previously discussed, introduces a challenging tension between local changes and large-scale consistency.In that respect, our algorithm is related to the techniques that operate in the photo realm (Bae, Paris, and Durand 2006a). But unlike these techniques that are dedicated to a specific scenario, our approach is generic and can handle a broader diversity of style images.

In terms of anime style transfer, Chen et al.(Chen, Lai, and Liu 2017) proposed a method to improve comic style transfer by training a dedicated CNN to classify comic/non-comic images. All these methods use a single style image for a content image, and the result have a consistent style.Unlike this, our model can generate images in different styles based on the reference image, providing greater flexibility.CartoonGAN(Chen, Lai, and Liu 2018a) proposes a GAN-based method that utilizes unpaired training sets to transform real-world photos into cartoon images. In this project, the model can also emulate comic images, converting real photos into anime pictures. In the experiments, we will compare the results of generated images.

Method

Our algorithm takes two images: an input image which is usually an ordinary photograph and a cartoonized and retouched reference image, the reference style image. We seek to transfer the style of the reference to the input while keeping the result photorealistic. Our approach augments the Neural Style algorithm(Gatys, Ecker, and Bethge 2016b) by introducing two core ideas.

- We propose a photorealism regularization term in the objective function during the optimization, constraining the reconstructed image to be represented by locally affine color transformations of the input to prevent distortions.
- We introduce an optional guidance to the style transfer process based on semantic segmentation of the inputs(Champandard 2016) to avoid the content-mismatch problem, which greatly improves the photorealism of the results.

For completeness, we summarize the Neural Style algorithm by Gatys et al.(Gatys, Ecker, and Bethge 2016b) that transfers the reference style image S onto the input image I to produce an output image O by minimizing the objective function:

$$\mathcal{L}_{\text{total}} = \sum_{\ell=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_s^\ell \tag{1}$$

$$\mathcal{L}_c^\ell = \frac{1}{2N_\ell D_\ell} \sum_{ij} (F_\ell[O] - F_\ell[I])_{ij}^2 \tag{2}$$

$$\mathcal{L}_s^\ell = \frac{1}{2N_\ell^2} \sum_{ij} (G_\ell[O] - G_\ell[S])_{ij}^2 \tag{3}$$

where L is the total number of convolutional layers and the l-th convolutional layer of the deep convolutional neural network.In each layer, there are N_l filters each with a vectorized feature map of size D_l. $F_l[\cdot] \in \mathbf{R}^{N_l * D_l}$ is the feature matrix with (i, j) indicating its index and the Gram matrix $G_l[\cdot] = F_l[\cdot]F_l[\cdot]^T \in \mathbf{R}^{N_l * N_l}$ is defined as the inner product between the vectorized feature maps. α_l and β_l are the weights to configure layer preferences and Γ is a weight that balances the tradeoff between the $content$ (Eq. 2) and the $sytle$ (Eq. 3)

Formally, we build upon the Matting Laplacian of Levin et al.(Levin, Lischinski, and Weiss 2008) who have shown how to express a grayscale matte as a locally affine combination of the input RGB channels.They describe a least-squares penalty function that can be minimized with a standard linear system represented by a matrix M_I that only depends on the input image I. Note that given an input image I with N pixels, M_I is $N * N$. We name $V_c[O]$ the vectorized version $(N * 1)$ f the output image O in channel c and define the following regularization term that penalizes outputs that are not well explained by a locally affine transform:

$$\mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T \mathcal{M}_I V_c[O] \tag{4}$$

Using this term in a gradient-based solver requires us to compute its derivative w.r.t. the output image. Since M_I is a symmetric matrix, we have: $\frac{\mathrm{d}\mathcal{L}_m}{\mathrm{d}V_c[O]} = 2\mathcal{M}_I V_c[O]$

A limitation of the style term (Eq. 3) is that the Gram matrix is computed over the entire image. Since a Gram matrix determines its constituent vectors up to an isometry, it implicitly encodes the exact distribution of neural responses, which limits its ability to adapt to variations of semantic context and can cause "spillovers". We address this problem with an approach akin to Neural Doodle(Bae, Paris, and Durand 2006b) and a semantic segmentation method(Chen et al. 2018) to generate image segmentation masks for the input and reference images for a set of common labels (sky, buildings, water, etc.). We add the masks to the input image

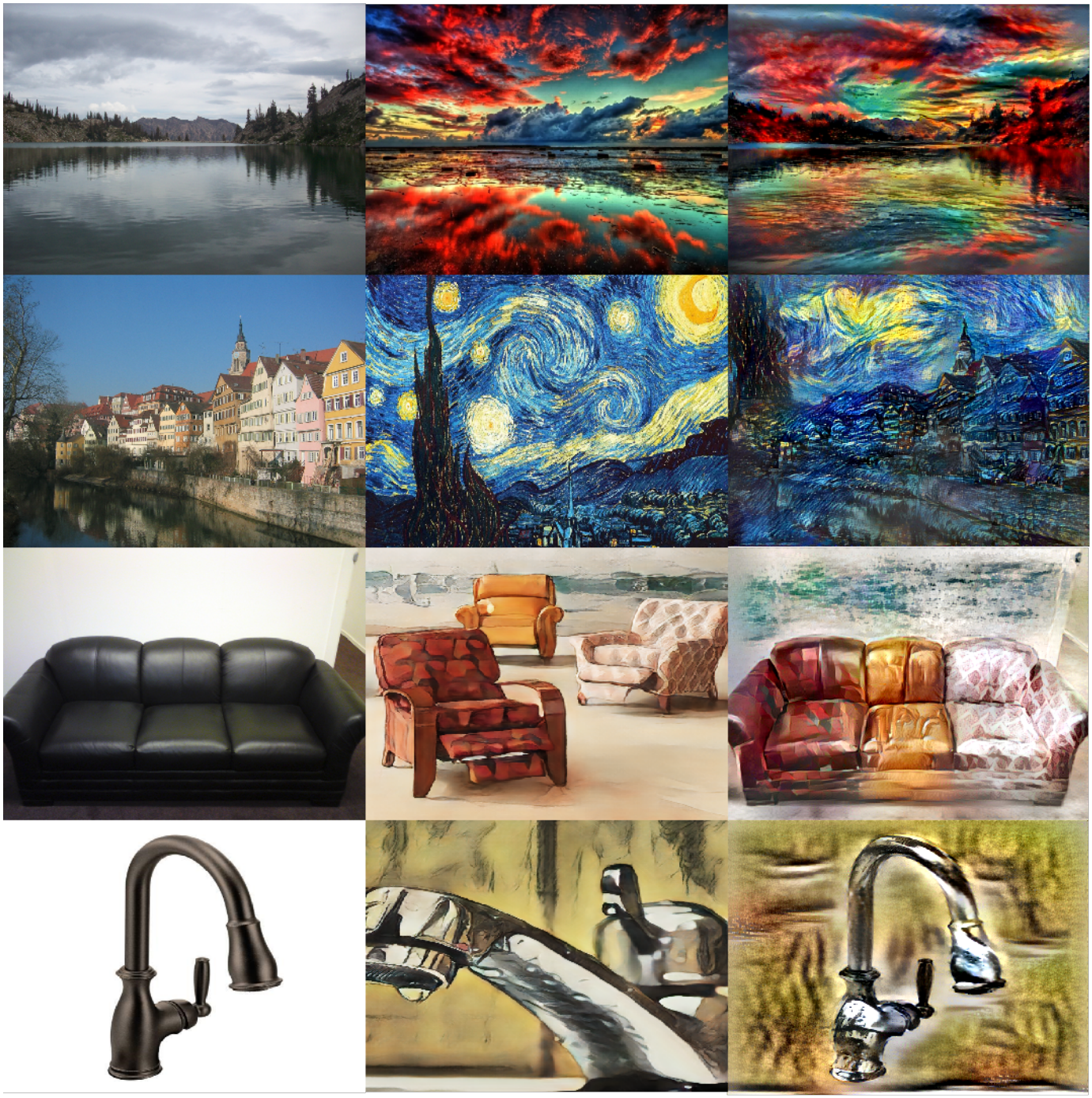

Figure 1: The left column is our input image, the middle column is our reference cartoon style image, and the rightmost column is our cartoon output result. It can be seen that our results can generate relatively accurate results according to the cartoon style benchmark under the given cartoon reference, achieving good results.

as additional channels and augment the neural style algorithm by concatenating the segmentation channels and updating the style loss as follows:

$$\mathcal{L}_{s+}^{\ell} = \sum_{c=1}^{C} \frac{1}{2N_{\ell,c}^2} \sum_{ij} (G_{\ell,c}[O] - G_{\ell,c}[S])_{ij}^2 \qquad (5)$$

$$F_{\ell,c}[O] = F_\ell[O]M_{\ell,c}[I] \quad F_{\ell,c}[S] = F_\ell[S]M_{\ell,c}[S] \quad (6)$$

where C is the number of channels in the semantic segmentation mask, $M_{l,c}[\cdot]$ denotes the channel c of the segmentation mask in layer l, and $G_{l,c}[\cdot]$ is the Gram matrix corresponding to $F_{l,c}[\cdot]$. We downsample the masks to match the feature map spatial size at each layer of the convolutional neural network.

We formulate the photorealistic style transfer objective by combining all 3 components together:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \lambda \mathcal{L}_m \quad (7)$$

where L is the total number of convolutional layers and l ndicates the l-th convolutional layer of the deep neural network. Γ is a weight that controls the style loss. α_l and β_l are the weights to configure layer preferences. λ is a weight that controls the photorealism regularization. L_c^l is the content loss (Eq. 2). L_{s+}^l is the augmented style loss (Eq. 5). L_m is the photorealism regularization (Eq. 4).

Experiments

We have conducted a series of experiments to validate the effectiveness of our approach. Prior to presenting the user study results, we first compared our method with previous works. Specifically, we compared our method with that of Yang Chen and colleagues (Chen, Lai, and Liu 2018b). This technique, at times, introduces distortions, which is undesirable in the context of style transfer scenes with outdoor backgrounds. It may lead to artifacts, such as the sky adopting the style of the ground. In contrast, our approach, incorporating realistic regularization and semantic segmentation, mitigates these issues, resulting in visually more satisfying outcomes.

Simultaneously, we conducted comparisons between our approach and global style transfer methods, such as the one proposed by Jie Chen (Chen, Liu, and Chen 2020) and colleagues. This technique employs global color mapping to match colors between input images and styles, limiting its authenticity in style transfer scenarios that require spatially varying color transformations. In contrast, our style transfer method is capable of handling context-sensitive color changes. Moreover, our algorithm directly replicates the style of the reference image, as opposed to analogy-based techniques. From a technical standpoint, our method is considered more practical.

The experimental results are shown in Figure 1. The left column is our input image, the middle column is our reference cartoon style image, and the rightmost column is our cartoon output result. It can be seen that our results can generate relatively accurate results according to the cartoon style benchmark under the given cartoon reference, achieving good results.

User Study: We conducted a user study to validate our work. Initially, we assessed realism by presenting users with a style image and four transfer outputs, including two previously mentioned global methods and our technique. Users were then asked to select the image most similar to the reference style image. Additionally, users were required to score the images on a scale ranging from absolute fidelity to absolute non-fidelity, without being shown the input images intentionally. This approach aimed to allow users to focus solely on the output images. We presented 20 comparisons, collecting an average of 35 responses for each. The study revealed that our algorithm produced the most faithful style transfer results over 80% of the time, confirming the realism achieved by our method.

Conclusions

In this paper, we have introduced a deep-learning approach that effectively transfers style from a reference image to a wide range of image content. By incorporating the Matting Laplacian and leveraging semantic segmentation, we have achieved satisfying photorealistic results in various scenarios, including changes in time of day, weather, season, and artistic edits. Our algorithm offers a practical solution for faithful style transfer, enhancing the visual appeal and realism of the stylized images. These advancements open up new possibilities for artistic expression and visual storytelling, with potential applications in image editing and beyond.

References

Bae, S.; Paris, S.; and Durand, F. 2006a. Two-scale tone management for photographic look. *ACM Transactions on Graphics (TOG)*, 25(3): 637–645.

Bae, S.; Paris, S.; and Durand, F. 2006b. Two-scale tone management for photographic look. *ACM Transactions on Graphics*, 25(3): 637–645.

Champandard, A. J. 2016. Semantic style transfer and turning two-bit doodles into fine artworks. *arXiv preprint arXiv:1603.01768*.

Chen, J.; Liu, G.; and Chen, X. 2020. AnimeGAN: a novel lightweight GAN for photo animation. In *International symposium on intelligence computation and applications*, 242–256. Springer.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; and Yuille, A. L. 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 834–848.

Chen, Y.; Lai, Y.-K.; and Liu, Y.-J. 2017. Transforming Photos to Comics Using Convolutional Neural Networks. In *2017 IEEE International Conference on Image Processing (ICIP)*, 2010–2014. IEEE Press.

Chen, Y.; Lai, Y.-K.; and Liu, Y.-J. 2018a. CartoonGAN: Generative Adversarial Networks for Photo Cartoonization. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9465–9474.

Chen, Y.; Lai, Y.-K.; and Liu, Y.-J. 2018b. Cartoongan: Generative adversarial networks for photo cartoonization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 9465–9474.

Gardner, J. R.; Upchurch, P.; Kusner, M. J.; Li, Y.; Weinberger, K. Q.; Bala, K.; and Hopcroft, J. E. 2016. Deep Manifold Traversal: Changing Labels with Convolutional Features. arXiv:1511.06421.

Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2016a. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2414–2423.

Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2016b. Image Style Transfer Using Convolutional Neural Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Hertzmann, A.; Jacobs, C. E.; Oliver, N.; Curless, B.; and Salesin, D. H. 2001. Image Analogies. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, 327–340. New York, NY, USA: Association for Computing Machinery. ISBN 158113374X.

Laffont, P.-Y.; Ren, Z.; Tao, X.; Qian, C.; and Hays, J. 2014. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on graphics (TOG)*, 33(4): 1–11.

Levin, A.; Lischinski, D.; and Weiss, Y. 2008. A Closed-Form Solution to Natural Image Matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 228–242.

Li, C.; and Wand, M. 2016. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2479–2486.

Pitie, F.; Kokaram, A. C.; and Dahyot, R. 2005. N-dimensional probability density function transfer and its application to color transfer. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, 1434–1439. IEEE.

Reinhard, E.; Adhikhmin, M.; Gooch, B.; and Shirley, P. 2001. Color transfer between images. *IEEE Computer graphics and applications*, 21(5): 34–41.

Selim, A.; Elgharib, M.; and Doyle, L. 2016. Painting style transfer for head portraits using convolutional neural networks. *ACM Transactions on Graphics (ToG)*, 35(4): 1–18.

Shih, Y.; Paris, S.; Barnes, C.; Freeman, W. T.; and Durand, F. 2014. Style transfer for headshot portraits.

Shih, Y.; Paris, S.; Durand, F.; and Freeman, W. T. 2013. Data-driven hallucination of different times of day from a single outdoor photo. *ACM Transactions on Graphics (TOG)*, 32(6): 1–11.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sunkavalli, K.; Johnson, M. K.; Matusik, W.; and Pfister, H. 2010. Multi-scale image harmonization. *ACM Transactions on Graphics (TOG)*, 29(4): 1–10.