Image Style Transfer
AdaIN

ICCV

2017

Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization

Xun Huang Serge Belongie
Department of Computer Science & Cornell Tech, Cornell University
{xh258,sjb344}@cornell.edu

Abstract

Gatys et al. recently introduced a neural algorithm that renders a content image in the style of another image, achieving so-called style transfer. However, their framework requires a slow iterative optimization process, which limits its practical application. Fast approximations with feed-forward neural networks have been proposed to speed up neural style transfer. Unfortunately, the speed improvement comes at a cost: the network is usually tied to a fixed set of styles and cannot adapt to arbitrary new styles. In this paper, we present a simple yet effective approach that for the first time enables arbitrary style transfer in real-time. At the heart of our method is a novel adaptive instance normalization (AdaIN) layer that aligns the mean and variance of the content features with those of the style features. Our method achieves speed comparable to the fastest existing approach, without the restriction to a pre-defined set of styles. In addition, our approach allows flexible user controls such as content-style trade-off, style interpolation, color & spatial controls, all using a single feed-forward neural network.

1. Introduction

The seminal work of Gatys *et al.* [16] showed that deep neural networks (DNNs) encode not only the content but also the *style* information of an image. Moreover, the image style and content are somewhat separable: it is possible to change the style of an image while preserving its content. The style transfer method of [16] is flexible enough to combine content and style of arbitrary images. However, it relies on an optimization process that is prohibitively slow.

Significant effort has been devoted to accelerating neural style transfer. [24, 51, 31] attempted to train feed-forward dilemma. Our approach can transfer arbitrary new styles in real-time, combining the flexibility of the optimization-based framework [16] and the speed similar to the fastest feed-forward approaches [24, 52]. Our method is inspired by the *instance normalization* (IN) [52, 11] layer, which is surprisingly effective in feed-forward style transfer. To explain the success of instance normalization, we propose a new interpretation that instance normalization performs style normalization by normalizing feature statistics, which have been found to carry the style information of an image [16, 30, 33]. Motivated by our interpretation, we introduce a simple extension to IN, namely *adaptive instance normalization* (AdaIN). Given a content input and a style input, AdaIN simply adjusts the mean and variance of the content input to match those of the style input. Through experiments, we find AdaIN effectively combines the content of the former and the style latter by transferring feature statistics. A decoder network is then learned to generate the final stylized image by inverting the AdaIN output back to the image space. Our method is nearly three orders of magnitude faster than [16], without sacrificing the flexibility of transferring inputs to arbitrary new styles. Furthermore, our approach provides abundant user controls at runtime, without any modification to the training process.

2. Related Work

Style transfer. The problem of style transfer has its origin from non-photo-realistic rendering [28], and is closely related to texture synthesis and transfer [13, 12, 14]. Some early approaches include histogram matching on linear filter responses [19] and non-parametric sampling [12, 15]. These methods typically rely on low-level statistics and often fail to capture semantic structures. Gatys *et al.* [16] for

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 1501-1510).

Image Style Transfer

AdaIN

Background & Goal

Previous research limitation

- Traditional framework(NST) requires a slow iterative optimization process, which limits its practical application
- Feed-forward network relies on a fixed set of styles and cannot adapt to arbitrary new styles.

Goal

- Apply styles to content images in real time
- Flexibility to adapt to new styles immediately

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 1501-1510).

Image Style Transfer

AdaIN

Feature Normalization

Mean of feature map

- Hue, Brightness

Variance of feature map

- Contrast, Texture

Feature normalization

- The mean and variance of each channel are computed to remove the original style, and they are replaced with the me an and variance of the target domain.

$$\gamma\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\beta$$

Eq 1. Feature Normalization

Image Style Transfer

AdaIN

Feature Normalization

Batch normalization

- Normalize by calculating the mean and variance for each channel over the spatial dimensions (H, W) within a batch

Instance normalization

- Unlike BN, normalization is performed for each instance (image)

$$\mathrm{BN}(x) = \gamma\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\beta$$

$$\mu_c(x) = \frac{1}{NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{nchw}$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}(x_{nchw}-\mu_c(x))^2+\epsilon}$$

Eq 2. Batch Normalization

$$\mathrm{IN}(x) = \gamma\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\beta$$

$$\mu_{nc}(x) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{nchw}$$

$$\sigma_{nc}(x) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(x_{nchw}-\mu_{nc}(x))^2+\epsilon}$$

Eq 3. Instance Normalization

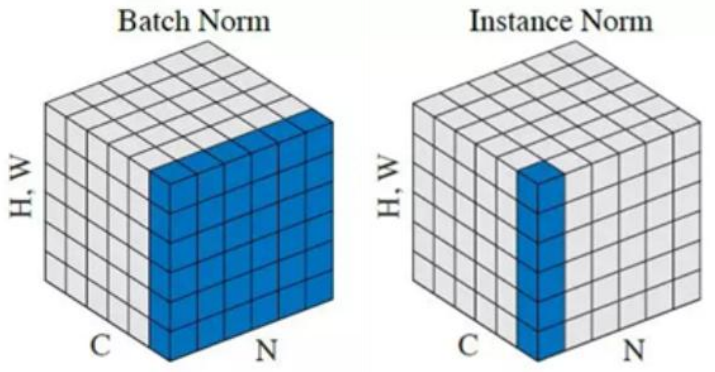


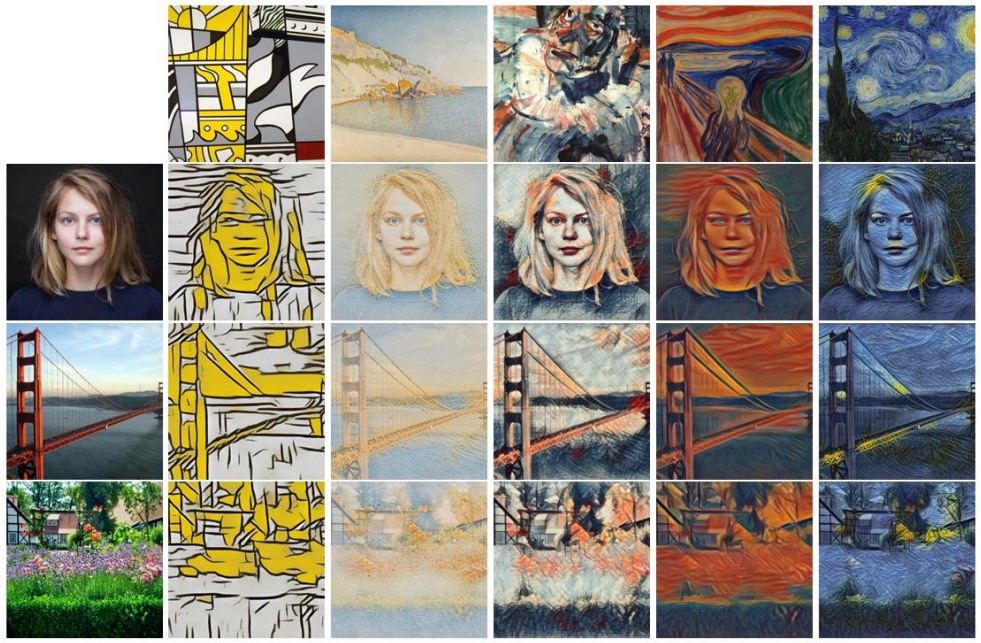
Image Style Transfer

AdaIN

Feature Normalization

Conditional Instance Normalization

- Learn affine parameters differently for each style



$$\mathrm{CIN}(x; s) = \gamma^s \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta^s$$

Eq 3. Conditional Instance Normalization

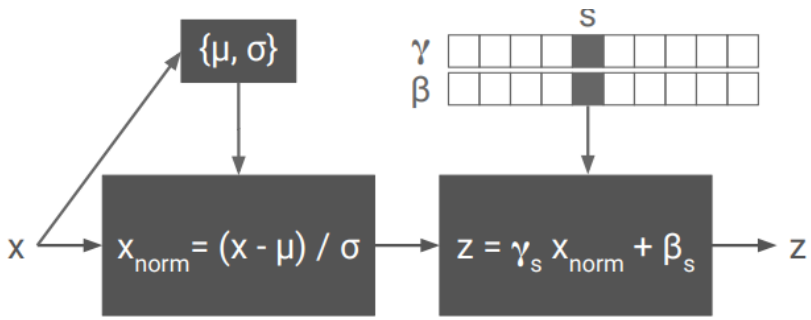


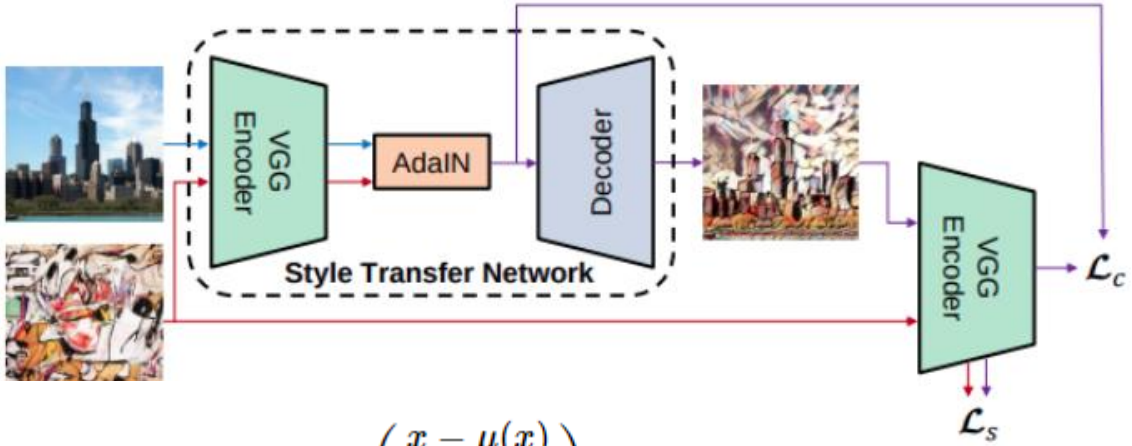
Image Style Transfer

AdaIN

Network architecture

Adaptive Instance Normalization

- AdaIN receives a content input x, style input y → aligns the mean and variance of x to match those of y
- No learnable affine parameters → adaptively computes the affine parameters from the style input
- Perform style transfer in the feature space by transferring feature statistics (channel-wise mean & variance)



$$\text{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 1501-1510).

Image Style Transfer

AdaIN

Loss functions

Content loss

- Content loss: Euclidean distance between the target features and the features of the output image

Style loss

- AdaIN layer only transfers the mean and standard deviation of the style features
- Style loss only matches these statistics

$$\mathcal{L}_c = \|f(g(t)) - t\|_2$$

Eq 7. Content loss

$$\mathcal{L}_s = \sum_{i=1}^{L} \|\mu(\phi_i(g(t))) - \mu(\phi_i(s))\|_2 \quad + \quad \sum_{i=1}^{L} \|\sigma(\phi_i(g(t))) - \sigma(\phi_i(s))\|_2$$

Eq 8. Style loss

$$\mathcal{L} = \mathcal{L}_c + \lambda \mathcal{L}_s$$

Eq 9. Total loss

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 1501-1510).

Image Style Transfer

AdaIN

Experiment Results

Method	Time (256px)	Time (512px)	# Styles
Gatys *et al.*	14.17 (14.19)	46.75 (46.79)	∞
Chen and Schmidt	0.171 (0.407)	3.214 (4.144)	∞
Ulyanov *et al.*	**0.011** (N/A)	**0.038** (N/A)	1
Dumoulin *et al.*	**0.011** (N/A)	**0.038** (N/A)	32
Ours	**0.018** (0.027)	**0.065** (0.098)	∞

Table 1. Speed comparison

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 1501-1510).

Image Style Transfer

AdaIN

Implications & Limitations

Implications

- present a simple adaptive instance normalization (AdaIN) layer that for the first time enables arbitrary style transfer in real-time

Limitations

- AdaIN layer only aligns the most basic feature statistics (mean and variance)

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision* (pp. 1501-1510).