Towards Ultra-Resolution Neural Style Transfer via
朝向超分辨率神经样式转移
Thumbnail Instance Normalization
缩略图实例标准化

Zhe Chen[1], Wenhai Wang[1], Enze Xie[2], Tong Lu[1], Ping Luo[2]
浙晨 1 号，文海王 1 号，恩泽协 2 号，通路 1 号，平路 2 号
[1]Nanjing University [2]The University of Hong Kong
1 南京大学 2 香港大学

chenzhe98@smail:nju:edu:cn, lutong@nju:edu:cn
Chenzhe98@smail: nju: edu: cn，lutong@nju: edu: cn

Abstract
摘要

We present an extremely simple Ultra-Resolution Style Transfer framework, termed URST, to flexibly process arbitrary high-resolution images (e.g., 10000 10000 pixels) style transfer for the first time. Most of the existing state-of-the-art methods would fall short due to massive memory cost and small stroke size when processing ultra-high reso-lution images. URST completely avoids the memory prob-lem caused by ultra-high resolution images by 1) dividing the image into small patches and 2) performing patch-wise style transfer with a novel Thumbnail Instance Normaliza-tion (TIN). Specifically, TIN can extract thumbnail's nor-malization statistics and apply them to small patches, en-suring the style consistency among different patches.

我们提出了一个非常简单的超分辨率样式转换框架，称为 URST，以灵活处理任意高分辨率图像(如 1000010000 像素)样式转换的第一次。在处理超高分辨率图像时，由于大量的内存消耗和小的笔画尺寸，大多数现有的最先进的方法都不能满足要求。URST 完全避免了超高分辨率图像引起的内存问题: 1)将图像分割成小块，2)使用新颖的缩略图实例规范化(TIN)执行逐块样式转换。具体来说，TIN 可以提取缩略图的标准化统计数据，并将其应用到小补丁上，确保不同补丁之间的样式一致性。

Overall, the URST framework has three merits compared to prior arts. 1) We divide input image into small patches and adopt TIN, successfully transferring image style with arbitrary high-resolution. 2) Experiments show that our URST surpasses existing SOTA methods on ultra-high reso-lution images benefiting from the effectiveness of the pro-posed stroke perceptual loss in enlarging the stroke size.

总的来说，URST 框架与现有技术相比具有三个优点: 1)将输入图像分割成小块，采用 TIN 技术，成功地实现了任意高分辨率图像样式的转换。2)实验结果表明，我们的 URST 方法在超高分辨率图像上优于现有的 SOTA 方法，其优点在于提出的笔画知觉损失在扩大笔画尺寸方面的有效性。

3) Our URST can be easily plugged into most existing style transfer methods and directly improve their perfor-mance even without training. Code is available at https: //github.com/czczup/URST.
我们的 URST 可以轻松插入大多数现有的风格转换方法，并直接提高其性能，即使没有培训。代码可在 https://github. com/czczczup/urst 获得。

1. Introduction
1. 简介

With the development of deep learning, neural style transfer has achieved remarkable success [3, 10, 14, 15, 17, 20, 24, 28, 30], but ultra-high resolution style transfer is rarely explored in these works. In natural scenes, ultra-high resolution images are often seen in large posters, pho-tography works, and ultra-high definition (e.g., 8K) videos. There are two main challenges when transferring the style of ultra-high resolution images: 1) The massive memory cost of ultra-high resolution images may exceed the GPU memory capacity. 2) Small stroke size may cause unpleas-

随着深度学习的发展，神经样式迁移已经取得了显著的成功[3,10,14,15,17,20,24,28,30]，但是超高分辨率的样式迁移在这些工作中很少被探索。在自然场景中，超高分辨率的图像经常出现在大型海报、摄影作品和超高清(例如 8K)视频中。传输超高分辨率图像的主要挑战有两个: 1)超高分辨率图像的海量存储成本可能超过 GPU 的存储容量。2)小笔画尺寸可能导致不愉快

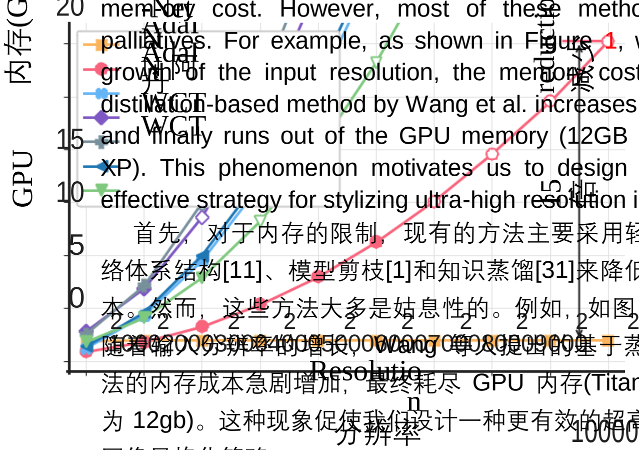
these res-olutions cannot be rendered on a 12GB GPU (Titan XP) due to memory limitation. By contrast, our URST not only can process images of arbitrary high-resolution, but also achieves a 15 re-duction in memory when stylizing an ultra-high resolution image of 10000 10000 pixels by using Wang et al. 's method [31].

图 1:不同风格转换方法的 GPU 内存比较。中空标记表明，由于内存限制，这些分辨率的图像不能在 12gb 的 GPU (Titan XP)上呈现。相比之下，我们的 URST 不仅可以处理任意高分辨率的图像，而且使用 Wang 等[31]的方法对 10000-10000 像素的超高分辨率图像进行风格化处理时，还可以在内存上减少 15 个像素。

ant dense textures in ultra-high resolution results.

蚂蚁密集纹理在超高分辨率的结果。

First, for the memory limitation, the existing methods mainly use lightweight network architecture [11], model pruning [1], and knowledge distillation [31] to reduce memory cost. However, most of these methods are palliatives. For example, as shown in Figure 1, with the growth of the input resolution, the memory cost of the distillation-based method by Wang et al. increases sharply and finally runs out of the GPU memory (12GB in Titan XP). This phenomenon motivates us to design a more effective strategy for stylizing ultra-high resolution images.

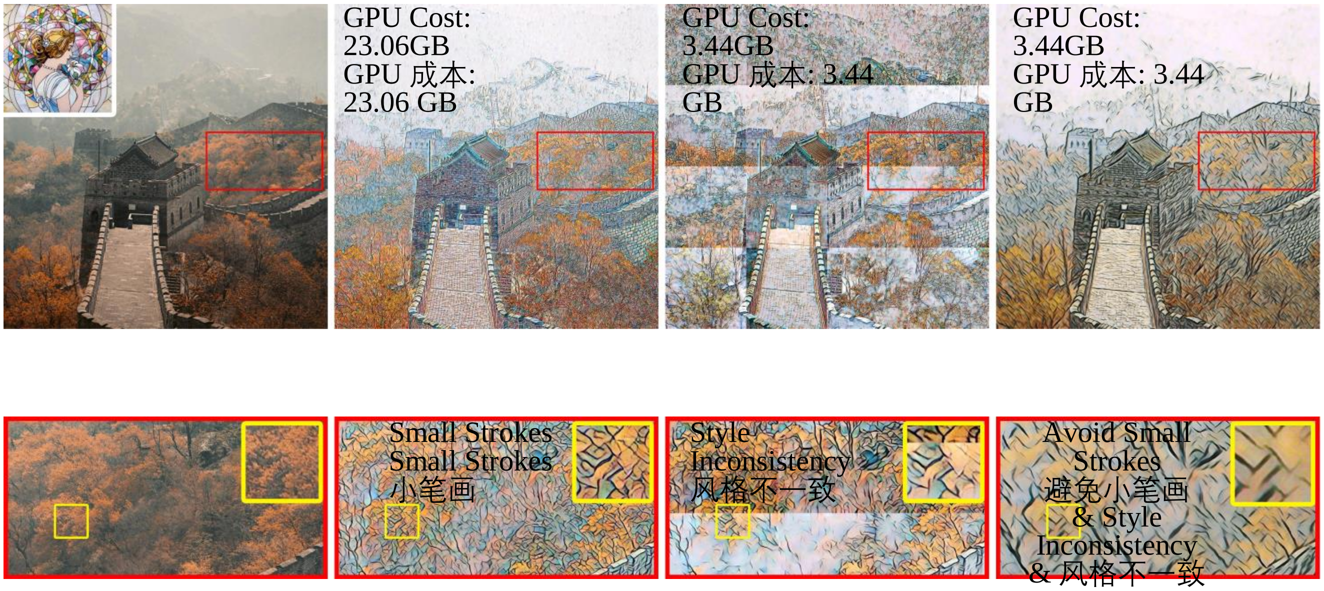


首先，对于内存的限制，现有的方法主要采用轻量级网络体系结构[11]、模型剪枝[1]和知识蒸馏[31]来降低内存成本。然而，这些方法大多是姑息性的。例如，如图 1 所示，随着输入分辨率的增长，Wang 等人提出的基于蒸馏的方法的内存成本急剧增加，最终耗尽 GPU 内存(Titan XP 中为 12gb)。这种现象促使我们设计一种更有效的超高分辨率图像风格化策略。

Figure 1: GPU memory comparison of different style transfer methods. The hollow markers indicate that images of

The second problem is that the brush strokes in ultra-high resolution stylized results are relatively small. Many previous works [12, 32] have shown that the style trans-fer network tends to produce stylized textures of the same stroke size. As a result, when given a high-resolution in-put, the model with small brush strokes would produce un-pleasant dense textures, as shown in Figure 2 (b). Enlarg-

第二个问题是超高分辨率风格化结果中的笔触相对较小。许多以前的作品[12,32]已经表明，风格转移网络倾向于产生相同笔画大小的风格化纹理。因此，当输入高分辨率时，使用小笔刷的模型会产生不愉快的致密纹理，如图 2(b)所示。放大



(a) Content & Style (b) IN (c) IN + Patch-wise (d) TIN + Patch-wise + $\mathcal{L}$ (ours)
(a)内容及样式(b)(c) IN + Patch-wise (d) TIN + Patch-wise + l (ours)

Figure 2: Ultra-high resolution stylized results (4000 4000 pixels) generated by different methods. (a) Content image and style image. (b) Result produced by the original instance normalization (IN), which costs 23.06GB GPU memory. Compared with our result (d), its brush strokes are relatively small. (c) Result produced by patch-wise style transfer with IN, which only costs 3.44GB, but its stylized patches are inconsistent in style. (d) Result produced by patch-wise style transfer with the proposed thumbnail instance normalization (TIN) and stroke perceptual loss L_{sp}, which also costs 3.44GB. Here, TIN is designed to ensure the style consistency among different patches, and L_{sp} is aimed at enlarging the stroke size. Based on these components, we can obtain ultra-high resolution stylized results with large brush strokes under limited memory resources.
图 2: 不同方法生成的超高分辨率风格化结果(40004000 像素)。内容图像和样式图像。(b)原始实例标准化(IN)产生的结果，耗资 23.06 GB GPU 内存。与我们的结果(d)相比，它的笔画相对较小。(c)使用 IN 进行补丁样式传输的结果，只需 3.44 GB，但其样式化补丁的样式不一致。(d)使用所提出的缩略图实例归一化(TIN)和笔画知觉损失 Lsp 进行贴片样式转换所产生的结果，该方法的成本也为 3.44 GB。在这里，TIN 旨在确保不同补丁之间的风格一致性，Lsp 旨在扩大笔画大小。基于这些组件，我们可以在有限的内存资源下用大笔刷获得超高分辨率的样式化结果。

ing the stroke size is a widely-used approach to address this problem. At present, the existing methods for enlarging the stroke size can be mainly divided into two categories. One is to train or inference with large style images [13, 17, 35]. Another solution is to enlarge the receptive field of the style transfer network [12, 32]. However, these methods tend to take extra time and memory, are not suitable for ultra-high resolution style transfer.
行程大小是解决这个问题的一种广泛使用的方法。目前，现有的扩大笔画大小的方法主要可以分为两类。一个是用大样式图像训练或推断[13,17,35]。另一个解决方案是扩大风格转移网络的接受范围[12,32]。然而，这些方法往往需要额外的时间和内存，不适合超高分辨率的风格转换。

To compensate the above limitations, this work proposes an Ultra-Resolution Style Transfer framework, termed URST, which can perform style transfer on arbitrary high-resolution images under limited memory resources. Differ-ent from previous methods [1, 12, 31, 32], our method 1) takes small patches instead of a full image as input, which makes it possible to process ultra-high resolution input as large as 12000 8000 pixels, as shown in Figure 9. 2) We replace the original instance normalization (IN) [29] by the proposed thumbnail instance normalization (TIN), to ensure the style consistency among different patches. As shown in Figure 2 (c), if we perform patch-wise style trans-fer with IN, style inconsistency among different patches would make them cannot be assembled into a pleasing im-age. 3) We propose a stroke perceptual loss as an auxiliary loss for neural style transfer, which motivates style transfer networks to keep large brush strokes.

为了弥补上述缺陷，本文提出了一种超分辨率风格传输框架 URST，该框架可以在有限的内存资源下对任意高分辨率图像进行风格传输。与以前的方法[1,12,31,32]不同，我们的方法 1)使用小的补丁而不是完整的图像作为输入，这使得处理超高分辨率的输入成为可能，最大可达 120008000 像素，如图 9 所示。2)将原来的实例规范化(IN)[29]替换为提出的缩略实例规范化(TIN)，以保证不同补丁之间的样式一致性。如图 2(c)所示，如果我们使用 IN 执行补丁样式转换，不同补丁之间的样式不一致将使它们无法组装成令人满意的图像。3)我们提出笔画感知损失作为神经样式转换的辅助损失，这促使样式转换网络保持大笔刷笔画。

Overall, the proposed URST has the following advan-tages:

总体而言，拟议的城市电脑售票网有以下优点:

- Our framework can process arbitrary high-resolution images with limited memory. As illustrated in Fig-ure 1, when stylizing an ultra-high resolution image of 10000 10000 pixels using Wang et al. 's method, our framework only requires 1.94GB memory, while the original method needs 30.25GB, 15 times larger. As we know, it is the first unconstrained resolution style transfer method.

我们的框架可以在内存有限的情况下处理任意高分辨率图像。如图 1 所示，当使用 Wang 等人的方法对 1000010000 像素的超高分辨率图像进行风格化时，我们的框架只需要 1.94 GB 的内存，而原来的方法需要 30.25 GB 的内存，是原来的 15 倍。正如我们所知，它是第一个无约束分辨率样式转换方法。

- Our framework can achieve high-quality style transfer of ultra-high resolution images. As shown in Figure 2 (d), our method uses larger brush strokes to depict the scene, which is much better than the effects presented in Figure 2 (b).

我们的框架可以实现超高分辨率图像的高质量风格转换。如图 2(d)所示，我们的方法使用更大的笔触来描绘场景，这比图 2(b)所示的效果要好得多。

- Our framework can be easily plugged into most ex-isting style transfer methods. Even without training, our framework can also obtain high-quality stylized re-sults.

我们的框架可以很容易地插入大多数现有的风格转换方法。即使没有训练，我们的框架也可以获得高质量的风格化结果。

2. Related Work
相关工作

Neural Style Transfer. Inspired by the success of con-volutional neural networks (CNNs), Gatys et al. first pro-posed a CNN-based style transfer algorithm [7], which opens up the new research field. To accelerate neural style transfer, Johnson et al. [14] and Ulyanov et al. [28] at-tempted to train a feed-forward network to learn a spe-cific artistic style. Compared with Gatys et al. 's method,

神经样式转换。受到卷积神经网络(cnn)成功的启发，Gatys 等人首次提出了一种基于 cnn 的样式转移算法[7]，开辟了新的研究领域。为了加速神经风格的转换，Johnson 等[14]和 Ulyanov 等[28]试图训练一个前馈网络来学习一种特定的艺术风格。与 Gatys 等人的方法相比，

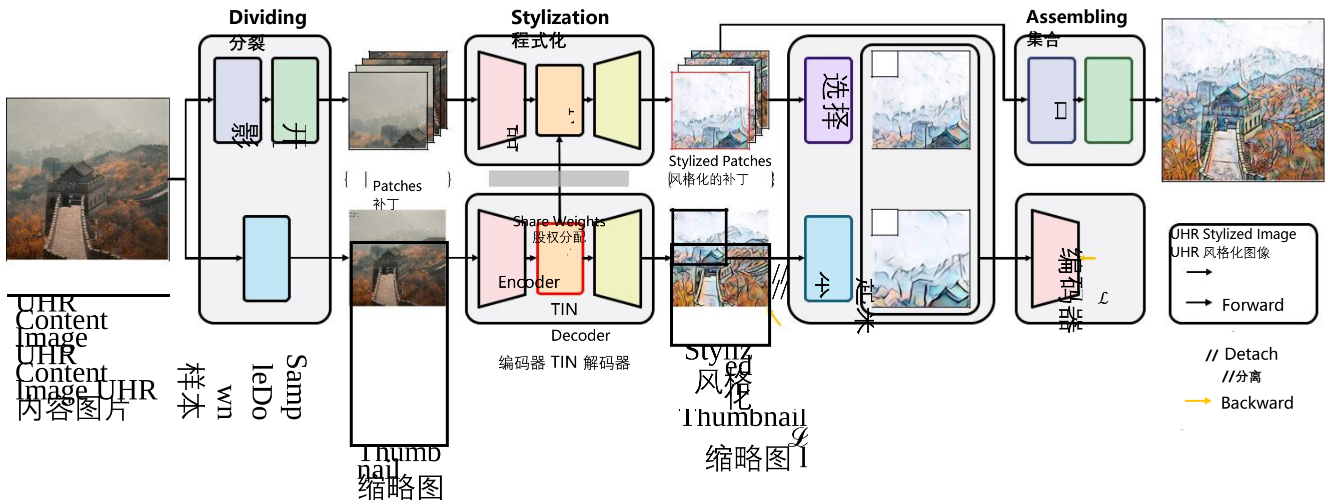


Figure 3: Overall architecture of the proposed URST framework. Its data flow is divided into three stages: dividing, stylization, and assembling. The core idea of the URST is to divide the ultra-high resolution (UHR) content image into small patches and perform patch-wise style transfer with the proposed TIN. The style transfer network in our framework can be different methods. In addition to the original loss L_O of the selected method, our URST includes an auxiliary loss function termed stroke perceptual loss L_{sp}, to enlarge the stroke size. Thanks to the above key designs, we for the first time build an ultra-high resolution style transfer system.
图 3: 拟议 URST 框架的总体架构。它的数据流分为三个阶段: 划分，样式化和组装。URST 的核心思想是将超高分辨率(UHR)内容图像分割成小块，并利用提出的 TIN 进行逐块样式转换。我们框架中的样式传输网络可以是不同的方法。除了选择的方法的原始损失 Lo，我们的 URST 包括一个辅助损失函数称为笔画知觉损失 Lsp，以扩大笔画大小。由于上述关键设计，我们首次建立了一个超高分辨率风格的传输系统。

these methods are more than 1,000 times faster. In recent years, to improve the efficiency of transferring new styles, researchers have proposed many multiple style trans-fer [2, 6, 16, 35] and arbitrary style transfer (i.e., universal style transfer) [5, 9, 10, 15, 17, 19, 25, 33] methods, which can render content images in multiple or arbitrary styles through a single model. Nowadays, neural style transfer has achieved great success, but due to massive memory cost and small stroke size, ultra-high resolution style transfer is still challenging.
这些方法的速度要快 1000 倍以上。近年来，为了提高新样式转换的效率，研究人员提出了多种多样式转换方法[2,6,16,35]和任意样式转换方法(即通用样式转换)[5,9,10,15,17,19,25,33]。目前，神经样式转换已经取得了很大的成功，但是由于大量的记忆代价和笔画尺寸较小，超高分辨率的样式转换仍然具有挑战性。

High-Resolution Neural Style Transfer. GPU memory is the main factor that restricts high-resolution style transfer. A direct way to alleviate this problem is to use a lightweight model. An et al. [1] proposed ArtNet, a lightweight network pruned from GoogLeNet [27] via a channel pruning method designed for neural style transfer. Jing et al. [11] developed a MobileNet-based lightweight style transfer network, lead-ing to a significant reduction of computation complexities compared with the original VGG encoder. The distillation-based method by Wang et al. [31] is another solution, which uses the pre-trained VGG-19 [26] as the teacher and a small encoder as the student, successfully rendering high-resolution images up to 6000 6000 pixels on a single 12GB GPU. Although these lightweight methods reduce the mem-ory cost for neural style transfer to a certain extent, they will exhaust the GPU memory when processing ultra-high reso-lution images (e.g., 10000 10000 pixels).
高分辨率神经样式传输。GPU 内存是限制高分辨率风格转换的主要因素。缓解这个问题的一个直接方法是使用轻量级模型。一个等人[1]提出了 ArtNet，一个从 GoogLeNet [27]修剪的轻量级网络，通过为神经样式转移设计的通道修剪方法。Jing 等[11]开发了一种基于 mobilenet 的轻量级传输网络，与最初的 VGG 编码器相比，大大降低了计算复杂度。Wang 等[31]提出的基于蒸馏的方法是另一种解决方案，它使用预先训练好的 VGG-19[26]作为教师，使用小型编码器作为学生，成功地在一个 12gb 的 GPU 上绘制高达 60006000 像素的高分辨率图像。虽然这些轻量级的方法在一定程度上降低了神经样式传输的内存成本，但是在处理超高分辨率图像(如 1000010000 像素)时，它们会耗尽 GPU 内存。

Stroke Size Control in Neural Style Transfer. Stroke size is an important perceptual factor highly related to the quality of style transfer results. Typically, a stylized result with large brush strokes tends to have a better appearance than the one with small brush strokes. Gatys et al. [8] first presented that the stroke size is related to the recep-
Stroke Size Control in Neural Style Transfer 神经样式传输中的笔画大小控制。笔画大小是一个重要的感知因素，与风格转换结果的质量高度相关。通常情况下，大笔刷的样式化

结果往往比小笔刷的样式化结果有更好的外观。Gatys 等[8]首先提出笔画大小与接收相关

tive field of the loss network, and they proposed a coarse-to-fine method to generate stylized results with large brush strokes. Wang et al. [32] proposed a hierarchical network to enlarge the stroke size and trained it with multiple losses of increasing scales. Similarly, Jing et al. [12] presented a style-specific network with multiple stroke branches, supervised by multi-scale style images, which achieves continuous stroke size control. Furthermore, Zhang et al. [35] proposed a multi-style generative network named MSG-Net, which controls the stroke size by using style images of different sizes for inference. These stroke size control methods are mainly designed for the style transfer under common image resolution (e.g., 1000 1000 pixels). It is difficult to apply them in ultra-high resolution scenarios, where the stylized results are often unsatisfying due to the small stroke size (see Figure 2 (b)).

他们提出了一种由粗到细的方法来生成大笔画的程式化结果。Wang 等[32]提出了一种分层网络来扩大笔画大小，并通过增加规模的多个损失来训练它。类似地，Jing 等[12]提出了一种具有多个笔画分支的特定风格网络，由多尺度风格图像监视，实现连续的笔画大小控制。此外，Zhang 等[35]提出了一种多样式生成网络 MSG-Net，它通过使用不同大小的样式图像进行推理来控制笔画的大小。这些笔画大小控制方法主要用于常见图像分辨率(例如 10001000 像素)下的样式转换。在超高分辨率场景中应用它们是困难的，在这种场景中，由于笔画尺寸较小，风格化的结果往往不令人满意(参见图 2(b))。

3. Proposed Method
3. 建议的方法

3.1. Overall Architecture
3.1 总体架构

The goal of the URST framework is to overcome the difficulties in GPU memory limit and small brush strokes when processing ultra-high resolution images (e.g., 10000 10000 pixels). It consists of three key designs: 1) A flexible pipeline termed patch-wise style transfer that can convert a high-cost style transfer task to multiple low-cost patch stylization. 2) A novel thumbnail instance normal-ization (TIN) layer that can extract thumbnail's normaliza-tion statistics and apply them to small patches, ensuring the style consistency among different patches. 3) A carefully defined stroke perceptual loss that focuses on the percep-tual differences in brush strokes, encouraging style transfer networks to keep large stroke size. Thanks to these versatile designs, our URST can be easily plugged into most existing methods [6, 10, 14, 15, 17, 30, 31, 35] to perform ultra-high

URST 框架的目标是在处理超高分辨率图像(如 1000010000 像素)时克服 GPU 内存限制和小笔刷的困难。它包括三个关键

设计: 1)一个灵活的管道，称为补丁式传输，可以将高成本的传输任务转换为多个低成本的补丁风格化。2)一种新颖的缩略图实例标准化(TIN)层，可以提取缩略图的标准化统计信息，并将其应用于小补丁，保证不同补丁之间的样式一致性。3)一个精心定义的笔画感知损失，关注笔刷笔画的感知差异，鼓励风格转换网络保持大笔画尺寸。由于这些多功能的设计，我们的 URST 可以很容易地插入到大多数现有的方法 [6,10,14,15,17,30,31,35]，以执行超高

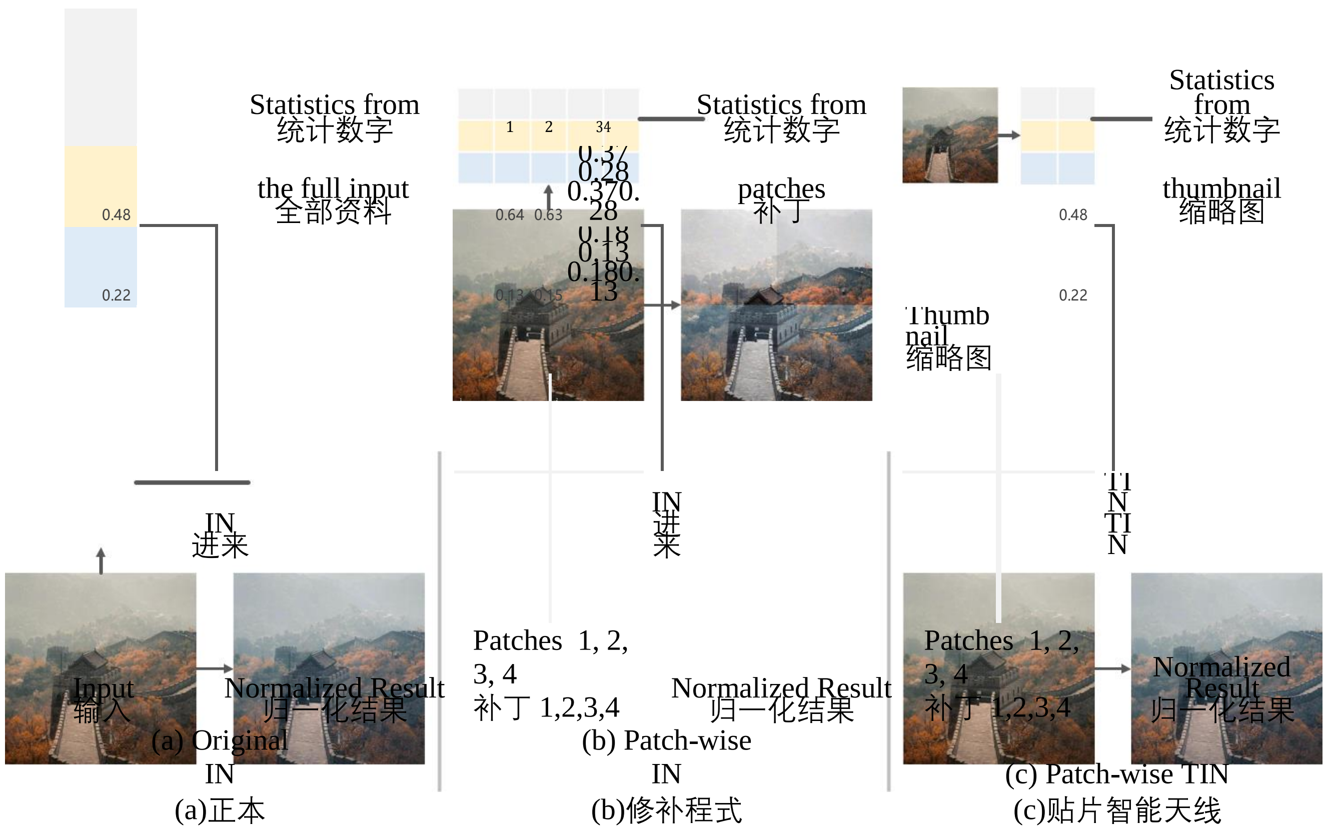


Figure 4: A simple example of IN and the proposed TIN. (a) We normalize the input as a whole. (b) We divide the input into four patches and normalize them individually. (c) We apply thumbnail's normalization statistics to these four patches, obtaining a similar output as (a). These results show that IN is not applicable to patch-wise style transfer.
图 4: IN 和拟议 TIN 的简单示例。(a)我们将输入作为一个整体标准化。(b)我们将输入划分为四个补丁，并分别对它们进行标准化。(c)我们将缩略图的标准化统计应用于这四个补丁，获得与(a)类似的输出。这些结果表明 IN 不适用于补丁样式转移。

resolution style transfer.
分辨率样式转换。

An overview of the URST framework is depicted in Fig-ure 3. Taking an ultra-high resolution content image I_c as input, the data flow of the URST can be divided into three stages: dividing, stylization, and assembling. 1) In the di-viding stage, we first generate a thumbnail image I_t for each content image, and then divide the content image I_c into a sequence of small patches fI_p^i j i = 1; 2; :::; Ng. 2) In the stylization stage, the thumbnail image I_t is the first to be fed into the style transfer network, to collect the normaliza-tion statistics across the network. Then, these normalization statistics are applying to stylize the small patches, obtain-
URST 框架的概述如图 3 所示。以超高分辨率内容图像集成电路为输入，将 URST 的数据流分为划分、程式化和装配三个阶段。1)在分割阶段，我们首先为每个内容图像生成一个缩略图 It，然后将内容图像 Ic 分割成一个小块序列 fIpi j i = 1; 2; : : ; Ng。2)在样式化阶段，首先将缩略图输入样式传输网络，收集整个网络的规范化统计数据。然后，这些标准化统计信息应用于小补丁的样式化，得到 -

ing the stylized patches $\hat{I}_p$ i = 1; 2; :::; N . Here, our style transfer network is not specific. Most existing meth-ods [6, 10, 14, 15, 17, 30, 31, 35] can be used as the style transfer network. 3) In the assembling stage, all stylize stylized
定义风格化的补丁 Ip i = 1; 2; : : : ; n。在这里，我们的风格传输网络不是特定的。格转换网络。3)在装配阶段，所有风格化的补丁都被装配成超高分辨率的风格化

image $\hat{I}_c$.
图像 Ic。

Since the style transfer network in our framework can be different methods (e.g., A various, for convenience, we de-fine the loss of the selected method as the original lo network with the original loss calculated on the stylized thumbnail. In addition to the ori termed stroke perceptual loss, to fur-ther improve the quality of ultra-high resolution differences
由于我们框架中的样式传递网络可以是损耗函数不同的不同方法(如 AdaIN [10]和 Lin 义为原始损耗 Lo。在培训期间，我们首先优化样式转移网络，在样式化缩略图上计算原 们引入了一种称为笔画感知损耗的辅助损耗函数。其核心思想是惩罚感知差异

in brush strokes between the stylized patch $\hat{I}_p$ and the up-
在风格化的补丁 Ip 和向上的笔触之间

sampled patch $\hat{I}_{tp}$ that cropped from the stylized thumbnail
从样式化缩略图裁剪的样本补丁 Itp

$\hat{I}_t$. It should be noticed that the upsampled patch $\hat{I}_{tp}$ plays
应该注意的是 Itp 播放的上升采样补丁

a role of the learning target. Therefore, the gradient flow of
一个角色的学习目标。因此，梯度流

$\hat{I}_{tp}$ is detached.
Itp 是分离的。

3.2. Patch-wise Style Transfer 3.2. 补丁风格转换

To process ultra-high

resolution images
high resolution c
with a stride o
overlapping patc
memory resource
After the loop of p

为了处理超高
分辨率的内容图像
像集成电路划分为
内存资源，这些补

tion, we obtain
Finally, we dis
patches and ass
我们得到了一系
弃这些程式化补
image $\hat{I}_c$.
图像 Ic。

Compared wi
input, our patch-
resolution image
transfer methods
LinearWCT [15].
Sec-tion 1, the
individual norma
work, we propose
this problem.

与以往使用完
以灵活地处理任
式转换方法，如广
但是，它有一个
算的个体标准化
中，我们提出了一

3.3. Thumbna
3.3 缩略图实例

IN is a widel
Given an input

IN 是神经样
RN c h w，IN 可

IN
IN

where (x); (x)
trainable affine

applicable to patch-wise style transfer, since stylized patches gen

erated by the IN
demonstrated i
其中(x) ; (x)2 R
然而，我们发现
生成的程式化补

Johnson *et al.* MSG-Net AdaIN WCT LinearWCT Wang *et al.*

Johnson 等人。 MSG-Net AdaIN WCT LinearWCT Wang 等人。

Figure 5: Stylization comparison of different pipelines and normalization methods (3000 3000 pixels). To verify the versatility of our approach, we conduct experiments on 6 representative style transfer methods, including Johnson et al. [14], MSG-Net [35], AdaIN [10], WCT [17], LinearWCT [15], and Wang et al. [31]. The style image and content image are the same with Figure 2. (a) shows the results directly generated by IN-based networks, which cost massive GPU memory. (b) and (c) present the results of patch-wise style transfer with IN and the proposed TIN, which cost much less memory than (a). It is worth noting that our results (c) are as high-quality as (a), which demonstrates the effectiveness of our TIN.

图 5: 不同管道的风格化比较和标准化方法(30003000 像素)。为了验证我们方法的通用性，我们对 6 种有代表性的风格转换方法进行了实验，包括 Johnson 等[14]、 MSG-Net [35]、 AdaIN [10]、 WCT [17]、 LinearWCT [15]和 Wang 等[31]。样式图像和内容图像与图 2 相同。(a)显示了基于 in- 的网络直接生成的结果，这需要大量的 GPU 内存。(b)和(c)提出了使用 IN 和所提出的 TIN 进行补丁样式转换的结果，这种方法的内存消耗比(a)小得多。值得注意的是，我们的结果(c)与(a)一样高质量，这证明了我们的 TIN 的有效性。

take a simple example to illustrate this problem. In Figure 4 (a), we normalize the input as a whole. In Figure 4 (b), we divide the input into four patches and normalize them individually. Comparing these two results reveals that the root cause of style inconsistency is the individual normalization statistics calculated from each patch.

以一个简单的例子来说明这个问题。在图 4(a)中，我们将输入作为一个整体进行规范化。在图 4(b)中，我们将输入划分为四个补丁，并分别对它们进行标准化。比较这两个结果显示，样式不一致的根本原因是从每个补丁计算的个体归一化统计量。

Based on the above analysis, we propose a simple variant to IN, termed thumbnail instance normalization (TIN). Our TIN receives a patch x 2 $R^{N\ C\ H\ W}$ and a thumbnail

在上述分析的基础上，我们提出了一种简单的 IN 变体，称为缩略实例规范化(TIN)。我们的 TIN 接收到一个补丁 x 2 RN c h w 和一个缩略图

t 2 $R^{N\ C\ H_t\ W_t}$ as input, and it can be formulated as:

2 RN c Ht Wt 作为输入，可以表示为:

TIN(x; t) = x (t) X (t) + : (2)

TIN (x; t) = (t) (t) + : (2)

Different from IN, here (t); (t) 2 R^N C are channel-wise mean and standard deviation of the thumbnail input t. In this way, our TIN is able to ensure the style consistency among different patches, as shown in Figure 4 (c).

与 IN 不同，这里(t; (t)2rn c 是缩略图输入 t 的通道均值和标准差。通过这种方式，我们的 TIN 能够确保不同补丁之间的样式一致性，如图 4(c)所示。

Similarly, instance whitening (IW) transformation [22] has the same problem, which is a standardization method based on second-order statistics (i.e., covariance matrix). It is also widely used in many neural style transfer meth-ods [15, 17, 31, 34]. Our TIN can be generalized to IW with minor modifications. We will discuss this in detail in the supplementary material.

同样，实例白化(IW)变换[22]也有同样的问题，这是一种基于二阶统计量(即协方差矩阵)的标准化方法。它也广泛用于许多神经样式转移方法[15,17,31,34]。我们的 TIN 可以通过小的修改推广到 IW。我们将在补充材料中详细讨论这一点。

3.4. Stroke Perceptual Loss

3.4 中风感知损失

This inspired us to propose an auxiliary loss function for enlarging the stroke size, named stroke perceptual loss:

这促使我们提出了一个扩大笔画大小的辅助损失函数，称为笔画知觉损失:

Lsp(I^p; I^tp) = l(I^t p)

Lsp (i ^ p; i ^ Fl(I^p) F L (i (3)

tp) = Fl (i ^ p) f ^ tp) 2 ; (3)

where Fl is the output feature map of the layer l in the VGG

层的输出特征映射在哪里 在 VGG 中

network. Îp is a stylized patch with small brush strokes, and

Ip 是一个带有小笔画的风格化补丁

Îtp is a stylized patch that cropped and upsampled from the

Itp 是一个样式化的补丁，从

stylized thumbnail Ît, which has large brush strokes.

样式化的缩略图 It，具有较大的笔画。

The stroke perceptual loss follows the design rules in [7]

笔画知觉损失遵循[7]中的设计规则

and [14]. Since the input pair (Îp; Îtp) has similar con-tent and style, but the stroke size is different, the proposed L_{sp} can mainly measure the perceptual differences in brush strokes. Therefore, optimize this loss is to encourage the

和[14]。由于输入对(Ip; Itp)具有相似的内容和风格，但笔画大小不同，所提出的 Lsp 主要能够测量笔画的知觉差异。因此，优化这种损失是为了鼓励

style transfer network to generate brush strokes as large as

风格转移网络，以生成画笔笔触的大小，如

that of the target Î$_{tp}$.

目标 Itp 的资料。

3.5. Total Loss

3.5 总损失

As discussed in Section 3.1, we define the loss function used in the selected method as the original loss L_o. On this basis, we add the stroke perceptual loss L_{sp} as an auxiliary

如第 3.1 节所讨论的，我们将所选方法中使用的损耗函数定义为原始损耗 Lo。在此基础上，我们添加笔画感知损失 Lsp 作为辅助

loss. Therefore, the total loss is expressed as:
因此，总损失表示为:

$$L = Lo + Lsp; \quad (4)$$

$$L = Lo + Lsp; \quad (4)$$

where is a weight to balance L_O and L_{SP}. In our experi-ments, is set to 1:0 by default.
在我们的实验中，在缺省情况下，Lo 和 Lsp 的权重设置为 1:0。

4. Experiments
实验

4.1. Experimental Settings
4.1. 实验环境

To verify the versatility of our URST, we apply it to 6 representative style transfer methods, including Johnson et al. [14], MSG-Net [35], AdaIN [10], WCT [17], Linear-WCT [15], and Wang et al. [31].

为了验证 URST 的通用性，我们将其应用于 6 种有代表性的风格转换方法，包括 Johnson 等[14]、MSG-Net [35]、 AdaIN [10]、 WCT [17]、 linear-WCT [15]和 Wang 等[31]。

In the testing phase, we perform ultra-high resolu-tion style transfer on photography works collected from pexels.com and unsplash.com. By default, we use a 1064 1064 pixels sliding window with a stride of 1000 to divide the input image, and the style image used in our framework is 1024 1024 pixels. When generating the thumbnail, we rescale the input image to have the shorter side of 1024 pixels.

在测试阶段，我们对从 pixels.com 和 unsplash. com 收集的摄影作品进行超高分辨率风格转换。默认情况下，我们使用 1000 步长的 10641064 像素的滑动窗口来划分输入图像，在我们的框架中使用的样式图像是 10241024 像素。当生成缩略图时，我们重新调整输入图像的大小，使其具有 1024 像素的较短一侧。

During training, our stroke perceptual loss is computed at the relu4 1 layer of the VGG network, and the weight

在训练过程中，我们的笔画感知损失是在 VGG 网络的 relu41 层和权重上计算的

is set as 1. Following common practices [3, 5, 10, 15], we use MS-COCO dataset [18] as content images and WikiArt dataset [21] as style images, both of which con-tain roughly 80,000 training samples. For style images, we rescale the shorter side to 256 pixels, and then randomly crop to 256 256 pixels. Differently, for content images, we first rescale the shorter side of each to 512 pixels and randomly crop a region of 512 512 pixels from it. Then, we randomly crop a patch and generate a thumbnail, both of which are 256 256 pixels. Following previous meth-ods [7, 10, 15, 28], we adopt a VGG-19 [26] pre-trained on ImageNet [4] as the loss network. All models are trained with a batch size of 8 on a Titan XP GPU, and other training settings are the same as the original settings in the selected style transfer methods [10, 15].

设置为 1。按照常规做法[3,5,10,15]，我们使用 MS-COCO 数据集[18]作为内容图像，使用 WikiArt 数据集[21]作为样式图像，两者都包含大约 80,000 个训练样本。对于样式图像，我们将较短的一侧重新缩放到 256 像素，然后随机裁剪到 256256 像素。不同的是，对于内容图像，我们首先将每个图像的较短一侧重新缩放为 512 像素，然后从中随机裁剪出 512512 像素的区域。然后，我们随机裁剪一个补丁并生成一个缩略图，两者都是 256256 像素。根据以往的方法[7,10,15,28]，我们采用了一个 VGG-19[26]预先训练的 ImageNet [4]作为损失网络。所有的模型都在 Titan XP GPU 上进行批量大小为 8 的培训，其他的培训设置与所选样式转换方法中的原始设置相同[10,15]。

4.2. Ablation Study
4.2. 消融研究

Thumbnail Instance Normalization. As discussed in Section 3.3, consistent normalization statistics are crucial for patch-wise style transfer. To verify this, we conduct experiments of patch-wise style transfer with IN and the proposed TIN, respectively. From Figure 5 (b) we can ob-serve that IN leads to the style inconsistency among differ-ent patches. Differently, our method avoids this problem by adopting TIN (see Figure 5 (c)). In addition, we find that our results are as high-quality as the results demonstrated in Figure 5 (a), while our memory cost is less than 5GB, showing that our TIN can approximate the IN statistics of the original ultra-high resolution image, enabling the low memory cost ultra-high resolution style transfer.
缩略图实例标准化。正如第 3.3 节所讨论的，一致的标准化统计对于补丁样式转移是至关重要的。为了验证这一点，我们分别用 IN 和提出的 TIN 进行补丁样式转移的实验。从图 5(b)中，我们可以观察到 IN 导致不同补丁之间的样式不一致。不同的是，我们的方法通过采用 TIN 来避免这个问题(见图 5(c))。此外，我们发现我们的结果与图 5(a)中的结果一样高质量，而我们的内存成本低于 5gb，这表明我们的 TIN 可以接近原始超高分辨率图像的 IN 统计，从而实现低内存成本的超高分辨率样式转移。

(a)内容和样式(b)缩略图统计(c)随机统计

Figure 6: Thumbnail statistics vs. random statistics. This comparison demonstrates that using thumbnail's normalization statistics is the key to the success of patch-wise style transfer.
图 6: 缩略图统计与随机统计。这个比较表明使用缩略图的标准化统计是成功的补丁样式转移的关键。

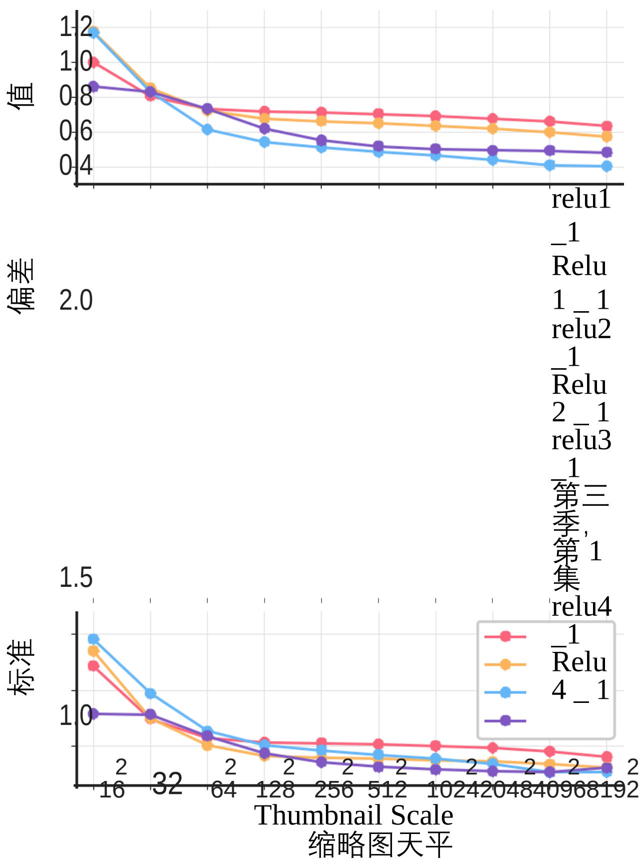


Figure 7: Mean and standard deviation of feature maps of the VGG-19 network under different input scales. It shows that with the growth of the thumbnail scale, the normalization statistics of feature maps tend to be stable.

图 7: 不同输入尺度下 VGG-19 网络特征图的平均值和标准差。它表明，随着缩略图规模的增长，特征图的归一化统计趋于稳定。

Moreover, we compare the stylized results generated by the model with our TIN and the model with random normalization statistics, in Figure 6. Although using random normalization statistics can also keep the style consistency among different patches, it destroys the style information extracted from the style image, resulting in the unexpected styles as shown in Figure 6 (c). In contrast, using TIN not only ensures the style consistency among different patches, but also maintains the information of the target style.

此外，我们将模型生成的程式化结果与 TIN 和具有随机归一化统计量的模型进行比较，如图 6 所示。尽管使用随机归一化统计信息也可以保持不同补丁之间的样式一致性，但它破坏了从样式图像中提取的样式信息，导致出现图 6(c)所示的意外样式。相比之下，使用 TIN 不仅可以确保不同补丁之间的样式一致性，而且可以保持目标样式的信息。

Thumbnail Size. To further study the relationship be-tween normalization statistics and thumbnail size, we rescale an ultra-high resolution image (8192 8192 pixels) to the thumbnails of different scales, and calculate their nor-
缩略图尺寸。为了进一步研究归一化统计量与缩略图尺寸之间的关系，我们将一幅超高分辨率图像(81928192 像素)重新缩放到不同尺寸的缩略图上，并计算其归一化统计量与缩略图尺寸之间的关系

Figure 8: Ablation study of the proposed stroke perceptual loss L_{sp}. Comparison of these stylized images (3000 3000 pixels) indicates that our L_{sp} can significantly enlarge the stroke size of the existing style transfer methods.

图 8: 拟议的中风感知损失 Lsp 的消融研究。这些样式化图像(30003000 像素)的比较表明，我们的 Lsp 可以显着扩大现有样式转移方法的笔画大小。

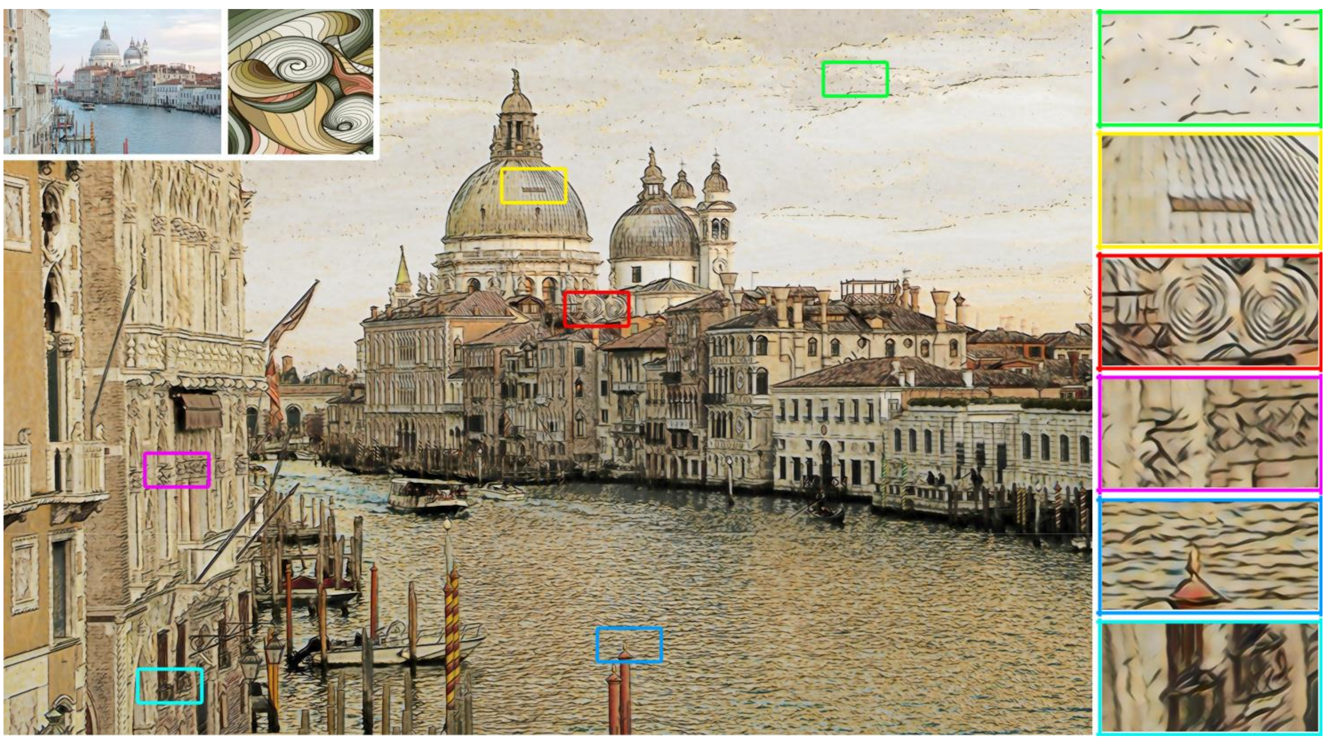


Figure 9: An ultra-high resolution stylized result (12000 8000 pixels), took about 2.5GB memory on a single Titan XP GPU (12GB).

图 9: 一个超高分辨率的样式化结果(120008000 像素)，在单个 Titan XP GPU (12gb)上占用大约 2.5 GB 的内存。

On the upper left are the content image and style image. Six close-ups (660 330 pixels) are shown on the right side of the stylized result.

左上角是内容图像和样式图像。6 个特写(660330 像素)显示在风格化结果的右侧。

malization statistics in the style transfer network. Specif-ically, we first feed these thumbnails to the encoder (i.e., VGG-19) of the style transfer network and obtain the output feature maps

of relu1 1, relu2 1, relu3 1, and relu4 1. Then, we calculate the mean and standard deviation of these

风格转移网络中的非均匀化统计。具体地说，我们首先将这些缩略图提供给样式传输网络的编码器(即 VGG-19)，并获得

relu11、 relu21、 relu31 和 relu41 的输出特征映射。然后，我们计算这些的平均值和标准差

feature maps, and plot them in Figure 7. Note that, when the thumbnail scale is equal to 8192 8192, the normalization statistics is the IN statistics. We see that with the growth of the thumbnail scale, the normalization statistics of feature maps tend to be stable. When the thumbnail scale is larger

特征映射，并将它们绘制在图 7 中。注意，当缩略图刻度等于 81928192 时，归一化统计是 IN 统计。我们看到随着缩略图比例的增长，特征映射的归一化统计趋于稳定。当缩略图比例变大时

Table 1: GPU memory comparison (in GB) for a single image of different resolutions. All results are tested by PyTorch [23] on a Titan XP GPU (12GB). " " denotes that images of these resolutions cannot be rendered due to memory limitation.
表 1: 不同分辨率的单个图像的 GPU 内存比较(以 GB 为单位)。所有结果都通过 PyTorch [23]在 Titan XP GPU (12gb)上进行测试。“”表示由于内存限制无法呈现这些分辨率的图像。

Resolution 分辨率	Johnson et al. [14] Johnson 等[14]		MSG-Net [35] MSG-Net [35]		AdaIN [10] AdaIN [10]		WCT [17] WCT [17]		LinearWCT [15] LinearWCT [15]		Wang et al. [31] Wang 等人 [31]	
	Original 原创的	Ours 我们的	Original 原创的	Ours 我们的	Original 原创的	Ours 我们的	Original 原创的	Ours 我们的	Original 原创的	Ours 我们的	Original 原创的	Ours 我们的
1000^2	1.21	2.12	2.55	3.21	1.40	2.17	1.84	4.61	2.79	3.44	0.90	1.94
2000^2	4.31	2.12	7.15	3.21	4.62	2.17	4.10	4.61	6.85	3.44	1.80	1.94
3000^2	9.48	2.12		3.21	9.95	2.17	7.87	4.61		3.44	3.27	1.94
4000^2		2.12		3.21		2.17		4.61		3.44	5.36	1.94
5000^2		2.12		3.21		2.17		4.61		3.44	8.00	1.94
6000^2		2.12		3.21		2.17		4.61		3.44	11.28	1.94
7000^2		2.12		3.21		2.17		4.61		3.44		1.94
8000^2		2.12		3.21		2.17		4.61		3.44		1.94
9000^2		2.12		3.21		2.17		4.61		3.44		1.94
10000^2		2.12		3.21		2.17		4.61		3.44		1.94

than 1024 1024 pixels, the TIN statistics are very close to IN statistics. This indicates that TIN can well approximate IN when the thumbnail scale is larger than 1024 1024. To balance speed and style transfer quality, we set the shorter side of the thumbnail to 1024 pixels by default.
除 10241024 像素外，TIN 的统计数字与 IN 的统计数字非常接近。这表明当缩略图比例大于 10241024 时，TIN 可以很好地近似 IN。为了平衡速度和样式转换质量，我们默认将缩略图的较短一侧设置为 1024 像素。

Stroke Perceptual Loss. As shown in Figure 8, using the proposed stroke perceptual loss L_{sp} as an auxiliary loss for neural style transfer can significantly enlarge the stroke size of these existing methods. Compared with the baseline re-sults (see Figure 8 (b) and 8 (d)), under the guidance of L_{sp}, these models learned to use thicker lines and sparser textures to depict the scenery, which helps to improve the quality of ultra-high resolution style transfer (see Figure 8
Stroke perception Loss 笔画感知损失。如图 8 所示，使用提出的笔画知觉损失 Lsp 作为神经风格传递的辅助损失，可以显著扩大这些现有方法的笔画大小。与基线结果(见图 8(b)和 8(d))相比，在 Lsp 的指导下，这些模型学会了使用较粗的线条和较稀疏的纹理来描绘风景，这有助于提高超高分辨率风格转换的质量(见图 8)
(c) and 8 (e)). Moreover, the computation/memory over-head of our L_{sp} is negligible, which is suitable for the train-ing of ultra-high resolution style transfer.
及 8(e))。此外，我们的 Lsp 的计算/内存开销可以忽略不计，适合于超高分辨率风格转换的训练。

4.3. Discussion
4.3 讨论

Different from previous methods [1, 31], our URST is a versatile framework, not designed for a specific method. It can easily extend most of the existing methods to support arbitrary high-resolution style transfer under limited mem-ory resources (12GB in Titan XP).
与以前的方法[1,31]不同，我们的 URST 是一个通用的框架，不是为特定的方法设计的。它可以轻松地扩展大多数现有的方法，以支持在有限的内存资源(Titan XP 中为 12gb)下的任意高分辨率样式传输。

To demonstrate the effectiveness of our URST, we re-port the memory cost for a single content image of different resolutions. As listed in Table 1, most existing methods cannot process high-resolution images (e.g., 4000 4000 pixels) with limited memory. Wang et al. [31] is a recently proposed distillation-based method designed for high-resolution style transfer, but it can only process up to 6000 6000 pixels. Unlike them, our URST can keep the memory cost below 5GB, and with the growth of the in-put resolution, our memory cost hardly increases. Theoret-ically, our URST supports style transfer of arbitrary high-resolution images.
为了证明 URST 的有效性，我们报告了不同分辨率的单个内容图像的内存成本。如表 1 所示，大多数现有的

方法不能处理高分辨率图像(例如，40004000 像素)与有限的内存。Wang 等[31]是最近提出的一种基于蒸馏的高分辨率风格转移方法，但它只能处理高达 60006000 像素。与它们不同，我们的 URST 可以将内存成本保持在 5gb 以下，并且随着输入分辨率的增长，我们的内存成本几乎不会增加。理论上，我们的 URST 支持任意高分辨率图像的样式转换。

Also, we evaluate the proposed URST on an ultra-high resolution image of 12000 8000 pixels (i.e., 96 megapix-els), as shown in Figure 9. This result is produced based on AdaIN and only costs 2.5GB of GPU memory. Further-more, it also shows that our URST has achieved superior performance in producing large brush strokes due to the effectiveness of the stroke perceptual loss. In conclusion, to our knowledge, this is the first time to build an uncon-strained resolution style transfer system on a single 12GB GPU (Titan XP).

此外，我们在 120008000 像素(即 9600 万像素)的超高分辨率图像上评估建议的 URST，如图 9 所示。这个结果是基于 AdaIN 的，只需要 2.5 GB 的 GPU 内存。此外，由于笔画感知损失的有效性，我们的 URST 在产生大笔画方面取得了更好的性能。总之，据我们所知，这是第一次在单个 12gb GPU (Titan XP)上构建无张力分辨率风格的传输系统。

5. Conclusion
5. 结论

In this work, we propose URST, a simple yet effective framework for arbitrary high-resolution style transfer. We perform patch-wise style transfer to process ultra-high res-olution input under limited memory resources, and develop a thumbnail instance normalization (TIN) to ensure the style consistency among different patches. Moreover, to enlarge the brush strokes in ultra-high resolution results, a stroke perceptual loss is introduced as an auxiliary loss for neural style transfer. Extensive experiments show that our URST surpasses existing SOTA methods on ultra-high resolution images and can be easily plugged into most existing meth-ods. Although we mainly study neural style transfer in this work, the instance normalization is also widely used in other low-level vision tasks. Therefore, the application of our TIN on other tasks is worth exploring in the future.

在这项工作中，我们提出了 URST，一个简单而有效的框架，任意高分辨率的风格传输。在有限的内存资源下，对超高分辨率输入进行补丁样式转换，并开发了一种缩略实例规范化(TIN)，以保证不同补丁之间的样式一致性。此外，为了扩大超高分辨率结果中的笔画，笔画感知损失被引入作为神经样式转换的辅助损失。大量的实验表明，我们的 URST 在超高分辨率图像上超越了现有的 SOTA 方法，并且可以很容易地插入到大多数现有的方法中。虽然我们在这项工作中主要研究神经样式转换，实例归一化也被广泛应用于其他低水平视觉任务。因此，我们的 TIN 在其他任务上的应用值得在未来探索。

References
参考文献

[1] J. An, T. Li, H. Huang, L. Shen, X. Wang, Y. Tang, J. Ma, W. Liu, and J. Luo. Real-time universal style transfer on

high-resolution images via zero-channel pruning. arXiv preprint arXiv:2006.09029, 2020. 1, 2, 3, 4, 8
安建安、李泰利、黄海伦、沈丽珊、王旭、唐咏华、马建华、刘华和罗建华。基于零通道剪枝的高分辨率图像实时通用风格转换。arXiv 预印 arXiv: 2006.09029,2020。1,2,3,4,8
[2] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua. Stylebank: An explicit representation for neural image style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1897–1906, 2017. 3
丁、陈一元、廖锦松、余锦荣和华国华。Stylebank: 神经图像样式转换的显式表示。在 IEEE 计算机视觉和模式识别会议的议事录中，页 1897-1906,2017。3

[3] T. Q. Chen and M. Schmidt. Fast patch-based style transfer of arbitrary style. In Proceedings of the NIPS Workshop on Constructive Machine Learning, 2016. 1, 6
陈天赐和施密特先生。基于快速补丁的任意样式的样式转换。在 2016 年 NIPS 建设性机器学习研讨会会议记录中。1,6
[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, 2009. 6
邓华东，r。索彻，l.j。李启雄及李飞飞。Imagenet: 一个大规模的分层图像数据库。在 IEEE 计算机视觉和模式识别会议的会议记录中，页 248-255,2009。6
[5] Y. Deng, F. Tang, W. Dong, W. Sun, F. Huang, and C. Xu. Arbitrary style transfer via multi-adaptation network. In Proceedings of the 28th ACM International Conference on Multimedia, pages 2719–2727, 2020. 3, 6
邓耀华、唐伟、董伟东、孙伟、黄伟和徐振明。多元适应网络中的任意风格转换。在第 28 届 ACM 国际多媒体会议的议事录中，页 2719-2727,2020。3,6
[6] V. Dumoulin, J. Shlens, and M. Kudlur. A learned representation for artistic style. In Proceedings of the International Conference on Learning Representations, 2017. 3, 4
杜穆林，j。 Shlens，m。艺术风格的学术表现。2017 年学习表现国际会议记录。3,4
[7] 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. 2, 5, 6
洛杉矶盖蒂斯 A.s. 埃克尔和 m. 贝格。使用卷积神经网络的图像样式转换。在 IEEE 计算机视觉和模式识别会议记录中，页 2414-2423,2016。2,5,6
[8] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and
盖蒂斯，a · s · 埃克尔，m · 贝格，a · 赫兹曼，以及
E. Shechtman. Controlling perceptual factors in neural style transfer. In Proceedings of the IEEE Conference on Com-puter Vision and Pattern Recognition, pages 3985–3993, 2017. 3
Shechtman.控制神经风格转换中的感知因素。在 IEEE 计算机视觉和模式识别会议记录中，页 3985-3993,2017。3
[9] S. Gu, C. Chen, J. Liao, and L. Yuan. Arbitrary style transfer with deep feature reshuffle. In Proceedings of the IEEE Con-ference on Computer Vision and Pattern Recognition, pages 8222–8231, 2018. 3
S。 Gu，c。 Chen，J.Liao，and l。带有深度特征重组的任意风格转换。在 IEEE 计算机视觉和模式识别会议进程中，页 8222-8231,2018。3
[10] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, pages 1501–1510, 2017. 1, 3, 4, 5, 6, 8
黄先生和黄先生。自适应实例规范化实时任意样式转换。在 IEEE 国际计算机视觉会议记录中，页 1501-1510,2017。1,3,4,5,6,8
[11] Y. Jing, X. Liu, Y. Ding, X. Wang, E. Ding, M. Song, and
Y Jing，x。 Liu，y。 Ding，x。 Wang，e. Ding，m。 Song，and
S. Wen. Dynamic instance normalization for arbitrary style transfer. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 4369–4376, 2020. 1, 3
阿雯。任意样式转换的动态实例规范化。在 AAAI 人工智能会议记录，卷 34，页 4369-4376,2020。1,3
[12] Y. Jing, Y. Liu, Y. Yang, Z. Feng, Y. Yu, D. Tao, and M. Song. Stroke controllable fast style transfer with adaptive receptive fields. In Proceedings of the European Conference on Com-puter Vision, pages 238–254, 2018. 1, 2, 3
刘燕京，刘燕阳，Z.Feng，Yu，d。 Tao，and m。 Song。具有适应性接受场的卒中可控快速风格转换。在《欧洲计算机视觉会议进程》2018 年第 238-254 页。1,2,3
[13] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song. Neural style transfer: A review. IEEE transactions on Visualization and Computer Graphics, 26(11):3365–3385, 2019. 2
京、杨、冯、叶、于、宋明。神经类型转换: 综述。IEEE 可视化与计算机图形学汇编，26(11) : 3365-3385,2019。2
[14] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, pages 694– 711, 2016. 1, 2, 3, 4, 5, 6, 8
J · 约翰逊，a · 阿拉希和 l · 菲菲。实时风格转换和超分辨率的感知损失。欧洲计算机视觉会议记录，第 694-711 页，2016。1,2,3,4,5,6,8
[15] X. Li, S. Liu, J. Kautz, and M.-H. Yang. Learning linear transformations for fast image and video style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vi-sion and Pattern Recognition, pages 3809–3817, 2019. 1, 3, 4, 5, 6, 8
李、刘、考茨和 M.-H。杨先生。快速图像和视频样式转换的学习线性变换。在 IEEE/CVF 计算机视觉和模式识别会议记录中，页 3809-3817,2019。1,3,4,5,6,8
[16] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang. Diversified texture synthesis with feed-forward networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3920–3928, 2017. 3
Y。 Li，c。 Fang，j。 Yang，z。 Wang，x。 Lu，and M.-H。杨。基于前馈网络的多样化纹理合成。在 IEEE 计算机视觉和模式识别会议记录中，页 3920-3928,2017。3
[17] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang. Universal style transfer via feature transforms. In Advances
李毅、方正、杨振强、王泽、吕克希和 M.-H。杨。通过特征转换实现通用风格转换。进步

in Neural Information Processing Systems, pages 386–396, 2017. 1, 2, 3, 4, 5, 6, 8
在神经信息处理系统，页 386-396,2017.1,2,3,4,5,6,8

[18] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ra-manan, P. Dollar,´ and C. L. Zitnick. Microsoft coco: Com-mon objects in context. In Proceedings of the European Con-ference on Computer Vision, pages 740–755. Springer, 2014. 6
T.y.林，m。梅尔，s。贝隆吉，j。海斯，p。佩罗纳，d。拉马南，p。美元和 c。Microsoft coco: com-common objects in context 微软可可: 上下文中的通用对象。在《欧洲计算机视觉会议进程》(Proceedings of the European conference on Computer Vision)740-755 页。Springer，2014.6

[19] M. Lu, H. Zhao, A. Yao, Y. Chen, F. Xu, and L. Zhang. A closed-form solution to universal style transfer. In Proceed-ings of the IEEE/CVF International Conference on Com-puter Vision, pages 5952–5961, 2019. 3
陆先生，赵先生，姚先生，陈先生，许先生和张先生。通用风格转换的封闭式解决方案。在 IEEE/CVF 计算机视觉国际会议的进程中，页 5952-5961,2019。3

[20] M. Lu, H. Zhao, A. Yao, F. Xu, Y. Chen, and L. Zhang. De-coder network over lightweight reconstructed feature for fast semantic style transfer. In Proceedings of the IEEE Inter-national Conference on Computer Vision, pages 2469–2477, 2017. 1
陆、赵、姚、许、陈、张。基于轻量级重构特性的解码器网络语义风格快速转换。在 IEEE 国际计算机视觉会议记录中，页 2469-2477,2017。1

[21] K. Nichol. Painter by numbers – kaggle. https://www.kaggle.com/c/painter-by-numbers, 2016. 6
作者: 尼科尔。《画家》作者: numbers-kaggle 2016.6

[22] X. Pan, X. Zhan, J. Shi, X. Tang, and P. Luo. Switchable whitening for deep representation learning. In Proceedings of the IEEE/CVF International Conference on Computer Vi-sion, pages 1863–1871, 2019. 5
潘、詹、史、唐、罗。用于深度表征学习的可切换美白。在 IEEE/CVF 计算机视觉国际会议记录中，页 1863-1871,2019。5

[23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury,
A。帕斯克，s。格罗斯，f。马萨，a。勒勒，j,
G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32:8026–8037, 2019. 8
查南，t. 基林，z. 林，n. 吉梅尔辛，l. 安提加等。Pytorch: 一个命令式的、高性能的深度学习库。神经信息处理系统的进展，32:8026-8037,2019。8

[24] F. Shen, S. Yan, and G. Zeng. Neural style transfer via meta networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8061– 8069, 2018. 1
沈志华、颜思恩和曾庆红。通过元网络进行神经样式转换。在 IEEE 计算机视觉和模式识别会议记录中，页 8061-8069,2018。1

[25] L. Sheng, Z. Lin, J. Shao, and X. Wang. Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In Pro-ceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8242–8250, 2018. 3
L. Sheng，z. Lin，j. Shao，and x. Wang.阿凡达网: 多尺度零镜头风格的特征装饰转换。在 IEEE 计算机视觉和模式识别会议的议事录中，页 8242-8250,2018。3

[26] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. 3, 6
K. Simonyan 和 a. Zisserman. 用于大规模图像识别的非常深的卷积网络. arXiv 预印本 arXiv: 1409.1556,2014.3,6

[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed,
C. 赛格迪，刘永佳，P.Sermanet，S.Reed,
D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015. 3
安圭洛夫、 d. Erhan、 v. Vanhoucke 和 a. Rabinovich。更深入地研究卷积。在 IEEE 计算机视觉和模式识别会议记录中，第 1-9 页，2015 年。3

[28] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In Proceedings of the International Con-ference on Machine Learning, pages 1349–1357, 2016. 1, 2, 6
乌里诺夫、列别捷夫、维达尔迪和兰皮茨基。纹理网络: 纹理和风格化图像的前馈合成。机器学习国际会议记录，页 1349-1357,2016。1,2,6

[29] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normal-ization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022, 2016. 2
正常化: 快速程式化缺失的成分 arXiv 预印本 arXiv: 1607.08022,2016.2

[30] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6924–6932, 2017. 1, 3, 4
D. Ulyanov a. Vedaldi 和 v. Lempitsky。改进的纹理网络: 在前馈风格化和纹理合成中最大化质量和多样性。在 IEEE 计算机视觉和模式识别会议记录中，页 6924-6932,2017。1,3,4

[31] H. Wang, Y. Li, Y. Wang, H. Hu, and M.-H. Yang. Collabo-rative distillation for ultra-resolution universal style transfer.
Wang，y. Li，y. Wang，h. H.，and M.-H。超分辨率通用风格转换的合作精华。

In Proceedings of the IEEE/CVF Conference on Computer
IEEE/CVF 计算机会议论文集

Vision and Pattern Recognition, pages 1860–1869, 2020. 1,
视觉与模式识别，页 1860-1869,2020.1,

2, 3, 4, 5, 6, 8
2,3,4,5,6,8

[32] X. Wang, G. Oxholm, D. Zhang, and Y.-F. Wang. Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-tion, pages 5239–5247, 2017. 1, 2, 3

X. Wang，g. Oxholm，d. Zhang，and Y.-F.王永平。多模态传递：一种用于快速艺术风格传递的分层深度卷积神经网络。在 IEEE 计算机视觉和模式识别会议记录中，页 5239-5247,2017。1,2,3

[33] Y. Yao, J. Ren, X. Xie, W. Liu, Y.-J. Liu, and J. Wang. Attention-aware multi-stroke style transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pat-tern Recognition, pages 1467–1475, 2019. 3

姚永仁，谢华华，刘华华，刘永强。Liu，and J.Wang 刘、王建民。注意多笔划风格转换。在 IEEE/CVF 计算机视觉和模式识别会议进程中，页 1467-1475,2019。3

[34] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha. Photorealistic style transfer via wavelet transforms. In Proceedings of the IEEE/CVF International Conference on Computer Vi-sion, pages 9036–9045, 2019. 5

余永定、陈子昂、姜炳浩和余永定。哈。基于小波变换的 photo 风格转换。在 IEEE/CVF 计算机视觉国际会议记录中，页 9036-9045,2019。5

[35] H. Zhang and K. Dana. Multi-style generative network for real-time transfer. In Proceedings of the European Confer-ence on Computer Vision Workshops, 2018. 2, 3, 4, 5, 6, 8

张海伦和戴娜。用于实时传输的多种生成网络。在 2018 年欧洲计算机视觉研讨会会议记录中。2,3,4,5,6,8