

Texture Networks Feed forward Synthesis of Textures and Stylized Images (纹理和样式化图像的前馈合成)

引言

尽管本文发行时间神经网络已经可以从单一纹理示例生成纹理和风格化图像，但现有的方法仍需要占用大量内存而且极为缓慢，因此作者提出一种方法，增大学习阶段的负担，减小使用时的损耗，训练紧凑的前馈卷积网络生成相同纹理的多个样本，并且可以将艺术风格从一个给定图像转移到任何其他图像。

文章发行时间段（2016）图像生成有以下几种方法：

1. 使用前馈计算从随机种子产生图像输出
2. 描述性地使用网络作为图像统计学——将图像生成简化为从匹配某些统计数据的图像集中随机抽样的问题（效果更好，但生成图像需要时间长）
 1. 纹理合成中，参考统计数据是从单个视觉纹理示例中提取的，目标是生成更多该纹理的示例。
 2. 风格转移中，目标是同时匹配第一幅图像的视觉风格和第二幅图像的视觉内容，这些都是使用一些低级统计数据捕获的。

于是作者研究了如何使用前馈生成网络实现描述性网络（如CNN）的合成和风格化能力，提出了一种生成方法，其速度快两个数量级，内存效率高一个数量级；设计了一种特别适合我们考虑的任务的新型多尺度生成架构。

背景

图像生成

图像生成可视为从一个分布中抽取样本的问题，比如说下面两个问题可以这样表示：

1. 纹理生成—— $x \sim p(x|x_0)$ (x_0 指纹理示例)
2. 风格迁移—— $x \sim p(x|x_0, x_1)$ (x_0 指内容图, x_1 指风格图)

因此人们提出的使用描述性网络（cnn等）进行提取的方法，将问题简化为寻找一个新图像，最

$$\operatorname{argmin}_{x \in \mathcal{X}} \|\Phi(x) - \Phi(x_0)\|_2^2.$$

小化双方统计距离这样一个最优化问题

也就是说，人们设计了这样一个网络，从随机初始化 z 开始，利用算法 A 进行优化，缩短双方的统计距，这样，只要我们的统计 ϕ 够好，我们就能产生一个足够好的图片

$$\underset{\mathbf{x} \in \mathcal{X}}{\text{localopt}}(\|\Phi(\mathbf{x}) - \Phi(\mathbf{x}_0)\|_2^2; \mathcal{A}, \mathbf{z}), \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \Sigma).$$

深度生成网络

另一种使用神经网络进行生成的方法是直接使用生成器，即输入 z ，输出新的图片 y ，其中 z 可以有规律的已知图像参数值，也可以是随机的初始化值

描述性纹理建模

那么我们该如何表现纹理呢？纹理一般是仅仅在局部的分布有规律，也就是说我们可以按照马尔科夫模型来对纹理建模，即原型纹理 x_0 上这些局部统计的空间平均近似其样本平均：

$$\phi(\mathbf{x}_0) = \frac{1}{|\Omega|} \sum_{i=1}^{|\Omega|} \psi \circ F(\mathbf{x}_0; i) \approx \underset{\mathbf{x} \sim p(\mathbf{x})}{E} [\psi \circ F_l(\mathbf{x}; 0)]. \quad ($$

其中 Ω 表示空间， ψ 是直方图操作符， F 是一组线性滤波器的输出

实验（具体内容）

模型整体结构：训练一个前馈网络，接受噪声 z 作为输入，生成纹理作为输出（风格迁移上的应用还接受 x 作为输入），并构建一个自动评估生成质量的损失函数（模型）

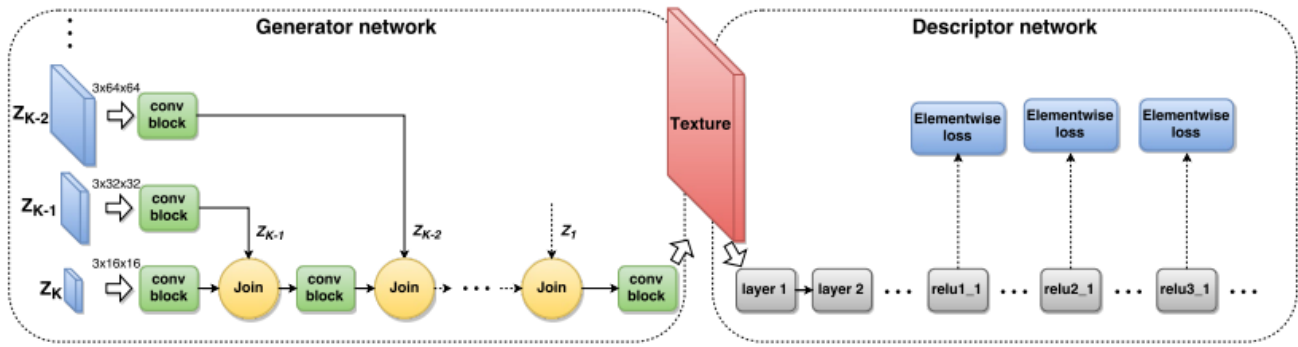
纹理和内容损失函数：

文中使用CNN测量原型纹理 x_0 和生成图像 x 之间的不匹配，使用Gram矩阵描述各层卷积核的卷积结果并作为纹理描述，并用如下公式计算纹理损失

$$\mathcal{L}_T(\mathbf{x}; \mathbf{x}_0) = \sum_{l \in L_T} \|G^l(\mathbf{x}) - G^l(\mathbf{x}_0)\|_2^2.$$

纹理生成器网络架构：

在这部分，作者先搭建了一个简单的模型架构：由卷积、非线性激活和上采样层组成，但后面发现多尺度架构效果更好，不仅有利于收敛而且参数更少



环形卷积：卷积的时候把图片卷起来，开头连上结尾，这样卷积到后面有地方要被漏掉时，可以用图片的前半部分补充（cv学过，但名字不一样）

多尺度生成器网络架构

输入：原型纹理 x_0 ，噪声 z ——包括 K 个随机张量 $z_i \in \mathbb{R}^{\frac{M}{2^i} \times \frac{M}{2^i}}$ （从均匀分布中独立同分布采样的）每个随机噪声张量首先被一系列卷积和非线性激活层处理，然后上采样两倍，最后作为额外的特征通道与下面尺度的部分处理张量连接。

上采样层：使用简单的最近邻插值法

在连接层之前插入批量归一化层，这显著有利于训练端到端训练

风格迁移网络架构

生成器网络除了输入 z 外还输入内容图 y ， z 直接应用于 y 上，在这里 z 增加到了 6（纹理生成是 5）

效果

由于这样的模型只需要单次评估，相较于之前迭代并反向传播的模型，稳重模型的使用速度快了近 500 倍，还减少了内存的使用