光电技术及应用

DOI：10.16818/j.issn1001－5868.2023071803

结合可逆网络的超高分辨率图像的风格迁移

林　真，郑茜颖*

（福州大学 物理与信息工程学院，福州 350108）

摘　要： 针对处理超高分辨率图像时面临的内存成本和风格迁移过程中过度风格化，提出了一种结合可逆网络的超高分辨率图像的风格迁移方法。该算法采用可逆的 Glow 模块作为基本单元构建可逆神经网络模块，并将图像分为小块处理；在风格迁移模块中使用具有通道注意力机制的残差模块和缩略图实例化归一化模块（TIN），以保证各模块风格一致；提出基于全局-局部的损失计算方式，能够有效地处理局部的结构特征。实验结果表明，相较于当前通用的各种神经风格迁移网络，所提算法不仅能够避免图像在编码和解码过程中的信息丢失问题，而且能以更低的内存成本实现更优的风格迁移。

关键词： 可逆网络；超高分辨率图像；风格迁移；通道注意力机制；Glow 模块

中图分类号： TP389.1　**文章编号：** 1001－5868(2023)05－0756－05

Super-Resolution Image Style Transfer Combined with Reversible Network

LIN Zhen，ZHENG Qianying

（College of Physics & Information Engineering，Fuzhou University，Fuzhou 350108，CHN）

Abstract： Aiming at the problem of memory consumption in ultra-high resolution image processing and the problem of over-stylization in the process of style transfer，a method of ultra-high resolution image style transfer combined with reversible network is proposed. The algorithm used the reversible Glow module as the basic unit to construct a reversible neural network module，and divided the image into small blocks for processing. In the style transfer module，a residual module with a channel attention mechanism and a thumbnail instantiation normalization module（TIN）were used to ensure that the styles of each module were consistent. A global-local loss calculation method was proposed，which could effectively deal with local structural features. Experimental results show that，compared with the current general-purpose neural style transfer network，this algorithm can not only avoid the information loss problem in the process of image encoding and decoding，but also achieve better style transfer performance at a lower memory cost.

Key words： reversible network；ultra-high resolution image；style transfer；channel attention mechanism；Glow module

0　引言

在神经风格迁移之前，风格迁移大都是基于非真实感渲染的风格化算法来实现特定的艺术风格转化。由于这些算法使用低级图像特征，所以无法准确表达图像的结构特征[1]。基于卷积神经网络的风格迁移算法，将图像的内容信息和样式信息进行编码，实现任意图像的内容和风格的组合。通用风格

收稿日期：2023－07－18.

基金项目：福建省科技重点产业引导项目(2020H0007).

* **通信作者：**郑茜颖　**E-mail：**zhengqy@vip.sina.com

迁移算法有 WCT[2]，AdaIN[3] 和 Avatar-Net[4]，使用预训练网络提取特征，实现优秀的风格迁移效果，但是受到内存的限制，预训练网络的大模型尺寸极大地限制了输入图像的分辨率。

为了实现超分辨率图像的风格迁移，提出基于模型压缩的风格迁移算法。Wang 等[5] 提出了一种新的知识蒸馏方法，从预训练的冗余模型中学习更小的模型，压缩模型的参数。Jing 等[6] 提出了基于 MobileNet 的轻量级网络，显著地降低计算复杂度。An 等[7] 提出了 ArtNet，它是基于 GoogLeNet，通过零通道剪枝得到的轻量级网络，通过删除深层特征的冗余通道来降低计算成本。但是这些算法在面对超高分辨率图片时依旧存在内存耗尽的情况。基于通过处理子样本然后合并成整体的思想，能够解决内存限制的问题。例如 Junginger 等[8] 使用 GAN 在子样本的级别上训练和评估域转换，但是算法在图像的局部处理中会出现变形的问题。Chen 等[9] 提出了 URST 和缩略图实例规范化（Thumbnail Instance Normalization，TIN），实现对任意分辨率图像的风格迁移，但是算法对保存内容图像的纹理特征的效果一般。

基于 URST 的框架，本文提出了基于可逆网络的超高分辨率真实图像的风格迁移方法，旨在使用较少的内存，完成超高分辨率的内容图像对风格图像的色彩特征的学习，得到的风格化图像能够保持内容图像的结构细节。为了增强对内容图像的关键特征的提取，本文提出了一个结合通道注意力机制的风格转换模块。为了缓解图像局部出现扭曲、变形的问题，本文提出了一种基于全局-局部的损失计算方式。经过大量实验表明，该方法实现了与当前通用的风格迁移算法相当的性能。

1　基于可逆网络的超高分辨率图像的风格迁移方法

本方法的总体框架包含四个部分：1）图像的切割和重组模块，完成将超高分辨率的图像切割成多个小块，以及将风格化后的小块重组成超高分辨率图像；2）可逆网络模块，实现图像的深度特征提取和将深度特征转换为图像的功能；3）风格转换模块，使用结合通道注意力机制的缩略图实例规范化层；4）设计全局-局部损失，不仅着重于全局的风格化效果，而且能够增强局部区域的感知效果。算法的整体流程如图 1 所示。

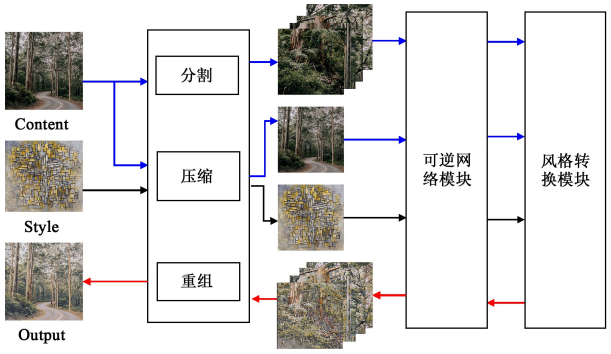

图 1　总体框架

1.1　图像的切割和重组模块

为了能在有限的 GPU 内存上处理超高分辨率的图像，本方法将一个超高分辨率的内容图像，依据图像的大小，计算出将内容图像划分为多个小块的滑动窗口大小，以及滑动窗口的使用步幅。在对图像小块进行风格化之前，本方法对图像小块的四周进行了等量填充操作。对完成风格化后的重叠块进行去除填充和重叠区域，并将它们重组成一个完整的风格化图像。

图 2 的 A 图展示了不采用对图像小块四周进行等量填充的风格化结果，在小块之间的拼接处可以看到明显的拼接痕迹。B 图展示了采用了对图像小块四周进行等量填充的风格化结果，小块拼接处无明显痕迹。由此可见，采用等量填充的方法，可以避免不同图像小块边界区域的不连贯的视觉感知。

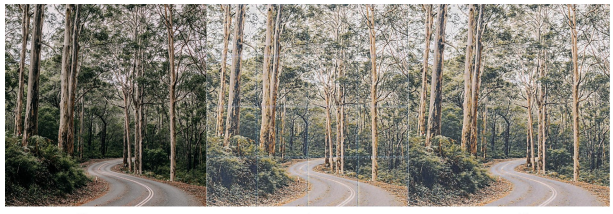

图 2　采用不同处理方式的风格化比较

1.2　可逆网络

本文方法的可逆网络是基于 Glow[10] 模型来设计的，Glow 模型的每一个模块都包含三个可学习的可逆变换链，即激活归一化层（Actnorm），可逆的 1×1 卷积层，仿射耦合层。激活归一化层使用每个通道的比例和偏置参数对激活执行仿射变换，是批量归一化的替代方法。本方法在可逆网络的仿射耦合层的设计上选用仿射耦合层的特例加性耦合层。本文设计的网络由两个大模块组成，每个大模块中都包含 K 个小模块，每个小模块都是按照激活归一

化层，可逆的 1×1 卷积层和加性耦合层的顺序组成。每个大模块间都由挤压操作层(Squeeze)来连接，挤压操作层可以实现将特征图拆分成小的补丁块，然后沿着通道维度连接补丁块。本文的可逆网络模块的具体框架图如图 3 所示。

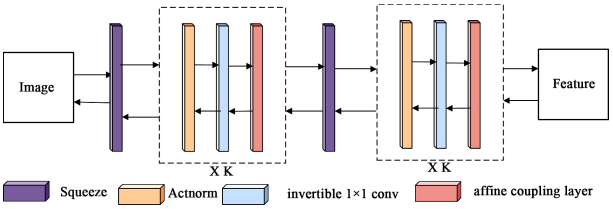

图 3 可逆网络模块

1.3 风格迁移模块

风格迁移模块的结构如图 4 所示。受 SENet[11] 的启发，利用一个全局平均池化层和两个线性层搭建的残差模块来实现通道注意力，之后结合 Chen 等提出的缩略图实例规范化来完成风格迁移，风格迁移过程可表示为：

$$f(x',t',s)=\sigma(s)\left(\frac{x'-\mu(t')}{\sigma(t')}\right)+\mu(s) \quad (1)$$

式中，x' 和 t' 表示分割的图像小块和缩略图经过可逆网络和通道注意力模块后的输出，s 表示风格图像经过可逆网络后的输出，$\mu(t')$ 和 $\sigma(t')$ 是 t' 的通道均值和标准差，$\mu(s)$ 和 $\sigma(s)$ 是 s 的通道均值和标准差。这种计算方式能够保证不同的分割图像小块的风格化图片之间的样式一致性。

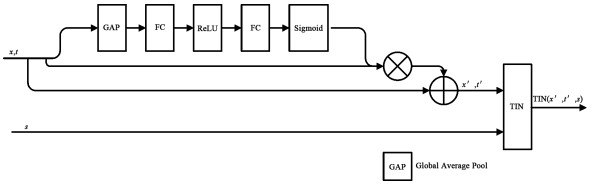

图 4 风格迁移模块

1.4 全局-局部损失

在传统的神经风格迁移算法中，经常出现一些整体上具有不错的视觉感知的风格化图像，在局部区域出现了扭曲、变形的问题。受到 Jiang 等[12] 的启发，在保证全局图像的风格化转换效果的基础上，为了保证局部区域也能得到不错的风格化转换效果，本方法提出了一种基于全局-局部的损失计算的训练方式。除了传统的计算全局图像的内容损失和风格损失之外，本方法还添加了一个局部图像的内容损失，其中局部图像是从输入的内容图像中随机裁剪得到局部块。

本方法使用预训练的 VGG-19 来计算全局和局部的损失函数来训练本方法的可逆网络。对于全局损失的计算：

$$L^{\text{Global}}=\alpha_1 L_{\text{C}}+\alpha_2 L_{\text{S}} \quad (2)$$

式中，L_{C} 是内容损失，L_{S} 是风格损失，α_1 和 α_2 分别是内容损失权重和风格损失权重。内容损失是内容图像的特征与输出图像特征之间的欧式距离。

$$L_{\text{C}}=\left\|f(g(h(d(I_{\text{C}}))))-f(I_{\text{C}})\right\|_2 \quad (3)$$

式中，I_{C} 表示调整好尺寸的内容图像，$d(\cdot)$，$g(\cdot)$，$f(\cdot)$，$h(\cdot)$ 分别代表着可逆网络的前向，可逆网络的反向，预训练的 VGG-19 网络，以及结合通道注意力的风格转换模块。

$$L_{\text{S}}=\sum_{i=1}^{L}\left\|f_i(g(h(d(I_{\text{C}}))))-f_i(I_{\text{S}})\right\|_2 \quad (4)$$

式中，I_{S} 表示调整好尺寸的风格图像，f_i 表示 VGG-19 网络中用于计算风格损失的层。在本方法的训练中，用的是 relu1_1，relu2_1，relu3_1，relu4_1 层。

对于局部损失，本方法通过随机裁剪的方式将输入的内容图像裁剪出 N 个固定大小的局部图片 I_{tj}。局部损失的计算如下：

$$L^{\text{local}}=\frac{1}{N}\sum_{j=1}^{N}(\beta_1\left\|f(g(h(d(I_{tj}))))-f(I_{tj})\right\|_2+\beta_2\sum_{i=1}^{L}\left\|f_i(g(h(d(I_{tj}))))-f_i(I_{\text{st}})\right\|_2) \quad (5)$$

式中，β_1 和 β_2 是局部剪切图像的内容损失权重和风格损失权重，I_{st} 是尺寸跟局部剪切图像一样的风格图像。总的损失计算为：

$$L=L^{\text{Global}}+L^{\text{Local}} \quad (6)$$

2 实验与分析

2.1 实验设置

从 pexels.com 和 wallpaper abyss 上收集的摄影作品和艺术图像来做内容图像和风格图像。在训练时，本方法将输入图像调整为 256×256，并且从每张图像上随机裁剪 5 个 128×128 的局部图像用来计算局部损失。图 3 所示的可逆网络模块中的 K 设置为 8，表示每个大模块中包含 8 个可逆的小模块。本方法采用 PyTorch 深度学习框架建立网络，所有实验均在单张 NVIDIA GeForce RTX 3080Ti 显卡上完成训练。可逆网络模块训练的优化器选用 Adam 优化器，批量大小为 1，最大迭代次

数为 120 000 次，初始的学习率设置为 1×10^{-4}，学习率衰减为 5×10^{-5}。全局图像和局部图像的损失计算中，内容损失权重和风格损失权重都是 0.1 和 1。

2.2　风格化结果

为了验证本方法的风格转换能力，本文将基于本方法得到的风格转移结果与当前通用的风格转移算法做对比，即 AdaIN，WCT，LinearWCT[13]，Artflow[14]和 URST。在测试阶段，选用 2 400×2 400 像素的图像来进行风格迁移，本文方法把内容图像分割成 512×512 大小的小块，并且在每个小块的四周打上了 96 个像素点大小的补丁。图 5 显示了所有比较算法的风格迁移结果，所有比较的算法在视觉效果上都能得到很不错的效果，只不过在艺术效果的侧重点上会有所不同。本文算法想实现的效果是在保留内容图像的所有细节的基础上学习风格图像的色彩分布。从结果上可以看出，WCT 算法可以学习到更加丰富的艺术效果，而 AdaIN，LinearWCT，Artflow，URST 以及本文算法可以保留更多的内容图像的细节，尤其是本文算法。

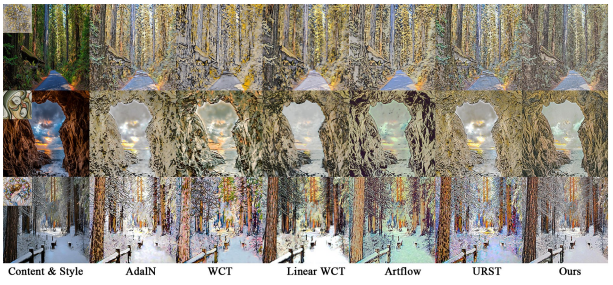

图 5　通用算法的风格化结果

受到 An 等[14]的启发，本文除了在视觉效果上对风格迁移结果进行比较，还进行了定量比较。本文总共进行了三个方面的比较，第一个方面是使用结构相似性指数法（Structural Similarity Index Measure，SSIM），比较输入的内容图像和风格化后的图像之间的相似度，从而衡量风格化过程中内容损失的程度。第二个方面是使用输入的风格图像和风格化后的图像之间的 Gram 矩阵的均方误差，来衡量算法在风格化过程中学习艺术效果的能力。第三个方面是比较所有算法在风格化过程中的显存使用量。本文选择了 10 种 2 400×2 400 的内容图像和 10 种风格图像，总计得到 100 张风格化图像来做定量测试。

结果如表 1 所示，加粗的数字代表最优结果。可见本文算法在 SSIM 和显存使用量上都达到了最优的效果。最优的 SSIM 表明了本文算法在保留输入内容图像的细节方面的能力是最好的。在 Gram 损失计算上，由于本文算法偏向于保持内容图像的真实性，所以在这个方面达不到最好的效果是合理的，而且在这个方面上我们的算法能达到跟 ArtFlow，LinearWCT 一样的效果。在显存使用量上，本文算法远低于大部分算法。本文采用 URST 算法中将图像分割后再风格化的策略，降低了显存的使用量，在模型架构上本文应用可逆网络作为编解码模块，相较 VGG-19 进一步降低了显存占用。

综上可得，本文算法使用较低的显存即可完成超高分辨率图像的风格化转换，同时具有与其他通用的风格迁移算法相当的风格迁移性能。

表 1　定量评价的结果

Method	SSIM	Gram Loss	Memory Usage/G
AdaIN	0.304 6	0.000 2	7.5
WCT	0.288 3	$6.161\,6\times10^{-6}$	10.5
LinearWCT	0.366 5	0.000 1	11.8
ArtFlow	0.306 1	0.000 1	11.8
URST	0.263 7	0.000 2	3.0
Ours	0.440 9	0.000 1	2.8

为了证明本文算法在超高分辨率的图像上的风格迁移能力，本文将和提出 TIN 和 URST 框架的风格迁移算法作对比。图 6 显示了两个算法在超高分辨率图像上的风格迁移结果，可以看到两个算法都能很好地完成内容图像的色彩风格迁移。表 2 展示了对比实验结果的定量比较，可见本文算法的 SSIM 是优于 URST 的 SSIM。这表明了对超高分辨率图像的风格迁移，本文算法也能够保留更多的内容图像的纹理细节，这也证实了使用可逆网络模型能够避免在风格迁移过程中的输入图像由于池化操作带来的空间信息损失的问题。

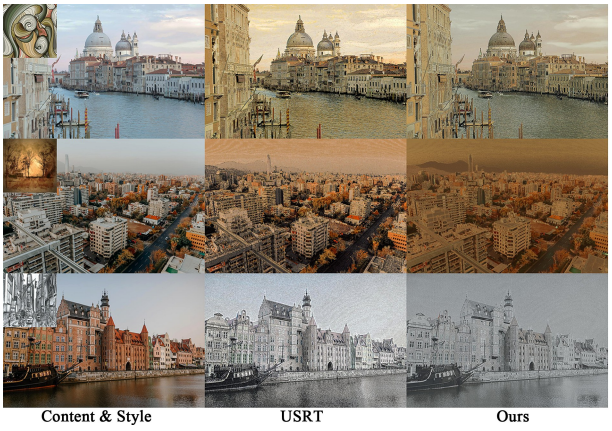

图 6　超高分辨率程式化结果（12 000×8 000 像素）

表 2 在超分辨率图像上的定量评价

Method	SSIM	Gram Loss
URST	0.415 5	0.000 4
Ours	0.457 5	0.000 2

3 结论

本文提出了一种结合可逆网络模块的风格转换算法,能够实现以有限的显存完成对超高分辨率的风格迁移过程。与可逆网络模块的结合,使得输入图像能够避免在编码过程中因为池化操作而带来的空间信息损失的问题。结合通道注意力机制的风格迁移模块,能高效地提取内容图像的关键特征,从而将风格图像的风格更恰当地迁移到内容图像的特征图上。为了避免图像局部纹理出现扭曲变形的问题,引入了基于全局和局部的损失计算。我们的算法能够避免在风格迁移过程中,过度学习风格图像的艺术效果的问题,很好的保留了内容图像的纹理细节。

参考文献:

[1] Jing Y, Yang Y, Feng Z, et al. Neural style transfer: A review[J]. IEEE Trans. on Visualization and Computer Graphics, 2019, 26(11): 3365-3385.

[2] Li Y J, Fang C, Yang J M, et al. Universal style transfer via feature transforms[J/OL]. ArXiv, 2017, 30: 1-11[2023-8-18]. https://arxiv.org/pdf/1705.08086.pdf.

[3] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C]// Proc. of the IEEE Inter. Conf. on Computer Vision, 2017: 1501-1510.

[4] Sheng L, Lin Z, Shao J, et al. Avatar-net: Multi-scale zero-shot style transfer by feature decoration[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 8242-8250.

[5] Wang H, Li Y, Wang Y, et al. Collaborative distillation for ultra-resolution universal style transfer[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2020: 1860-1869.

[6] Jing Y, Liu X, Ding Y, et al. Dynamic instance normalization for arbitrary style transfer[C]// Proc. of the AAAI Conf. on Artificial Intelligence, 2020: 4369-4376.

[7] An J, Li T, Lou J B, et al. Real-time universal style transfer on high-resolution images via zero-channel pruning[J]. ArXiv, 2020, 2006: 09029[2023-8-18]. https://arxiv.org/pdf/2006.09029.pdf.

[8] Junginger A, Hanselmann M, Strauss T, et al. Unpaired high-resolution and scalable style transfer using generative adversarial networks[J]. ArXiv, 2018, 1810: 05724[2023-8-18]. https://arxiv.org/pdf/1810.05724.pdf.

[9] Chen Z, Wang W H, Xie E Z, et al. Towards ultra-resolution neural style transfer via thumbnail instance normalization[C]// Proc. of the AAAI Conf. on Artificial Intelligence, 2022: 393-400.

[10] Kingma D P, Dhariwal P. Glow: Generative flow with invertible 1 × 1 convolutions[J]. Advances in Neural Information Processing Systems, 2018, 31: 10236-10245.

[11] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[12] Jiang Y, Gong X, Liu D, et al. Enlightengan: Deep light enhancement without paired supervision[J]. IEEE Trans. on Image Processing, 2021, 30: 2340-2349.

[13] Li X, Liu S, Kautz J, et al. Learning linear transformations for fast image and video style transfer[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019: 3809-3817.

[14] An J, Huang S, Song Y, et al. Artflow: Unbiased image style transfer via reversible neural flows[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2021: 862-871.

作者简介:

林 真(1999—),男,硕士研究生,主要研究方向为深度学习、图像处理;

郑茜颖(1970—),女,博士,副教授,主要研究方向为计算机视觉、图像处理。