Stanford
University

Exploring Style Transfer

Noah Makow & Pablo Hernandez – CS 231N

Introduction

Human's cognitive ability to separate style from content in the visual field is something that is taken for granted yet enables much of the rich, emotional understanding that comes with vision. While much recent work in computer vision has focused on understanding content, recent research has investigated the ability of computers to understand style in conjunction with, and perhaps separately from, the content of an image.

Problem Statement

We explore alterations to Johnson's fast neural style algorithm and explore the impact on efficiency and perceptual quality. We achieve different perceptual effects such as replicating multiple style images, transferring style while preserving the color of the content image, and combining style transfer with semantic segmentation. To evaluate each model, we consider space and time complexity and perceptual quality.

Datasets

We chose 50 random images from the Microsoft COCO dataset, and 5 different style images from famous artwork. The 5 style images are displayed below:

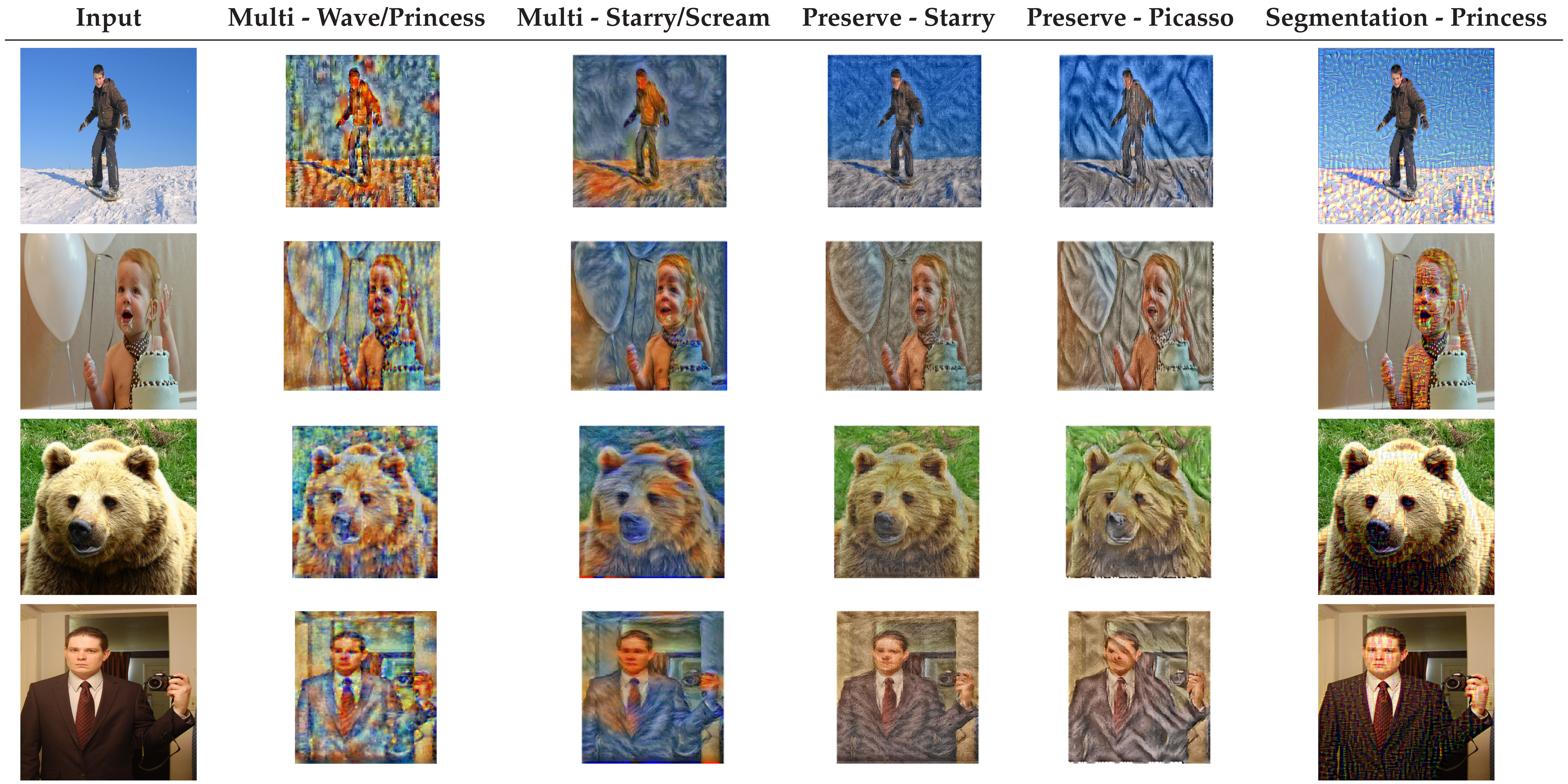
Algorithms & Models

The baseline method for fast neural style transfer involves two separate components: an *image transformation network* f_W and a *loss network* ϕ that is used to define several loss functions $l_1, \ldots, l_k$. The image transformation network is a deep residual convolutional network parametrized by weights W, transforming images x onto images $\hat{y}$ via the mapping $\hat{y} = f_W(x)$. Each loss function computes a scalar value measuring the difference between the output image $\hat{y}$ and a target image y_i. The image transformation network is trained to minimize a weighted combination of loss functions:

$$W^* = \arg\min_W E_{x,\{y_i\}}\left[\sum_{i=1} \lambda_i l_i(f_W(x), y_i)\right]$$

The loss network ϕ is used to define a *feature reconstruction loss* l^{ϕ}_{feat} and a *style reconstruction loss* l^{ϕ}_{style} that measure the differences in content and style per between images. For each input image x we have a content target y_c and a style target y_s. In the case of style transfer, y_c is the input image x and y_s is the stylized or artistic image of which we would like to extract the style. One network is trained per style target.

Results



Conclusion

Our extensions to the model have worked relatively well in terms of producing perceptually pleasing images. Combining the style from multiple images can dampen the effect of either style image. Future work will involve tuning hyperparameters to allow the output images to look more stylized, as well as exploring other extensions such as spatially guided style transfer and luminance-only style-transfer. Questions of speed and efficiency are raised when implementing this to a consumer facing application such as real-time video processing. Facebook, for example, is implementing style transfer in some of its messenger face filters, which is a interesting application of this technology.

References

[1] Matthias Bethge Leon A. Gatys, Alexander S. Ecker. A neural algorithm of artistic style.

[2] Li Fei-Fei Justin Johnson, Alexandre Alahi. Perceptual losses for real-time style transfer and super-resolution.