Depth-Preserving Style Transfer

Xiuming Zhang

Massachusetts Institute of Technology

`xiuming@mit.edu`
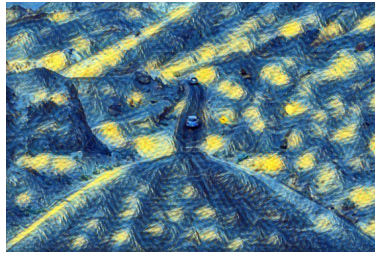
Teammates: Ruizhi Liao & Yu Xia

| (A) Scene with large variations in depth | (B) Johnson *et al.*, 2016 | (C) Our depth-preserving results |

When the input scene exhibits large variations in depth (A), the current state-of-the-art method tends to destroy the layering and lose the depth variations, producing a "flat" stylization result (B). This paper aims to address this issue by incorporating depth preservation into the loss function such that variations in depth and layering are preserved in the stylized image (C).

Abstract

Style transfer is defined as the process that given a content image and a style image it tries to migrate the style from the style image to the content image. Though it is not clear what the exact definition of style is, pattern transforming and matching are generally accepted.

In this work we present a novel method which preserve the depth information of the content image while migrating the style.

1. Introduction

convolutional neural network (CNN)

2. Related Work

The core of our method is incorporating depth preservation losses into the image transformation neural network. Therefore, we review related literature on both neural network-based image style transfer and single-image depth estimation.

2.1. Image Style Transfer with Neural Networks

Style transfer can be considered as a more general form of texture transfer, where one transfers texture from one im-age (style image) to another image (content image). Ideally, semantics of the content image should not be altered in this process. In texture transfer, it is usually the low-level features that are utilized, e.g., in [1].

With the recent prevalence of deep neural networks, researchers started exploring how high-level features extracted by neural networks can be utilized for the task of style transfer. For instance, Gatys *et al.* perform image style transfer by synthesizing a new image that matches both contents of the content image and styles of the style image [4]. In particular, they extract content representations from the content image and style representations from the style image using the VGG network [6]. Since the VGG network is trained to perform object recognition and localization tasks, the layers deep down the network hierarchy capture object information (i.e., the contents) of the content image and are insensitive to the exact pixel values. Therefore, outputs from these deep layers serve as good content targets that the synthesized image tries to achieve at varying levels of resolution. As for style, they adopt a feature space built on filter responses in any layer of the network [2]. By design, the feature space captures texture information without global arrangement. Finally, they minimize a weighted sum of the content and style loss under a CNN framework, where forward and backward passes are iteratively performed. Build-

ing upon this work, the authors recently devised a way of preserve the original colors in the content image [3]. However, the high computational cost still remains as a drawback in [4].

To reduce the computational burden and generate visually similar-quality results, Johnson *et al.* [5] train a feed-forward image transform network to approximate solutions to the optimization problem posed in [4]. In particular, their system consists of a deep residual CNN as the image transform network and the pretrained VGG network [6] as the fixed loss network. For each style image, the image transform network is trained to apply this style to a content image while minimizing the style and content losses as measured by the loss network. This method produces reasonably good results with low computational cost, but tends to lose the depth variations and destroy layering in the content image as illustrated in the teaser figure.

2.2. Depth

3. Methods

4. Experiments

5. Discussion & Conclusion

References

[1] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346. ACM, 2001.

[2] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 262–270, 2015.

[3] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman. Preserving color in neural artistic style transfer. *arXiv preprint arXiv:1606.05897*, 2016.

[4] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.

[5] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *arXiv preprint arXiv:1603.08155*, 2016.

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