
AI Technology Cultural Products

人工智能技术 文艺作品篇



Victory won't come to us unless we go to it.

作者: 王晗

邮箱: wanghan@stu.scu.edu.cn

Version: 0.1

目 录



1	风格迁移，创作属于你的名画	1
1.1	风格迁移的前世今生	1
1.2	什么是风格	1
1.2.1	Gram 矩阵描述风格	1
1.2.2	Gram 矩阵的应用	2
1.3	通过深度学习进行风格迁移	3
1.4	总结	3
	参考文献	3
2	对抗生成，从无到有的创作	4
	参考文献	4

第 1 章 风格迁移，创作属于你的名画



1.1 风格迁移的前世今生

风格图像迁移指的是讲图像 A 的风格转换到图像 B 中，得到新的图像。其开山之作为《Image style transfer using convolutional neural networks》[1]。然后在短短的一段时间内，世界范围内的研究者在风格迁移领域发表了大量的论文，其中经典的一篇论文是名为《Perceptual losses for real-time style transfer and super-resolution》[2]。

1.2 什么是风格

1.2.1 Gram 矩阵描述风格

Definition 1.1 Gram 矩阵

n 维欧式空间中任意 $k \leq n$ 个向量 $\alpha_1, \alpha_2, \dots, \alpha_k$ 的内积所组成的矩阵

$$\Delta(\alpha_1, \alpha_2, \dots, \alpha_k) = \begin{pmatrix} (\alpha_1, \alpha_1) & (\alpha_1, \alpha_2) & \dots & (\alpha_1, \alpha_k) \\ (\alpha_2, \alpha_1) & (\alpha_2, \alpha_2) & \dots & (\alpha_2, \alpha_k) \\ \dots & \dots & \dots & \dots \\ (\alpha_k, \alpha_1) & (\alpha_k, \alpha_2) & \dots & (\alpha_k, \alpha_k) \end{pmatrix}$$

称为 k 个向量 $\alpha_1, \alpha_2, \dots, \alpha_k$ 的格拉姆矩阵 (Gram 矩阵)，它的行列式称为 Gram 行列式。

Gram 矩阵可以看做特征之间的偏心协方差矩阵（即没有减去均值的协方差矩阵），在特征图中，每个数字都来自于一个特定滤波器在特定位置的卷积，因此每个数字代表一个特征的强度，而 Gram 计算的实际上是两两特征之间的相关性，哪两个特征是同时出现的，哪两个是此消彼长的等等，同时，Gram 的对角线元素，还体现了每个特征在图像中出现的量，因此，Gram 有助于把握整个图像的大体风格。由此产生了表示风格的 Gram 矩阵，要度量两个图像风格的差异，只需比较他们 Gram 矩阵的差异即可。

总之，Gram 矩阵用于度量各个维度自己的特性以及各个维度之间的关系。内积之后得到的多尺度矩阵中，对角线元素提供了不同特征图各自的信息，其余元素提供

了不同特征图之间的相关信息。这样一个矩阵，既能体现出有哪些特征，又能体现出不同特征间的紧密程度。

1.2.2 Gram 矩阵的应用

在《Image style transfer using convolutional neural networks》[1] 出来之前，Gatys 还做了这个工作《Texture Synthesis Using Convolutional Neural Networks》[3]，他们发现如果让隐藏层的特征用协方差来进行约束，可以得到较好的纹理生成，他们发现如果用协方差（也就是 Gram 矩阵）来进行约束隐藏层特征的话，重建出来的特征虽然有些会保持，但是有些可能位置会打散。比如最右侧的一张图，人还是人，但是重建出来相当于“拼图”效果了。这是因为协方差本身就是去除了位置信息。那么既然协方差可以用于纹理生成，那么如果我们加上“让生成图的隐藏层特征与原图尽量一样，另一方面让生成图的打散特征与画的打散特征尽量相似”，这就是用神经网络做风格转换的最初想法。这也比较符合“风格”的定义，毕竟风格不应该具有位置信息，一种风格应该是与位置无关的。

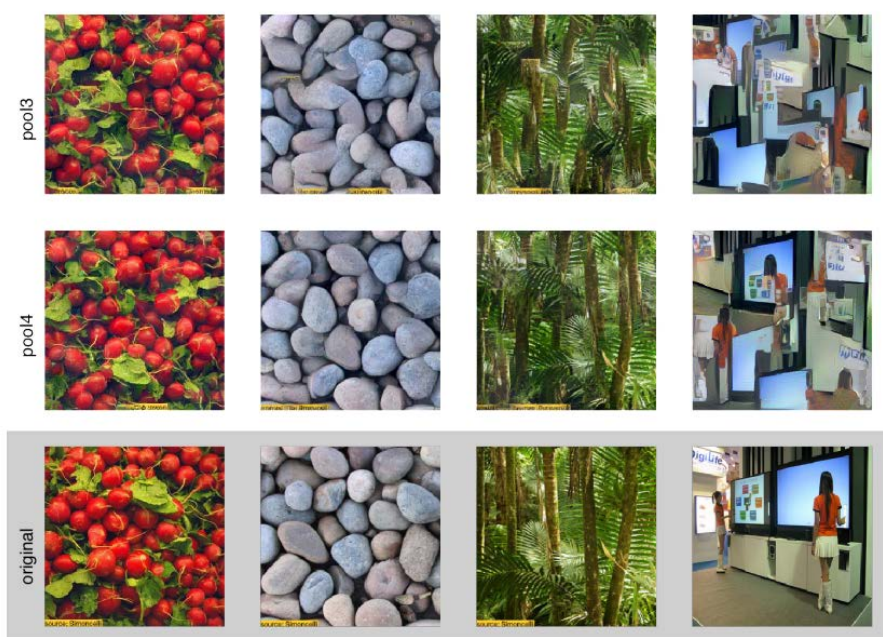


图 1.1: 纹理

需要注意的是 Gatys 的几篇论文没有解释为什么用 Gram 矩阵，其实可以这样认为，协方差就是一种二阶统计信息，我们要求输出图的什么信息与风格图相近，肯定不是特征图上单纯的逐点的相近，Gram 矩阵描述的就是全局特征的自相关，如果输出图与风格图的这种自相关相近，那么差不多是我们所理解的“风格”。当然，其实也可以用很多其他的统计信息进行描绘风格。这也就是后面有用直方图的，甚至直接简化成“均值 + 方差”进行描绘风格的。



1.3 通过深度学习进行风格迁移

下面约定 p 为风格图, a 为待转换的图, 即内容图。比如 p 为某张梵高的画, a 为某个场景, 生成为 f , 即具有梵高的画的风格的场景图。首先定义两个损失, l_{style} 和 $l_{content}$, 前者希望 f 和 p 在“风格”上尽量一致, 后者则希望 f 与 a 在内容上尽量一致。

$$l(a, f, p) = \alpha * l_{style}(p, f) + \beta * l_{content}(a, f) \quad (1.1)$$

其中, 公式 1.1 的 α 和 β 为两个损失的平衡参数。我们希望 l 尽量小, 采用梯度下降即可优化。所用的 CNN 网络是 VGG-19。开始训练时, 随机生成 a 同等大小的随机噪声图 x , 通过指定不同的层作为 content 损失的提取层 L_c 以及 style 损失的提取层 L_s , 使得 x 在 L_c 层得到的内容损失 $l_{content}(L_c, x)$ 与 $l_{content}(L_c, a)$ 尽量一样。同时, 使得 x 在 L_s 层得到的风格损失 $l_{style}(L_s, x)$ 与 $l_{style}(L_s, p)$ 也尽量一样。其中 $l_{content}$ 的具体形式可以由 MSE 来指定, 而 l_{style} 可由对应的 Gram 矩阵来计算。

1.4 总结

参考文献

- [1] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, IEEE, 2016.
- [2] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European Conference on Computer Vision*, pp. 694–711, Springer, 2016.
- [3] L. Gatys, A. S. Ecker, and M. Bethge, “Texture synthesis using convolutional neural networks,” in *Advances in Neural Information Processing Systems*, pp. 262–270, 2015.



第 2 章 对抗生成，从无到有的创作



参考文献