An Evaluation of The Impact of Dataset Bias in Pretrained VGG Network on The Performance of Neural Network Based Style Transfer

Zichen Gong[1] Hiroshi Watanabe[1]

[1]Graduate School of Fundamental Science and Engineering, Waseda University

1. Introduction

Many style transfer methods with perceptually satisfying performance followed the basic idea of the pioneering work of Gatys et al. [1]. This approach utilizes a VGG16 or VGG19 network pretrained on ImageNet to extract style and content features of the input images. It evaluates and minimizes the corresponding losses to transfer the style from one image to another. However, a recent style transfer of Sanakoyeu & Kotovenko et al. [2] achieves more impressive outputs while discarding the pretrained VGG network (an adversarial discriminator is used instead). They argue that the data bias introduced by ImageNet database may influence the ultimate performance of style transfer. In this paper, we conduct an experiment to evaluate and verify the impact of ImageNet bias on the quality of stylized output images.

2. Experiments

2.1 Style transfer model

We employ Stylebank [3] as the experimental style transfer model since it is developed and constructed along with a pretrained VGG network. It can conveniently train multiple styles in one time within relatively short time compared with other feed-forward generator networks for style transfer such as [4, 5], while showing competitive and even better performance in stylization quality.

2.2 Evaluation metric

There was no quantitative evaluation method of the stylized outputs until Sanakoyeu & Kotovenko et al. [2] proposed a metric named 'Style transfer deception rate'. Here, a VGG16 network trained from scratch to classify 624 artists on Wikiart was employed to calculate the fraction of those correctly classified as the artworks of a particular style (belongs to corresponding artist) among all stylized images. The higher the deception rate is, the better performance of the style transfer model achieves.

2.3 Methodology

We respectively train the Stylebank on MS-COCO training set which is composed of scenes pictures taken in real life. Thus, they have similar dataset bias with ImageNet database. We used another training set composed of illustrations, cartoon scenes images and various modern artworks collected on the Internet, representing a dataset with bias quite different from ImageNet. Each training process includes 5 target styles of different artworks (320,000 iterations for each dataset), and 5,000 test images in accordance to corresponding training set. Finally, we evaluate their outputs using the metric mentioned above. The results are shown in Table 1. Note that the target styles contain three different artworks of Claude Monet.

Table 1. Style transfer deception rate of Stylebank trained on different dataset. (See Appendix for the detailed results.)

Style / Model	Claude Monet	Edvard Munch	Vincent Van Gogh	Mean Deception rate
Trained on MS-COCO	0.10087	0.02700	0.31220	0.12836
Trained on illustrations	0.03980	0.03420	0.06820	0.04436

2.4 Analysis and discussion

As shown in Table 1, the deception rate decreased dramatically by 65.44% when the Stylebank was trained on illustrations dataset which potentially introduces bias that different from ImageNet instead of training on the dataset with similar bias. The result indicates that it is possibly difficult for a VGG16/19 network pretrained on ImageNet to accurately and properly extract features when applying to a dataset with bias different from ImageNet such as

dataset consists of illustrations, cartoon images or artworks. Consequently, it could lead to the difficulty of optimizing the losses in detail thus degrade the stylization. Furthermore, we conjecture that using a VGG16/19 network pretrained on ImageNet as the feature extractor in style transfer model can even limit the performance of the whole model since it might have similar difficulty in extracting detailed features of the target style images.

Someone may argue that the decreased deception rate might be induced by the classifier itself. It is worthy to note that the aim of the classifier used for metric is to classify the styles instead of contents. The Wikiart dataset used for training the classifier contains varieties of artworks across the art history so that images with a vast range of different contents and various colors and strokes are included. Therefore, we consider that the classifier used for metric on the whole is able to focus on the style features. Thus, it can generally give a correct evaluation result.

3. Conclusion

We conduct an experiment to verify and confirm that the dataset bias introduced by ImageNet could have a negative effect on the stylization outputs. The result suggests that one may train the VGG network from scratch with his own dataset for a proper feature extractor in the style transfer task based on neural networks, especially when his dataset is quite different from ImageNet.

4. Reference

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," arXiv:1508.06576v2, Sep. 2015.

[2] A. Sanakoyeu, D. Kotovenko, S. Lang, and B. Ommer, "A Style-Aware Content Loss for Real-time HD Style Transfer," European Conference on Computer Vision (ECCV2018), No.O-2C-03, Sep. 2018.

[3] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, " Stylebank: An Explicit Representation for Neural Image Style Transfer." Proc. CVPR. Vol. 1, No. 3, July 2017.

[4] J. Johnson, A. Alahi, and F. Li, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution." European Conference on Computer Vision 2016 (ECCV'16), No. P-1B-47, Oct. 2016.

[5] V. Dumoulin, J. Shlens, and M. Kudlur, " A Learned Representation for Artistic Style," arXiv:1610.07629v5, Feb. 2017.

5. Appendix

Detailed results of Table 1.
Style transfer deception rate of Stylebank trained on different dataset.

Trained on MS-COCO

Style	Number of correctly classified	Number of total test images	Calculated Deception rate
Claude Monet	1513	15000	0.10087
Edvard Munch	135	5000	0.02700
Vincent Van Gogh	1561	5000	0.31220
Mean Deception rate = 3209/25000 = 0.12836			

Trained on illustrations

Style	Number of correctly classified	Number of total test images	Calculated Deception rate
Claude Monet	597	15000	0.03980
Edvard Munch	171	5000	0.03420
Vincent Van Gogh	341	5000	0.06820
Mean Deception rate = 1109/25000 = 0.04436			