Improved Texture Networks: Maximizing Quality and Diversity in Feed-forward Stylization and Texture Synthesis
改进的纹理网络: 最大化质量和多样性 前馈风格化与纹理合成

Dmitry Ulyanov
Dmitry Ulyanov
Skolkovo Institute of Science and Technology & Yandex
Skolkovo Institute of Science and Technology & Yandex
dmitry.ulyanov@skoltech.ru
Dmitry.ulyanov@skoltech.ru

Andrea Vedaldi
Andrea Vedaldi
University of Oxford
牛津大学
vedaldi@robots.ox.ac.uk
Vedaldi@robots.ox.ac.uk

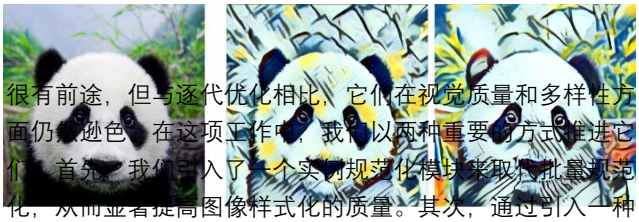
Victor Lempitsky
Victor Lempitsky 维克多·兰皮茨基
Skolkovo Institute of Science and Technology
斯科尔科沃科学技术研究所
lempitsky@skoltech.ru
Lempitsky@skoltech.ru

arXiv: 1701.02096 v2[cs.CV] 11 Jun 2017
日获发

Abstract
摘要

The recent work of Gatys et al., who characterized the style of an image by the statistics of convolutional neural network filters, ignited a renewed interest in the texture gen-eration and image stylization problems. While their image generation technique uses a slow optimization process, re-cently several authors have proposed to learn generator neural networks that can produce similar outputs in one quick forward pass. While generator networks are promis-ing, they are still inferior in visual quality and diversity compared to generation-by-optimization. In this work, we advance them in two significant ways. First, we introduce an instance normalization module to replace batch normal-ization with significant improvements to the quality of image stylization. Second, we improve diversity by introducing a new learning formulation that encourages generators to sample unbiasedly from the Julesz texture ensemble, which is the equivalence class of all images characterized by cer-tain filter responses. Together, these two improvements take feed forward texture synthesis and image stylization much closer to the quality of generation-via-optimization, while retaining the speed advantage.

Gatys 等人最近的工作，通过卷积神经网络滤波器的统计特征表征了图像的风格，重新点燃了人们对纹理生成和图像风格化问题的兴趣。虽然他们的图像生成技术使用一个缓慢的优化过程，最近几个作者已经提出学习发生器神经网络，可以在一个快速前进通道产生类似的输出。虽然生成器网络很有前途，但与逐代优化相比，它们在视觉质量和多样性方面仍然逊色。在这项工作中，我们以两种重要的方式推进它们。首先，我们引入了一个实例规范化模块来取代批量规范化，从而显著提高图像样式化的质量。其次，通过引入一种新的学习公式，鼓励生成器从 Julesz 纹理集合中无偏采样，从而提高图像的多样性。 Julesz 纹理集合是具有一定滤波响应的所有图像的等价类，具有一定的拥有属性。这两个改进使得前馈纹理合成和图像风格化更接近于通过优化生成的质量，同时保留了速度优势。



1. Introduction
引言

The recent work of Gatys et al. [4, 5], which used deep neural networks for texture synthesis and image stylization to a great effect, has created a surge of interest in this area. Following an earlier work by Portilla and Simoncelli [15], they generate an image by matching the second order mo-ments of the response of certain filters applied to a reference texture image. The innovation of Gatys et al. is to use non-

Gatys 等[4,5]最近的工作，使用深层神经网络进行纹理合成和图像风格化取得了很大的效果，引起了人们对这一领域的兴趣。继 Portilla 和 Simoncelli [15]的早期工作之后，他们通过匹

配用参纹图的些波响的阶来成像。Gatys等的新使非

应于考理像某滤器应二矩生图。人创是用

(a) Panda. (b) (c)
(a)熊猫(b)(c)

(d) Style. (e) (f)
(d)款式(e)(f)

Figure 1: Which panda stylization seems the best to you? Definitely not the variant (b), which has been produced by a state-of-the-art algorithm among methods that take no longer than a second. The (e) picture took several minutes to generate using an optimization process, but the quality is worth it, isn't it? We would be particularly happy if you chose one from the rightmost two examples, which are com-puted with our new method that aspires to combine the qual-ity of the optimization-based method and the speed of the fast one. Moreover, our method is able to produce diverse stylizations using a single network.

图 1: 你觉得哪种熊猫风格最好？绝对不是变体(b)，它是由最先进的算法在不超过一秒钟的方法中产生的。使用优化过程生成(e)图片需要几分钟，但是质量是值得的，不是吗？如果您从最右边的两个例子中选择一个，我们将非常高兴。这两个例子是用我们的新方法计算出来的，这种新方法旨在将基于优化的方法的优点和快速的方法的速度结合起来。此外，我们的方法能够使用单一的网络产生不同的风格化。

linear convolutional neural network filters for this purpose. Despite the excellent results, however, the matching pro-cess is based on local optimization, and generally requires a considerable amount of time (tens of seconds to minutes) in order to generate a single textures or stylized image.

线性卷积神经网络滤波器。尽管结果很好，但是匹配过程是基于局部优化的，通常需要相当长的时间(几十秒到几分钟)才能生成一个单一的纹理或风格化的图像。

In order to address this shortcoming, Ulyanov et al. [19] and Johnson et al. [8] suggested to replace the optimiza-tion process with feed-forward generative convolutional networks. In particular, [19] introduced texture networks

为了解决这个问题，Ulyanov 等[19]和 Johnson 等[8]提出用前馈生成卷积网络代替优化过程。特别是，[19]引入了纹理网络

The 源代码可以在 https://github.com/DmitryUlyanov/texture_nets DmitryUlyanov/纹理网 source code is available at https://github.com/

to generate textures of a certain kind, as in [4], or to apply a certain texture style to an arbitrary image, as in [5]. Once trained, such texture networks operate in a feed-forward manner, three orders of magnitude faster than the optimiza-tion methods of [4, 5].
生成特定类型的纹理，如[4]所示，或者对任意图像应用特定纹理样式，如[5]所示。一旦经过训练，这样的纹理网络以前馈方式运行，比[4,5]的优化方法快三个数量级。

The price to pay for such speed is a reduced performance. For texture synthesis, the neural network of [19] generates good-quality samples, but these are not as diverse as the ones obtained from the iterative optimization method of [4]. For image stylization, the feed-forward results of [19, 8] are qualitatively and quantitatively worse than iterative optimization. In this work, we address both limitations by means of two contributions, both of which extend beyond the applications considered in this paper.

这种速度的代价是性能降低。对于纹理合成，[19]的神经网络生成了高质量的样本，但是这些样本不像[4]的迭代优化方法所得到的样本那样多样化。对于图像风格化，[19,8]的前馈结果在定性和定量上比迭代优化更差。在这项工作中，我们通过两个贡献来解决这两个限制，这两个贡献都超出了本文所考虑的应用程序。

Our first contribution (section 4) is an architectural change that significantly improves the generator net-works. The change is the introduction of an instance-normalization layer which, particularly for the stylization problem, greatly improves the performance of the deep net-work generators. This advance significantly reduces the gap in stylization quality between the feed-forward models and the original iterative optimization method of Gatys et al., both quantitatively and qualitatively.

我们的第一个贡献(第 4 部分)是一个架构上的改变，它显著地改进了生成器网络。这种改变是引入了一个实例规范化层，特别是针对程式化问题，极大地提高了深层网络生成器的性能。这一进展大大缩小了前馈模型与 Gatys 等人的原始迭代优化方法在定量和定性上的差距。

Our second contribution (section 3) addresses the lim-ited diversity of the samples generated by texture networks. In order to do so, we introduce a new formulation that learns generators that uniformly sample the Julesz en-semble [20]. The latter is the equivalence class of im-ages that match certain filter statistics. Uniformly sampling this set guarantees diverse results, but traditionally doing so required slow Monte Carlo methods [20]; Portilla and Simoncelli, and hence Gatys et al., cannot sample from this set, but only find individual points in it, and possibly just one point. Our formulation minimizes the Kullback-Leibler divergence between the generated distribution and a quasi-uniform distribution on the Julesz ensemble. The learning objective decomposes into a loss term similar to Gatys et al. minus the entropy of the generated texture sam-ples, which we estimate in a differentiable manner using a non-parametric estimator [12].

我们的第二个贡献(第 3 部分)解决了由纹理网络生成的样本的有限多样性。为了做到这一点，我们引入了一个新的公式，学习统一采样 Julesz en-ensemble 的生成器[20]。后者是匹配某些过滤器统计的图像的等价类。统一采样这个集合保证了不同的结果，但传统上这样做需要慢蒙特卡罗方法[20]；Portilla 和 Simoncelli，因此 Gatys 等人，不能从这个集合采样，但只能找到单个点，可能只有一个点。我们的公式使生成的分布与 Julesz 集合上的准均匀分布之间的 Kullback-Leibler 分歧最小化。学习目标分解成一个类似于 Gatys 等人的损失项，减去生成纹理样本的熵，我们使用非参数估计器以可微的方式估计纹理样本[12]。

We validate our contributions by means of extensive quantitative and qualitative experiments, including com-paring the feed-forward results with the gold-standard optimization-based ones (section 5). We show that, com-bined, these ideas dramatically improve the quality of feed-forward texture synthesis and image stylization, bringing them to a level comparable to the optimization-based ap-proaches.

我们通过广泛的定量和定性实验验证了我们的贡献，包括将前馈结果与基于金标准的优化结果进行比较(第 5 节)。我们表明，结合这些思想，可以显著提高前馈纹理合成和图像风格化的质量，使它们达到与基于优化的方法相当的水平。

2. Background and related work
背景和相关工作

Julesz ensemble. Informally, a texture is a family of vi-sual patterns, such as checkerboards or slabs of concrete, that share certain local statistical regularities. The concept
Julesz 合唱团。非正式地说，纹理是一系列视觉模式，比如棋盘或混凝土板，它们共享某些局部统计规律。概念

was first studied by Julesz [9], who suggested that the visual system discriminates between different textures based on the average responses of certain image filters.
Julesz [9]首先对此进行了研究，他认为视觉系统根据某些图像滤波器的平均响应来区分不同纹理。

The work of [20] formalized Julesz' ideas by introducing the concept of Julesz ensemble. There, an image is a real function x : ! R^3 defined on a discrete lat-tice = f1; : : : ; Hg f1; : : : ; W g and a texture is a distribution p(x) over such images. The local statistics of an image are captured by a bank of (non-linear) filters Fl : X ! R, l = 1; : : : ; L, where Fl(x; u) denotes the response of filter Fl at location u on image x. The image x is characterized by the spatial average of the filter responses

通过引入 Julesz 集成的概念，[20]将 Julesz 的思想形式化。在那里，一个图像是一个真正的函数 x: ！R3 定义在一个离散格子上，纹理是这些图像上的分布 p (x)。图像的局部统计被一组(非线性)滤波器捕获 Fl: x！R，l = 1; : : ; l，其中 Fl (x; u)表示滤波器 Fl 在图像 x 上的位置 u 处的响应。图像 x 的特征是过滤器响应的空间平均值拥有属性

l(x) = u2 Fl(x; u)=j j. The image is perceived as a particular texture if these responses match certain characteristic values l. Formally, given the loss function,

L (x) = u2 Fl (x; u) = j j.如果这些响应符合特定的特征值，图像被认为是一个特定的纹理。正式地，给定损失函数，

$$L(x) = \sum_{l=1}^{L} (l(x) - l)^2 \qquad (1)$$

the Julesz ensemble is the set of all texture

$$\text{images } T = \{x \in X : L(x) = 0\}$$

Julesz 集合是所有纹理图像的集合 t = fx2 x: l

(x) g

that approximately satisfy such constraints. Since all tex-tures in the Julesz ensemble are perceptually equivalent, it is natural to require the texture distribution p(x) to be uniform over this set. In practice, it is more convenient to consider the exponential distribution
大致满足这些约束条件。由于 Julesz 集合中的所有纹理在感知上是等价的，因此很自然地要求纹理分布 p (x)在这个集合上是均匀的。实际上，考虑指数分布更方便

$$p(x) = \frac{e^{\frac{1}{T}\sum_{l=1}^{L}(\lambda_l - l(x))^2}}{\int_{R} e^{\frac{1}{T}\sum_{l=1}^{L}(\lambda_l - l(y))^2} dy} \qquad (2)$$

where T > 0 is a temperature parameter. This choice is mo-tivated as follows [20]: since statistics are computed from spatial averages of filter responses, one can show that, in the limit of infinitely large lattices, the distribution p(x) is zero outside the Julesz ensemble and uniform inside. In this manner, eq. (2) can be though as a uniform distribu-tion over images that have a certain characteristic filter re-sponses = (1; : : : ; L).
其中 t > 0 是温度参数。这种选择的动机如下[20]：由于统计是从滤波器响应的空间平均值计算出来的，人们可以证明，在无限大格子的极限下，分布 p (x)在 Julesz 系综外为零，在 Julesz 系综内为均匀。以这种方式，等式。(2)可以作为一个均匀分布的图像，具有一定的特征滤波响应 = (1; : : ; l)。

Note also that the texture is completely described by the filter bank F = (F_1; : : : ; F_L) and their characteristic re-sponses . As discussed below, the filter bank is generally fixed, so in this framework different textures are given by different characteristics .

还要注意，纹理完全由滤波器组 f = (F1; : : ; FL)及其特征响应来描述。正如下面讨论的，滤波器组通常是固定的，所以在这个框架中，不同的纹理由不同的特征给出。

Generation-by-minimization. For any interesting choice of the filter bank F , sampling from eq. (2) is rather challeng-ing and classically addressed by Monte Carlo methods [20]. In order to make this framework more practical, Portilla and Simoncelli [16] proposed instead to heuristically sam-ple from the Julesz ensemble by the optimization process
按最小化生成。对于滤波器组 f 的任何有趣的选择，从等式中取样。(2)是相当具有挑战性的，并且通过蒙特卡罗方法经典地解决[20]。为了使这个框架更加实用，Portilla 和 Simoncelli [16]提出用优化过程从 Julesz 集成中启发式采样

$$x = \operatorname{argmin}_{x \in X} L(x) : \qquad (3)$$

If this optimization problem can be solved, the minimizer x is by definition a texture image. However, there is no reason why this process should generate fair samples from the distribution p(x). In fact, the only reason why eq. (3) may not simply return always the same image is that the op-timization algorithm is randomly initialized, the loss func-tion is highly non-convex, and search is local. Only because of this eq. (3) may land on different samples x on different runs.

Deep filter banks. Constructing a Julesz ensemble requires choosing a filter bank F . Originally, researchers considered the obvious candidates: Gaussian derivative filters, Gabor filters, wavelets, histograms, and similar [20, 16, 21]. More recently, the work of Gatys et al. [4, 5] demonstrated that much superior filters are automatically learned by deep convolutional neural networks (CNNs) even when trained for apparently unrelated problems, such as image classification. In this paper, in particular, we choose for L(x) the style loss proposed by [4]. The latter is the distance between the empirical correlation matrices of deep filter responses in a CNN.[1]

Stylization. The texture generation method of Gatys et al. [4] can be considered as a direct extension of the texture generation-by-minimization technique (3) of Portilla and Simoncelli [16]. Later, Gatys et al. [5] demonstrated that the same technique can be used to generate an image that mixes the statistics of two other images, one used as a texture template and one used as a content template. Content is captured by introducing a second loss $L_{cont.}(x; x_0)$ that compares the responses of deep CNN filters extracted from the generated image x and a content image x_0. Minimizing the combined loss L(x) + $L_{cont.}(x; x_0)$ yields impressive artistic images, where a texture , defining the artistic style, is fused with the content image x_0.

Feed-forward generator networks. For all its simplicity and efficiency compared to Markov sampling techniques, generation-by-optimization (3) is still relatively slow, and certainly too slow for real-time applications. Therefore, in the past few months several authors [8, 19] have proposed to learn generator neural networks g(z) that can directly map random noise samples z p_z = N (0; I) to a local minimizer of eq. (3). Learning the neural network g amounts to minimizing the objective

$$g = \operatorname*{argmin}_{g} E_{p_z} L(g(z)):$$

[1]Note that such matrices are obtained by averaging local non-linear filters these are the outer products of filters in a certain layer of the neural network. Hence, the style loss of Gatys et al. is in the same form as eq. (1).

注意，这些矩阵是通过对局部非线性滤波器进行平均得到的；它们是神经网络某一层滤波器的外积。因此，Gatys 等人的样式损失与等式的形式相同。(1).

$$G = \operatorname*{argmin}_{G} E_{pz} l (g (z)) :$$

如果这个最佳化问题可以解决，那么最小值 x 定义为纹理图像。然而，这个过程没有理由从分布 p (x)中生成公平的样本。事实上，公式的唯一原因。(优化算法是随机初始化的，损失函数是高度非凸的，搜索是局部的，不可能总是返回相同的图像。仅仅是因为这个公式。(3)可以降落在不同的样本 x 在不同的运行。

深度过滤器组。构建 Julesz 集成需要选择滤波器组 f。最初，研究人员考虑了明显的候选者: 高斯导数滤波器，Gabor 滤波器，小波，直方图，和类似的[20,16,21]。最近，Gatys 等[4,5]的工作表明，深度卷积神经网络(cnn)能够自动学习更优秀的滤波器，即使在训练显然不相关的问题，如图像分类时也是如此。在本文中，特别是我们选择 l (x)提出的样式损失[4]。后者是 CNN 中深度滤波响应的经验相关矩阵之间的距离

程式化。Gatys 等[4]的纹理生成方法可以看作是 Portilla 和 Simoncelli [16]的最小化纹理生成技术(3)的直接延伸。后来，Gatys 等[5]证明了同样的技术可以用来生成一个混合了其他两个图像的统计信息的图像，一个用作纹理模板，另一个用作内容模板。通过引入第二个损失 Lcont 来捕获内容。(x; x0) ，比较从生成的图像 x 和内容图像 x0 提取的深度 CNN 滤波器的响应。最小化组合损失 l (x) + Lcont。(x; x0)产生令人印象深刻的艺术图像，其中定义艺术风格的纹理与内容图像融合。

前向发电机网络。尽管与马尔可夫采样技术相比，逐步优化(3)具有简单和高效的特点，但对于实时应用来说，它仍然相对缓慢，而且肯定太慢了。因此，在过去的几个月中，一些作者[8,19]已经提出学习发生器神经网络 g (z) ，可以直接映射随机噪声样本 z pz = n (0; i)到方程的局部最小值。(3). 学习神经网络 g 等于最小化目标

While this approach works well in practice, it shares the same important limitation as the original work of Portilla and Simoncelli: there is no guarantee that samples gener-ated by g would be fair samples of the texture distribu-tion (2). In practice, as we show in the paper, such samples tend in fact to be not diverse enough.

虽然这种方法在实践中运行良好，但它与 Portilla 和 Simoncelli 的原始工作具有同样重要的局限性: 不能保证由 g 生成的样本是纹理分布的公平样本(2)。在实践中，正如我们在论文中所展示的，这样的样本实际上往往不够多样化。

Both [8, 19] have also shown that similar generator net-works work also for stylization. In this case, the generator g(x_0; z) is a function of the content image x_0 and of the random noise z. The network g is learned to minimize the sum of texture loss and the content loss:

这两个[8,19]也表明，类似的发电机网络也工作程式化。在这种情况下，生成器 g (x0; z)是内容图像 x0 和随机噪声 z 的函数。网络 g 用于最小化纹理损失和内容损失的总和:

$$g = \arg\min_G [(g(x; z)) + \text{Lco}_{\text{nt.}} (g(x ; z); x)]: \quad (5)$$

$$g = \arg\min_G [(g (xZ)) + \text{继续}_{\circ} (g (x; z) ; x)] : \quad (5)$$

Alternative neural generator methods. There are many other techniques for image generation using deep neural networks.
使用深层神经网络生成图像的技术还有很多。

The Julesz distribution is closely related to the FRAME maximum entropy model of [21], as well as to the concept of Maximum Mean Discrepancy (MMD) introduced in [7]. Both FRAME and MMD make the observation that a prob-ability distribution p(x) can be described by the expected values $= E_{x\ p(x)}[(x)]$ of a sufficiently rich set of statistics (x). Building on these ideas, [14, 3] construct generator neural networks g with the goal of minimizing the discrepancy between the statistics averaged over a batch of

Julesz 分布与[21]的 FRAME 最大熵模型以及[7]中引入的最大均值差异(MMD)概念密切相关。FRAME 和 MMD 都观察到，概率能力分布 p (x)可以用一组足够丰富的统计量(x)的期望值 = Ex p (x)[(x)]来描述。在这些想法的基础上，[14,3]构造了发电机神经网络 g，其目标是最小化平均数据之间的差异

generated images
生成的图像

$$\frac{1}{N}\sum_{i=1}^{N} (g(zi))=N \text{ and the statistics av-}$$

$$\frac{1}{I}\sum_{i=1}^{I} (g (zi)) = n \text{ 和统计 av-}$$

training set
训练用其

$$\frac{1}{M}\sum_{i=1}^{M} (x_i) = m \text{ 结果}$$

eraged over a
发生在

network s
网络

are called
被称为

(MMN) .

(MMN) .

Moment Matching Networks

Moment Matching Networks 矩匹配网络

An important alternative methodology is based on the concept of Generative Adversarial Networks (GAN; [6]). This approach trains, together with the generator network g(z), a second adversarial network f( ) that attempts to dis-tinguish between generated samples g(z); z N (0; I) and real samples x $p_{data}(x)$. The adversarial model f can be used as a measure of quality of the generated samples and used to learn a better generator g. GAN are powerful but no-toriously difficult to train. A lot of research is has recently focused on improving GAN or extending it. For instance, LAPGAN [2] combines GAN with a

Laplacian pyramid and DCGAN [17] optimizes GAN for large datasets.

一个重要的替代方法是基于生成性对抗网络(GAN; [6])的概念。这种方法与生成器网络 g (z)一起训练第二个敌对网络 f ()，它试图区分生成的样本 g (z) ; z n (0; i)和实际样本 x pdata (x)。对抗模型 f 可以用来衡量生成的样本的质量，并用来学习一个更好的发电机。 GAN 是强大的，但并不是很难训练。最近很多研究都集中在改进 GAN 或者扩展 GAN 上。例如，LAPGAN [2]将 GAN 与 Laplacian 金字塔相结合，DCGAN [17]优化 GAN 用于大型数据集。

3. Julesz generator networks
3. Julesz 发生器网络

This section describes our first contribution, namely a method to learn networks that draw samples from the Julesz

本节描述我们的第一个贡献，即一种学习从 Julesz 中抽取样本的网络的方法

(4) ensemble modeling a texture (section 2), which is an intractable problem usually addressed by slow Monte Carlo methods [21, 20]. Generation-by-optimization, popularized by Portilla and Simoncelli and Gatys et al., is faster, but can only find one point in the ensemble, not sample from it,

集成建模纹理(第 2 节)，这是一个通常用慢速蒙特卡罗方法解决的难以处理的问题[21,20]。由 Portilla、 Simoncelli 和 Gatys 等人推广的通过优化生成算法速度更快，但是只能在集成中找到一个点，而不能从集成中取样,

其中 d = 3wh 是图像的分量 x 2 r3wh 的个数。
H (q) ND x
l = 1
距离 i 可以用来近似熵如下:
这类似于[11]的重参数化技巧，也用于[8,19]构造他们的学习目标。

第二项，负熵，更难准确估计，但存在简单的估计量。在我们的场景中特别吸引人的是 Kozachenko-Leonenko 估计量[12]。该估计器考虑一批 n 样本 x1; : : ; xn q (x)。然后，对于每个样本 xi，它计算到批中最近的邻居的距离 i:

(x) = t,

with scarce sample diversity, particularly when used to train feed-forward generator networks [8, 19].
样本多样性稀缺，特别是用于训练前馈发电机网络时[8,19]。

Here, we propose a new formulation that (x)=T , allows to train generator networks that sample the Julesz ensemble, gener-ating images with high visual fidelity as well as high diver-sity.

在这里，我们提出了一个新的公式，允许训练生成器网络的样本朱利叶斯集合，生成具有高视觉保真度和高多样性的图像。

A generator network [6] maps an i.i.d. noise vector z N (0; I) to an image x = g(z) in such a way that x is ideally a sample from the desired distribution p(x). Such generators have been adopted for texture synthesis in [19], but without guarantees that the learned generator g(z) would indeed sample a particular distribution.

一个发生器网络[6]将一个 i.i.i.d 噪声向量 z n (0; i)映射到一个图像 x = g (z)，使得 x 理想地是一个来自期望分布 p (x)的样本。这样的生成器已经在[19]中被用于纹理合成，但是没有保证学习生成器 g (z)确实会对特定的分布进行采样。

Here, we would like to sample from the Gibbs distri-bution (2) defined over the Julesz ensemble. This distri-

This is similar to the reparametrization trick of [11] and is also used in [8, 19] to construct their learning objectives.

The second term, the negative entropy, is harder to es-timate accurately, but simple estimators exist. One which is particularly appealing in our scenario is the Kozachenko-Leonenko estimator [12]. This estimator considers a batch of N samples x1; : : : ; xn q(x). Then for each sample xi, it computes the distance i to its nearest neighbour in the batch:

bution can be written compactly as p(x) = Z 1e

The distances i can be used to approximate the entropy as follows:

$$H(q) \quad N^{D} \sum_{i=1}^{N}$$

where D = 3W H is the number of components of the im-ages x 2 $R^{3 W H}$.

在这里，我们想从 Julesz 集合上定义的 Gibbs 分布(2)中取样。这个分布可以写成 p (x) = z 1e
where Z = e (x)=1 dx is an intractable
Z = 在哪 normalization
里 E (x) = tdx 是一个棘手的归一化问题 constant.
常数。 R

Denote by q(x) the distribution induced by a generator network g. The goal is to make the target distribution p and the generator distribution q as close as possible by minimiz-ing their Kullback-Leibler (KL) divergence:

用 q (x)表示由发电机网络 g 诱导的分布。目标是使目标分布 p 和生成器分布 q 尽可能接近，通过最小化它们的 Kullback-Leibler (KL)散度:

$$KL(q||p) = \int q(x) \ln \frac{q(x)}{p(x)} dx$$

KL (qjjp) = z q (x) ln qp (x) Dx

$$= \int q(x) \ln \frac{q(x) Z}{e^{-T(x)}} dx$$

$$= \int q(x) \ln q(x) + \ln(Z) + T(x) \, dx$$

$$= \mathbb{E}_{x \sim q(x)} \ln q(x) + \ln(Z) + \mathbb{E}_{x \sim q(x)} T(x)$$

$$= -H(q) + \text{const.} + \mathbb{E}_{x \sim q(x)} T(x)$$

=1 E (x) + E ln q(x) + ln(Z) (6)
T x q(x)
T x Q
T x (x) L X q (x)
1 (x) H(q) + const.
= E (x) H (q) + 常数。
T x q(x)
T x Q
T x (x) L

Hence, the KL divergence is the sum of the expected value of the style loss L and the negative entropy of the generated distribution q.
因此，KL 散度是风格损失 l 的期望值和生成的分布 q 的负熵的总和。

The first term can be estimated by taking the expectation over generated samples:
第一项可以通过对生成的样本采取期望值来估计:

$$\mathbb{E}_{x \sim q(x)} L(x) = \mathbb{E}_{z \sim N(0;I)} L(g(z)): \quad (7)$$

E l (x) = e l (g (z)) : (7)
x q(x) z N(0;I)
X q (x) z n (0; i)

An energy term similar to (6) was recently proposed in [10] for improving the diversity of a generator network in a adversarial learning scheme. While the idea is superfi-cially similar, the application (sampling the Julesz ensem-ble) and instantiation (the way the entropy term is imple-mented) are very different.

最近在[10]中提出了一个类似于(6)的能量术语，用于在对抗性学习方案中改善发电机网络的多样性。虽然这个想法在表面上是相似的，但是应用程序(对 Julesz 集成进行取样)和实例化(熵项的实现方式)是非常不同的。

Learning objective. We are now ready to define an objective function E(g) to learn the generator network g. This is given by substituting the expected loss (7) and the en-tropy estimator (9), computed over a batch of N generated images, in the KL divergence (6):

学习目标。我们现在准备定义一个目标函数 e (g)来学习生成器网络 g。这是通过替换期望损失(7)和熵估计(9)给出的，在一批 n 生成的图像上计算，在 KL 散度(6) :

E(g) = 1 N

E (g) = 1n h 1

ln j6=i k

Ln j6 = i ki (10)

k i j Ki (10)

g(z) g(z

min G) g(z

最小(z)) g (z)

The drawin

batch itself is obtained by g N samples

批次 本身 获得 画画 N 个样品

z1; : : : ;

zn N (0; I) from the noise distribution of the

Z1; : : ; 来自噪声分布的 n (0; i)

min

最

小 x x :

jk

开

玩

j=i

i = J = 笑

I = i k i 的 (8)
(8)

6

ln i + const.

— 在 i + 常数中。

(9)

zn

generator. The first term in eq. (10) measures how closely the generated images g(zi) are to the Julesz ensemble. The second term quantifies the lack of diversity in the batch by mutually comparing the generated images.

发电机。方程的第一项。(10)测量生成的图像 g (zi)与 Julesz 集合的距离。第二个术语通过相互比较生成的图像来量化批处理中多样性的缺乏。

Learning. The loss function (10) is in a form that allows optimization by means of Stochastic Gradient Descent (SGD). The algorithm samples a batch z1; : : : ; zn at a time and then descends the gradient:

学习。损失函数(10)的形式允许通过随机梯度下降(SGD)进行优化。该算法一次采样一批 z1; : : ; zn，然后下降梯度:

N

1 Xh dL dg(zi)

1 Xh dL dg (zi)

N dx> d >

Dx > d >

1=

1

I =

1

dg(z dg(zj

Dg Dg

(z) (zj)

(g(zi)

(g g(zji))> d > d > (11)

i (zi)) G (zji) > D > D > i (11)

where is the vector of parameters of the neural network g, the tensor image x has been implicitly vectorized and ji is the index of the nearest neighbour of image i in the batch.

其中 g 的参数向量是神经网络 g 的参数向量，张量图像 x 已经隐式向量化，ji 是图像 i 在批处理中最近邻的指标。

4. Stylization with instance normalization

实例规范化的程式化

The work of [19] showed that it is possible to learn high-quality texture networks g(z) that generate images in a Julesz ensemble. They also showed that it is possible to

[19]的工作表明，有可能学习高质量的纹理网络 g (z) ，生成图像的 Julesz 集合。他们还表明，可以

learn good quality stylization networks g(x0; z) that apply the style of a fixed texture to an arbitrary content image x0.

学习高质量的样式化网络 g (x0; z) ，将固定纹理的样式应用于任意内容图像。

Nevertheless, the stylization problem was found to be harder than the texture generation one. For the stylization

尽管如此，风格化的问题被发现比纹理生成的问题更难

task, they found that learning the model from too many example content images x0, say more than 16, yielded poorer

qualitative results than using a smaller number of such ex-amples. Some of the most significant errors appeared along

他们发现，从太多例子中学习模型的内容图像 x0，比如说超过 16，比使用较少数量的例子得到的定性结果更差。一些最显著的错误出现了

the border of the generated images, probably due to padding and other boundary effects in the generator network. We conjectured that these are symptoms of a learning problem too difficult for their choice of neural network architecture.

生成图像的边界，可能是由于生成器网络中的填充和其他边界效应。我们推测，这些是学习问题的症状，对于他们选择神经网络结构来说太难了。

A simple observation that may make learning simpler is that the result of stylization should not, in general, depend on the contrast of the content image but rather should match the contrast of the texture that is being applied to it. Thus, the generator network should discard contrast information in the content image x_0. We argue that learning to discard contrast information by using standard CNN building block is unnecessarily difficult, and is best done by adding a suit-able layer to the architecture.

一个简单的观察可能会使学习变得更简单，那就是程式化的结果通常不应该取决于内容图像的对比度，而是应该与应用到内容图像的纹理的对比度相匹配。因此，生成器网络应该丢弃内容图像中的对比度信息 x0。我们认为，通过使用标准的 CNN 构建块来学习丢弃对比度信息是不必要的困难，最好的方法是在体系结构中添加一个适合的层。

To see why, let x 2 $R^{N\ C\ W\ H}$ be an input tensor containing a batch of N images. Let x_{nijk} denote its nijk-th element, where k and j span spatial dimensions, i is the feature channel (i.e. the color channel if the tensor is an RGB image), and n is the index of the image in the batch. Then, contrast normalization is given by:

为了明白为什么，让 x2 RN c w h 是包含一批 n 图像的输入张量。让 xnijk 表示它的 nijk-th 元素，其中 k 和 j 跨空间维数，i 是特征通道(即如果张量是 RGB 图像，则为颜色通道)，n 是批图像的索引。然后，对比度归一化是通过:

$$y_{nijk} = \frac{x_{nijk}}{\sqrt{\epsilon + \sigma_{ni}^2}}; \quad \sigma_{ni}^2 = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H}(x_{nilm} - \mu_{ni})^2; \quad \mu_{ni} = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H} x_{nilm} \tag{12}$$

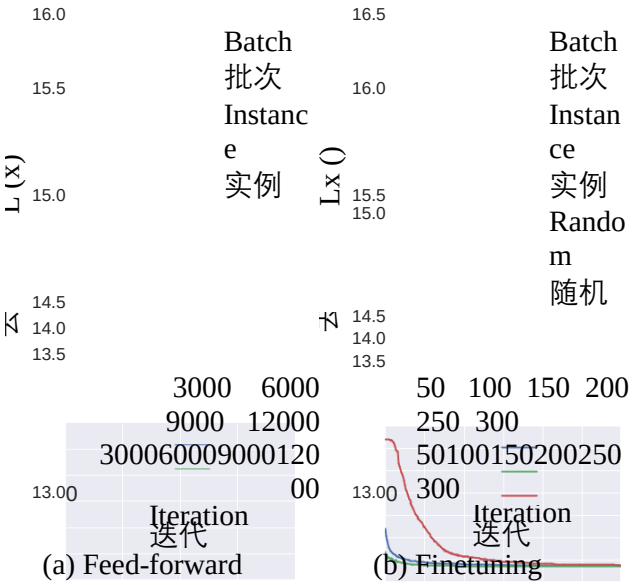
Ynijk = 尼 ni2 Ni 2 = Xnijk 尼 p 2 + ; 1 P niW H 1 niW h 尼 = HW =1 m=1 Xnilm; (12) 1 W H ni2 Ni 2 = HW Xl Xl X (xnilm ni)2: =1 m=1 Xnilm

It is unclear how such as function could be implemented as

目前还不清楚如何将这样的功能实现为

Figure 2: Comparison of normalization techniques in image stylization. From left to right: BN, cross-channel LRN at the first layer, IN at the first layer, IN throughout.

图 2: 图像风格化中标准化技术的比较。从左到右: BN，第一层的跨通道 LRN，第一层的 IN，整个 IN。



(a) Feed-forward history.

(a)前馈历史记录。

(b) Finetuning history.

(b)微调历史。

(c) Content.

(c)内容。

(d) StyleNet IN.

(d) StyleNet IN.

(e) IN finetuned.

(e)调节良好。

a sequence of standard operators such as ReLU and convo-
一系列标准操作符，如 ReLU 和 convo -
lution.
溶液。

On the other hand, the generator network of [19] does
另一方面，[19]的生成器网络则是如此
contain a normalization layers, and precisely batch normal-
包含一个正常化层，并精确批量正常-
ization (BN) ones. The key difference between eq. (12) and
公式(12)和
batch normalization is that the latter applies the normaliza-
批量标准化是指后者应用标准化-

tion to a whole batch of images instead of single ones:
对整批图像进行处理，而不是单个图像:

$$y_{nijk} = \frac{x_{nijk} - \mu_{ij}}{\sqrt{\sigma_{ij}^2 + \epsilon}}; \quad \mu_{ij} = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H} x_{nilm}; \quad \sigma_{ij}^2 = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H}(x_{nilm} - \mu_{ij})^2 \tag{13}$$

$$Y_{nijk} = \frac{X_{nijk} - \mu_{ij}}{\sqrt{\sigma_{ij}^2 + \epsilon}}; \quad \mu_{ij} = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H} X_{nilm}; \quad \sigma_{ij}^2 = \frac{1}{HW}\sum_{l=1}^{W}\sum_{m=1}^{H}(X_{nilm} - \mu_{ij})^2 \tag{13}$$

We argue that, for the purpose of stylization, the normal-
我们认为，为了程式化的目的，正常的 -
ization operator of eq. (12) is preferable as it can normalize

(f) Style. (g) StyleNet BN. (h) BN finetuned.
(f)款式。 (g) StyleNet BN. (h)对 BN 进行微调。

Figure 3: (a) learning objective as a function of SGD iter-
图 3: (a)学习目标作为 SGD iter-的函数
ations for StyleNet IN and BN. (b) Direct optimization of
(b)对 StyleNet IN 和 BN 的直接优化
the Gatys et al. for this example image starting from the re-
Gatys 等人为这个例子图像从重新开始
sult of StyleNet IN and BN. (d,g) Result of StyleNet with
StyleNet IN 和 BN 的结果(d，g)
instance (d) and batch normalization (g). (e,h) Result of
实例(d)和批量标准化(g)。(e，h)
finetuing the Gatys et al. energy.
细化 Gatys 等人的能量。

batch as a whole. Note in particular that this means that in-
批处理作为一个整体。特别要注意的是，这意味着
stance normalization is applied throughout the architecture,
姿态标准化应用于整个架构,
not just at the input image—fig. 2 shows the benefit of doing
不只是在输入图像-图 2 显示的好处做
so.
所以。

Another similarity with BN is that each IN layer is fol-
与 BN 的另一个相似之处在于，每一个 IN 层都是 fol-
lowed by a scaling and bias operator s x + b. A difference
由一个标度和偏差算子 s x + b 减小
is that the IN layer is applied at test time as well, unchanged,
在测试时也应用 IN 层，不变,
whereas BN is usually switched to use accumulated mean
而 BN 通常改用累计平均数
and variance instead of computing them over the batch.
和方差，而不是在批处理中计算它们。

IN appears to be similar to the layer normalization
IN 似乎与层标准化类似

方程(12)的归一化算符较好，因为它可以归一化

each individual content image x0.
每个单独的内容图像 x0。

While some authors call layer eq. (12) contrast normal-
虽然有些作者称层方程(12)对比正常-
ization, here we refer to it as instance normalization (IN)
这里我们称之为实例规范化(IN)

since we use it as a drop-in replacement for batch nor-
因为我们用它来代替批次或
malization operating on individual instances instead of the
在单个实例上操作的非正常化，而不是

method introduced in [1] for recurrent networks, although
[1]中引入的方法用于递归网络，尽管
it is not clear how they handle spatial data. Like theirs, IN
不清楚他们如何处理空间数据。像他们一样，IN
is a generic layer, so we tested it in classification problems
是一个通用层，所以我们在分类问题中测试了它
as well. In such cases, it still work surprisingly well, but not
在这种情况下，它仍然令人惊讶地工作良好，但不是
as well as batch normalization (e.g. AlexNet [13] IN has 2-
以及批量标准化(例如 AlexNet [13] IN 有 2-

3% worse top-1 accuracy on ILSVRC [18] than AlexNet BN).
ILSVRC [18]的最高精度比 AlexNet BN 差 3% 。

5. Experiments
5. 实验

In this section, after discussing the technical details of the method, we evaluate our new texture network architec-tures using instance normalization, and then investigate the ability of the new formulation to learn diverse generators.

在本节中，在讨论了该方法的技术细节之后，我们使用实例归一化来评估新的纹理网络结构，然后研究新公式学习不同生成器的能力。

5.1. Technical details
5.1 技术细节

Network architecture. Among two generator network architectures, proposed previously in [19, 8], we choose the residual architecture from [8] for all our style transfer exper-iments. We also experimented with architecture from [19] and observed a similar improvement with our method, but use the one from [8] for convenience. We call it StyleNet with a postfix BN if it is equipped with batch normalization or IN for instance normalization.
网络架构。在前面[19,8]中提出的两种发电机网络体系结构中，我们从[8]中选择剩余的体系结构进行所有的风格转换实验。我们还试验了来自[19]的架构，并观察到我们的方法有类似的改进，但为了方便使用[8]的架构。如果它配备了批量标准化或 IN 例如标准化，我们称之为带有后缀 BN 的 StyleNet。

For texture synthesis we compare two architectures: the multi-scale fully-convolutional architecture from [19] (Tex-tureNetV1) and the one we design to have a very large re-ceptive field (TextureNetV2). TextureNetV2 takes a noise vector of size 256 and first transforms it with two fully-connected layers. The output is then reshaped to a 4 4 image and repeatedly upsampled with fractionally-strided convolutions similar to [17]. More details can be found in the supplementary material.

对于纹理合成，我们比较了两种体系结构：来自[19]的多尺度全卷积体系结构(Tex-tureNetV1)和我们设计的具有非常大的接收场的体系结构(TextureNetV2)。TextureNetV2 获取大小为 256 的噪声矢量，并首先用两个完全连接的层转换它。然后将输出重塑为 44 图像，并以类似于[17]的分数步进卷积重复上样。更多细节可以在补充材料中找到。

Weight parameters. In practice, for the case of > 0, entropy loss and texture loss in eq. (10) should be weighted properly. As only the value of T is important for opti-mization we assume = 1 and choose T from the set of three values (5; 10; 20) for texture synthesis (we pick the higher value among those not leading to artifacts – see our discussion below). We fix T = 10000 for style transfer ex-periments. For texture synthesis, similarly to [19], we found useful to normalize gradient of the texture loss as it passes back through the VGG-19 network. This allows rapid con-vergence for stochastic optimization but implicitly alters the objective function and requires temperature to be adjusted. We observe that for textures with flat lightning high entropy weight results in brightness variations over the image fig. 7. We hypothesize this issue can be solved if either more clever distance for entropy estimation is used or an image prior in-troduced.
重量参数。在实践中，对于 > 0 的情况，熵损失和纹理损失在等式中。(10)应该适当加权。由于只有 t 值对于优化是重要的，我们假设为 = 1，并从三个值(5; 10; 20)中选择 t 值进行纹理合成(我们从那些不会产生伪影的值中选择较高的值 -- 参见我们下面的讨论)。我们将 t = 10000 固定为风格转换实验。对于纹理合成，类似于[19] ，我们发现当纹理损失通过 VGG-19 网络返回时，对纹理损失的梯度进行标准化是有用的。这允许随机优化的快速收敛，但隐含地改变了目标函数，并需要调整温度。我们观察到，对于具有扁平闪电的纹理，高熵权导致图像亮度变化图 7。我们假设这个问题可以解决，如果使用更聪明的距离熵估计或一个图像之前介绍。

5.2. Effect of instance normalization
5.2 实例归一化的效果

In order to evaluate the impact of replacing batch nor-malization with instance normalization, we consider first the problem of stylization, where the goal is to learn a gen-erator x = g(x0; z) that applies a certain texture style to the content image x0 using noise z as “random seed”. We

为了评价用实例归一化替代批量归一化的效果，我们首先考虑了样式化问题，其目标是学习一个生成器 x = g (x0; z) ，该生成器使用噪声 z 作为“随机种子”对内容图像 x0 应用某种纹理样式。我们

set = 0 for which generator is most likely to discard the noise.
Set = 0，哪个发电机最有可能丢弃噪声。

The StyleNet IN and StyleNet BN are compared in fig. 3. Panel fig. 3.a shows the training objective (5) of the networks as a function of the SGD training iteration. The objective function is the same, but StyleNet IN converges much faster, suggesting that it can solve the stylization problem more easily. This is confirmed by the stark difference in the qualitative results in panels (d) end (g). Since the StyleNets are trained to minimize in one shot the same objective as the iterative optimization of Gatys et al., they can be used to initialize the latter algorithm. Panel (b) shows the result of applying the Gatys et al. optimization starting from their random initialization and the output of the two StyleNets. Clearly both networks start much closer to an optimum than random noise, and IN closer than BN. The difference is qualitatively large: panels (e) and (h) show the change in the StyleNets output after finetuning by iterative optimization of the loss, which has a small effect for the IN variant, and a much larger one for the BN one.

图 3 比较了 StyleNet IN 和 StyleNet BN。面板图 3。A 显示了作为 SGD 培训迭代函数的网络培训目标(5)。目标函数是相同的，但 StyleNet IN 收敛得更快，这表明它可以更容易地解决程式化问题。这通过面板(d)末端(g)中定性结果的明显差异得到证实。由于 StyleNets 被训练成在一个镜头中最小化与 Gatys 等人的迭代优化相同的目标，因此它们可以用来初始化后一种算法。面板(b)显示了从随机初始化和两个 StyleNets 的输出开始应用 Gatys 等优化的结果。很明显，两个网络都比随机噪声更接近最优值，而 IN 比 BN 更接近最优值。差异是定性的大：面板(e)和(h)显示的变化，在 StyleNets 输出后，通过迭代优化的损失，有一个小的影响 IN 变体，一个更大的影响 BN 变体。

Similar results apply in general. Other examples are shown in fig. 4, where the IN variant is far superior to BN and much closer to the results obtained by the much slower iterative method of Gatys et al. StyleNets are trained on im-ages of a fixed sized, but since they are convolutional, they can be applied to arbitrary sizes. In the figure, the top tree images are processed at 512 512 resolution and the bot-tom two at 1024 1024. In general, we found that higher resolution images yield visually better stylization results.

类似的结果适用于一般情况。其他例子如图 4 所示，其中 IN 变体远远优于 BN，并且更接近 Gatys 等人用慢得多的迭代法得到的结果。样式表是在固定大小的图像上训练的，但是由于它们是卷积的，所以它们可以应用于任意大小。在图中，顶部的树木图像以 512512 分辨率处理，底部的图像以 10241024 分辨率处理。一般来说，我们发现更高分辨率的图像产生更好的视觉风格化结果。

While instance normalization works much better than batch normalization for stylization, for texture synthesis the two normalization methods perform equally well. This is consistent with our intuition that IN helps in normalizing the information coming from content image x_0, which is highly variable, whereas it is not important to normalize the texture information, as each model learns only one texture style.

虽然实例规范化在风格化方面比批量规范化要好得多，但在纹理合成方面，两种规范化方法的效果是一样的。这与我们的直觉是一致的，即 IN 有助于规范化来自内容图像 x0 的信息，这是高度可变的，而规范化纹理信息并不重要，因为每个模型只学习一种纹理样式。

5.3. Effect of the diversity term
5.3 多样性项的影响

Having validated the IN-based architecture, we evaluate now the effect of the entropy-based diversity term in the objective function (10).

在验证了基于信息网络的体系结构之后，我们现在评估了基于熵的多样性项在目标函数(10)中的作用。

The experiment in fig. 5 starts by considering the problem of texture generation. We compare the new high-capacity TextureNetV2 and the low-capacity TextureNetsV1 texture synthesis networks. The low-capacity model is the same as [19]. This network was used there in order to force the network to learn a non-trivial dependency on the input noise, thus generating diverse outputs even though the learning objective of [19], which is the same as eq. (10) with diversity coefficient = 0, tends to suppress diversity. The results in fig. 5 are indeed diverse, but sometimes of low quality. This should be contrasted

图 5 中的实验从考虑纹理生成问题开始。我们比较了新的高容量 TextureNetV2 和低容量 Tex-tureNetsV1 纹理合成网络。低容量模型与[19]相同。这个网络是用来迫使网络学习一个非平凡的依赖于输入噪声，从而产生不同的输出，即使学习目标[19]，这是相同的方程。(10)多样性系数 = 0，倾向于抑制多样性。图 5 中的结果确实是多样的，但有时质量低。这应该进行对比

Figure 4: Stylization results obtained by applying different textures (rightmost column) to different content images (leftmost column). Three methods are compared: StyleNet IN, StyleNet BN, and iterative optimization.
图 4: 通过将不同的纹理(最右边的列)应用到不同的内容图像(最左边的列)而获得的风格化结果比较了三种方法: StyleNet IN、 StyleNet BN 和迭代优化。

with TextureNetV2, the high-capacity model: its visual fi-delity is much higher, but, by using the same objective func-tion [19], the network learns to generate a single image, as expected. TextureNetV2 with the new diversity-inducing objective (> 0) is the best of both worlds, being both high-quality and diverse.
使用 TextureNetV2，大容量模型：它的视觉保真度要高得多，但是，通过使用相同的目标函数[19] ，网络学习生成一个单一的图像，如预期的那样。具有新的多样性诱导目标(> 0)的 TextureNetV2 是两个世界中最好的，既高质量又多样化。

The experiment in fig. 6 assesses the effect of the diversity term in the stylization problem. The results are similar to the ones for texture synthesis and the diversity term effec-tively encourages the network to learn to produce different results based on the input noise.
图 6 中的实验评估了多样性项在程式化问题中的作用。结果与纹理合成的结果相似，多样性项有效地鼓励网络学习基于输入噪声产生不同的结果。

One difficultly with texture and stylization networks is
纹理和风格化网络的一个难点是

that the entropy loss weight must be tuned for each learned texture model. Choosing too small may fail to learn a diverse generator, and setting it too high may create artifacts, as shown in fig. 7.
必须调整每个学习纹理模型的熵损失权重。选择太小可能无法学习多样化的生成器，设置太高可能产生伪影，如图 7 所示。

6. Summary
总结

This paper advances feed-forward texture synthesis and stylization networks in two significant ways. It introduces instance normalization, an architectural change that makes training stylization networks easier and allows the training process to achieve much lower loss levels. It also introduces a new learning formulation for training generator networks

本文从两个重要方面提出了前馈纹理合成和风格化网络。它引入了实例规范化，一个架构上的变化，使训练样式化网络更容易，并允许训练过程达到更低的损失水平。它还引入了一种新的训练发电机网络的学习公式

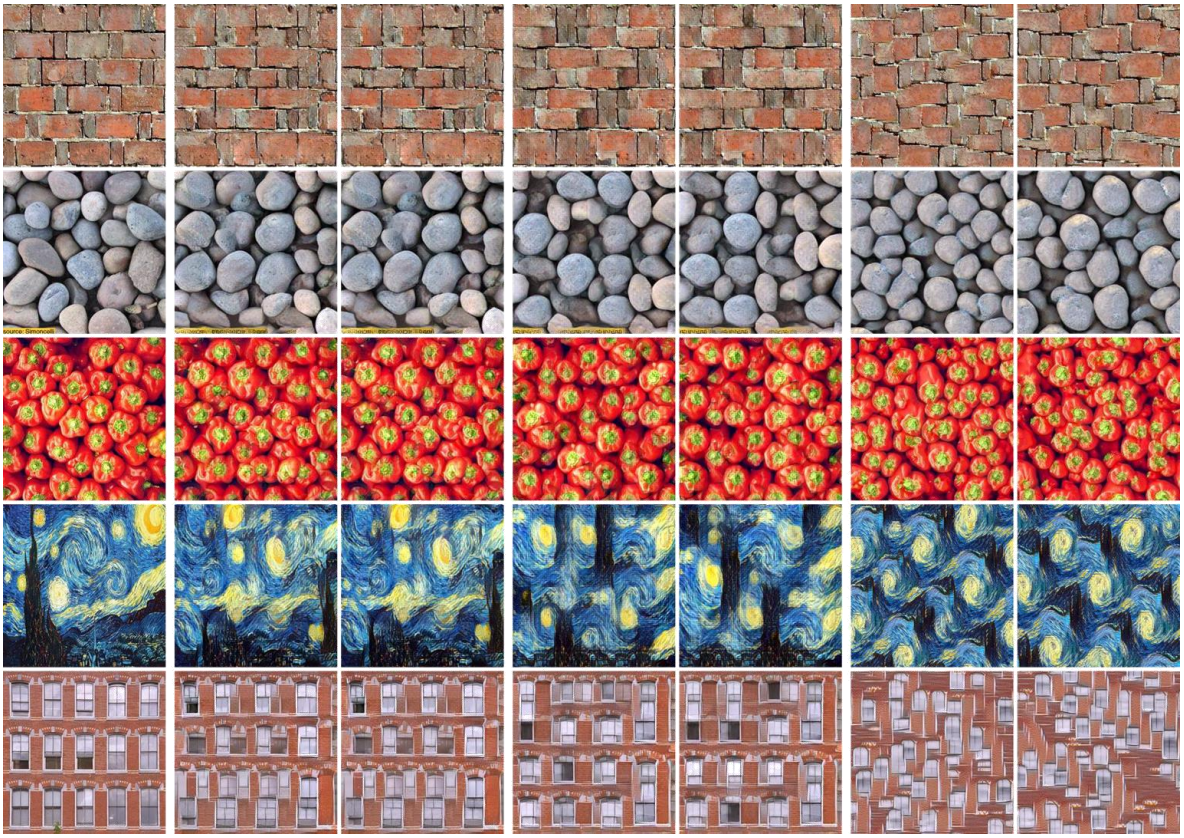


Figure 5: The textures generated by the high capacity Texture Net V2 without diversity term (= 0 in eq. (10)) are nearly identical. The low capacity TextureNet V1 of [19] achieves diversity, but has sometimes poor results. TextureNet V2 with diversity is the best of both worlds.

图 5: 高容量 Texture Net V2 生成的纹理没有多样性项(= 0，单位为 eq)。(10))几乎相同。[19]的低容量 TextureNet V1 实现了多样性，但有时效果不佳。具有多样性的 TextureNet V2 是两个世界中最好的。

Figure 6: The StyleNetV2 $g(x_0; z)$, trained with diversity > 0, generates substantially different stylizations for different values of the input noise z. In this case, the lack of stylization diversity is visible in uniform regions such as the sky.

图 6: stylenetv2g (x0; z)，经过多样性 > 0 的训练，对于输入噪声 z 的不同值产生了实质上不同的风格化。在这种情况下，在天空等统一区域可以看到缺乏样式化多样性。

to sample uniformly from the Julesz ensemble, thus explicitly encouraging diversity in the generated outputs. We show that both improvements lead to noticeable improve-ments of the generated stylized images and textures, while keeping the generation run-times intact.

从 Julesz 系列中统一采样，从而明确地鼓励生成输出的多样性。我们表明，这两个改进导致显着的改善生成的风格化图像和纹理，同时保持生成运行时间完整。

References
参考文献

[1] L. J. Ba, R. Kiros, and G. E. Hinton. Layer normalization. CoRR, abs/1607.06450, 2016. 5

L.j。 Ba，r。 Kiros，和 g。 Hinton。层规范化。 CoRR，abs/1607.06450,2016.5

[2] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. In NIPS, pages 1486–1494, 2015. 3

E. l. Denton，s. Chintala，a. Szlam，and r. Fergus.深度生成图像模型使用一个拉普拉斯金字塔的不利萨利亚网络。在 NIPS，页 1486-1494,2015。3

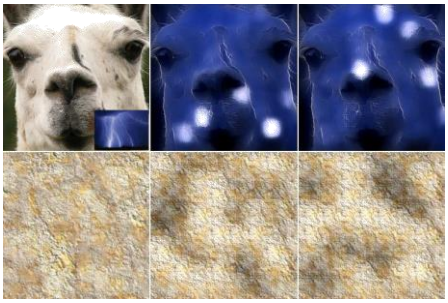


Figure 7: Negative examples. If the diversity term is too high for the learned style, the generator tends to generate artifacts in which brightness is changed locally (spotting) instead of (or as well as) changing the structure.

图 7：负面例子。如果多样性项对于学习风格来说太高，那么生成器倾向于生成一些工件，其中亮度局部改变(斑点)，而不是(或者)改变结构。

[3] G. K. Dziugaite, D. M. Roy, and Z. Ghahramani. Training generative neural networks via maximum mean discrepancy optimization. In UAI, pages 258–267. AUAI Press, 2015. 3

G. k. Dziugaite d. m. Roy 和 z. Ghahramani。通过最大均值差异优化训练生成神经网络。在 UAI 中，页 258-267。AUAI 出版社，2015。3

[4] L. Gatys, A. S. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In Advances in Neu-ral Information Processing Systems, NIPS, pages 262–270, 2015. 1, 2, 3

盖蒂斯，a · s · 埃克尔和 m · 贝格。使用卷积神经网络的纹理合成。In Advances In Neu-ral Information Processing Systems，NIPS，page 262-270,2015 神经信息处理系统进展，NIPS，262-270 页，2015 年。1,2,3

[5] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. CoRR, abs/1508.06576, 2015. 1, 2, 3

洛杉矶盖蒂斯 A.s. 埃克尔和 m. 贝格。艺术风格的神经算法。CoRR，abs/1508.06576,2015.1,2,3

[6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, I · j · 古德菲勒，j · 普吉-阿巴迪，m · 米尔扎，b · 许，D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Informa-tion Processing Systems (NIPS), pages 2672–2680, 2014. 3, 4

Warde-Farley，S.Ozair，A.c. Courville，and y. Bengio.生成性对抗网。神经信息处理系统进展(NIPS)，页 2672-2680,2014。3,4

[7] A. Gretton, K. M. Borgwardt, M. Rasch, B. Scholkopf,¨ and 格雷顿，K.m。博格沃特，m。拉希，b A. J. Smola. A kernel method for the two-sample-problem. In Advances in neural information processing systems,NIPS, pages 513–520, 2006. 3

J.Smola.两个样本问题的核方法。在神经信息处理系统的进步，NIPS，页 513-520,2006。3

[8] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II, pages 694–711, 2016. 1, 2, 3, 4, 6

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。2016 年 10 月 11 日至 14 日，在荷兰阿姆斯特丹召开的第 14 届欧洲会议，会议录，第二部分，页 694-711,2016。1,2,3,4,6

[9] B. Julesz. Textons, the elements of texture perception, and their interactions. Nature, 290(5802):91–97, 1981. 2

纹理知觉的要素及其相互作用。自然，290(5802) : 91-97,1981.2

[10] T. Kim and Y. Bengio. Deep directed generative models with energy-based probability estimation. arXiv preprint arXiv:1606.03439, 2016. 4

T. Kim 和 y. Bengio. Deep 定向生成模型与基于能量的概率估计. arXiv 预印 arXiv: 1606.03439,2016.4

[11] D. P. Kingma and M. Welling. Auto-encoding variational bayes. CoRR, abs/1312.6114, 2013. 4

《自动编码变分贝叶斯》，CoRR，abs/1312.6114,2013.4

[12] L. F. Kozachenko and N. N. Leonenko. Sample estimate of the entropy of a random vector. Probl. Inf. Transm., 23(1-2):95–101, 1987. 2, 4

L. f. Kozachenko 和 n. n. Leonenko。随机向量熵的样本估计。问题。参考译文。传送，23(1-2)：95-101,1987。2,4

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1106–1114, 2012. 5

A. Krizhevsky，i. Sutskever 和 G.e. Hinton。用深度卷积神经网络进行图像网分类。在 NIPS，页 1106-1114,2012。5

[14] Y. Li, K. Swersky, and R. S. Zemel. Generative moment matching networks. In Proc. International Conference on Machine Learning, ICML, pages 1718–1727, 2015. 3

Y · 李、k · 斯沃斯基和 r · s · 泽梅尔。生成矩匹配网络。在 Proc 中。国际机器学习会议，ICML，页 1718-1727,2015。3

[15] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. IJCV, 40(1):49–70, 2000. 1

J. Portilla 和 E.p. Simoncelli。基于复小波系数联合统计的参数纹理模型。IJCV，40(1) : 49-70,2000.1

[16] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. IJCV, 2000. 2, 3

J · 波蒂亚和 e · p · 西蒙切利。基于复小波系数联合统计的参数纹理模型。IJCV，2000.2,3

[17] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, abs/1511.06434, 2015. 3, 6

A. Radford，l. Metz 和 S.Chintala。无监督代表表象学习与深层卷积生成对抗网络。CoRR，abs/1511.06434,2015.3,6

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015. 5

O. Russakovsky，J.Deng，h. S.，J.Krause，s. Satheesh，S.Ma，z. Huang，a. Karpathy，a. Khosla，m. Bernstein，A.c. Berg，and l. Fei-Fei.ImageNet 大规模视觉识别挑战。国际计算机视觉杂志(IJCV)，115(3) : 211-252,2015。5

[19] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, pages 1349–1357, 2016. 1, 2, 3, 4, 5, 6, 8

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。第 33 届国际机器学习会议记录，ICML 2016，纽约，美国，2016 年 6 月 19-24 日，页 1349-1357,2016。1,2,3,4,5,6,8

[20] S. C. Zhu, X. W. Liu, and Y. N. Wu. Exploring texture ensembles by efficient markov chain monte carlotoward a atrichromacyo theory of texture. PAMI, 2000. 2, 3

朱克强，刘克华，吴恩恩。利用有效的马尔可夫链蒙特卡罗探索纹理集合，构建一个关于纹理的 atrichromacyo 理论。帕米，2000。2,3

[21] S. C. Zhu, Y. Wu, and D. Mumford. Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling. IJCV, 27(2), 1998. 3

朱咏武和 d. Mumford。过滤器，随机场和最大熵(FRAME)：走向纹理建模的统一理论。IJCV，27(2)，1998.3