

DOI:10.19651/j.cnki.emt.1802112

基于新的风格损失函数的图像风格转换方法

钱燕芳 王 敏

(河海大学 计算机与信息学院 南京 211100)

摘要: 虽然基于深度学习的图像风格转换方法已经取得了很大的进展,但是这些方法都没有考虑到生成图像的线条扭曲现象,为此提出直方图损失和转换 Gramian 矩阵相结合的方法。图像的直方图信息可以判断出图像质量的好坏,在图像风格转换中使用直方图损失,不仅可以增强图像,还可以使生成的图像更加稳定。转换 Gramian 矩阵类似于 Gram 矩阵,但是前者提取出图像纹理信息更加完整,还考虑到了图像的空间排列信息。实验结果表明,这两种方法的结合不仅能使生成的图像没有线条扭曲,还能减少图像生成的迭代次数。

关键词: 深度学习; 图像风格转换; Gram 矩阵; 转换 Gramian 矩阵

中图分类号: TN919.81 文献标识码: A 国家标准学科分类代码: 520.2040

Method of image style transfer based on new style loss function

Qian Yanfang Wang Min

(College of Computer and Information, HoHai University, Nanjing 211100, China)

Abstract: Although great progress has been made in image style transfer based on deep learning, these methods were not took into account the distortion of lines in the generated image. Therefore, new method combining histogram loss with transformed Gramian matrix is proposed. The use of histogram loss in image style transfer is not only enhanced the image, but also made the generated image more stable. Transformed Gramian matrix is similar to Gram matrix, but the former is extracted more complete texture information, and also took into account the spatial arrangement of an image information. The experimental results show that combination of two methods can not only make the generated image without line distortion, but also reduce the number of iterations in image generation.

Keywords: deep learning; image style transfer; Gram matrix; transformed Gramian matrix

0 引言

随着人们对图像美化的要求越来越高,图像滤镜等技术也得到越来越多的学者的关注,尤其是图像的艺术风格化,图像的艺术风格化就是将一幅普通图像转换成具有抽象画、中国水墨风、或者动漫风格的图像。传统的图像风格转换方法可以分为 3 类:基于笔触的风格化、基于纹理合成的风格化和对物理过程建模的风格化^[1]。但是这 3 类方法都比较繁琐。

随着深度学习的崛起,Gatys 等^[2]第一次发现深度学习也可以应用于图像的风格转换。首先可以用 Gram 矩阵来表示一幅艺术图像的风格信息,因为 Gram 矩阵表示的是特征映射之间的相关性,最后可以得到较好的纹理生成。而且可以将提取图像的内容特征与纹理特征的工作分开进行,然后再把图像的内容与风格融合起来,就可以得到想要的艺术风格图像^[3]。Prisma 软件就是对 Gatys 等的研究成

果改进优化后得到的。目前,Prisma 不仅可以对图像进行风格转换,还可以对视频进行风格转换,但是视频长度只有几十秒。

后来,Li 等^[4]提出了在卷积神经网络中加入马尔科夫随机场的图像风格转换方法,利用 MRF 可以将特征映射分成很多的小块,然后找小块之间的匹配,这种方法可以更好地保留局部信息。这个方法还在损失函数加入正则化项,为了使合成的图像更平滑。这个正则化项,即总变差,相当于在图像的每个像素的水平方向和垂直方向求差值的平方。随后,他们对这个方法做了改进^[5],改进后的办法可以做到实时的图像风格转换。

由于 Gatys 等的方法生成图像的速度非常慢,所以 Ulyanov 等^[6]提出了一个新的深度学习方法用来纹理合成和图像风格化,这个方法是将计算的负荷移到学习阶段。给出一个纹理图像,训练紧致的前馈卷积网络,产生相同纹理的多个任意大小的样本,并从一个给定图像根据艺术风

格转给任意其他图像。这个方法在时间上达到了 500 倍的提速。文献[7]提出了一种基于前馈网络的图像风格转换方法,加入了感知损失函数,速度提升了 3 个数量级。

随后,Li 等^[8]首次将风格转换应用到了人脸属性上,还可以对性别进行判别。还有 Chen 等^[9]提出了快速风格转换。Dumoulin 等^[10]提出了 N 个风格共用一个模型的方式,缓解了模型存储耗费空间大的问题。这个方法还可以实现同一内容图像融合多种风格图像的图像转换方法。

2016 年,Berger 等^[11]提出一种新的纹理生成方法,即转换 Gramian,这个方法很好地保留了纹理的空间排列信息。2017 年,Risser 等^[12]提出在图像风格转换中加入直方图损失来增加生成图像的稳定性与可控性。后来,Luan 等^[13]提出了图像写实正则化,这个方法可以实现局部颜色转换,使生成的图像看上去像照片。因为 Gram 矩阵是表示整幅图像的协方差,但是没有考虑到局部信息,所以该方法加入了语义分割。几个月后,Huang 等^[14]基于 GAN 网络提出加入 AdaIN 层,损失函数变成了均值和方差,而且可以做到实时转换,但是生成的图像会有块状效果。为了提高图像风格转换的实用性,Liao 等^[15]提出一种新的将两张图片直接进行视觉属性迁移的方法。这个方法针对的是两张具有不同内容但是有相似语义的图像,基本可以保留两张图中的内容与结构,同时融入参考图像的视觉属性,任何随手拍的照片都可以进行转换。

本文主要是将图像直方图损失和新的风格损失结合,新的损失函数比 Gatys 等提出的收敛速度更快,只需要迭代 5 次即可,而且生成的图像也比较稳定,没有线条扭曲以及风格特征分散的现象。

1 直方图损失

因为图像风格转换问题没有客观评价标准,通常是指根据人眼来判断生成的图像是否足够美观。所以找到一个能比较精确地度量两幅图像的差距的损失函数很重要。最初,Gatys 提出用 MSE 距离表示内容图像的损失,用 Gram 矩阵表示风格图像的损失。但是 Gram 矩阵并不能正确地表示一幅图像的纹理,因为有可能存在完全不同的图像,有着同样的 Gram 矩阵,毕竟 Gram 矩阵是从图像的全局信息考虑的。所以为了使生成的图像更加稳定,需要加入直方图损失来控制。

如果只是直接计算输入图像 \mathbf{S} 与输出图像 \mathbf{O} 的直方图的 L2 损失,那么优化过程中,对直方图损失项求导就会发现处处都是 0,这样就对优化起不到任何作用。所以在直方图匹配的基础上重新定义这个直方图损失。首先,对特征激活合成的图像与原图像做直方图匹配,这样得到的特征映射就与原纹理图像 \mathbf{S} 匹配,在反向传播过程中,必须对每个特征映射进行匹配。

用 \mathbf{O}_i^l 表示第 i 个特征映射在第 l 层输出的特征映射,计算 \mathbf{O}_i^l 的归一化直方图,并将其与原始输入图像 \mathbf{S} 的直方

图匹配,得到重组后的特征映射,用 $R(\mathbf{O}_i^l)$ 表示,那么可以得到直方图损失:

$$L_{\text{hist}} = \sum_{l=1}^L \gamma_l \| \mathbf{O}_i^l - R(\mathbf{O}_i^l) \|_F^2 \quad (1)$$

式中: L 表示网络层数; γ_l 表示各网络层在直方图损失中所占的比例。

反向传播计算导数时,发现 $R(\mathbf{O}_i^l)$ 的导数几乎处处都为 0,为了计算效率,可以将其看作一个常量。

2 转换 Gramian 矩阵

2.1 Gram 矩阵

Gatys 等最先提出用 Gram 矩阵来表示一幅纹理图像的纹理信息。第 l 层的 Gram 矩阵用 $\mathbf{G}^l \in \mathbf{R}^{N_l \times N_l}$,其中, N_l 表示第 l 层特征映射的个数,则 \mathbf{G}^l 中的元素 G_{ij}^l 是第 l 层第 i 个特征映射与第 j 个特征映射的内积。

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l = F_{ii}^l, F_{jj}^l \quad (2)$$

式中: F_{ij}^l 表示第 l 层第 i 个卷积核在第 j 个位置处的输出。所以,可以得到总的风格损失:

$$L_{\text{gram}} = \sum_{l=1}^L w_l \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \quad (3)$$

其中, \mathbf{G}^l 和 \mathbf{A}^l 分别表示参考风格图像和输出图像第 l 层的 Gram 矩阵; M_l 表示第 l 层卷积核的大小,等于卷积核的高乘以卷积核的宽; w_l 则表示各网络层在风格损失中所占的比例。

在特征映射中,每个值都是一个卷积核在特定位置上的卷积结果,观察式(2)可知,Gram 矩阵中的每个值都表示两个特征映射的内积,如果两个特征映射同一维度的值都比较小,那么得到的值就更小,反之,得到的值更大,所以 Gram 矩阵可以表现出两个特征之间的相关性,比如哪两个特征同时出现的,哪两个特征是此消彼长的。所以 Gram 矩阵可以很好地表示一幅图像的风格。

2.2 转换 Gramian 矩阵

虽然 Gram 矩阵可以提取图像的全局静态特征,但是忽略了图像的全局排列信息,所以生成的纹理是分散的,还会出现线条扭曲现象。为了解决这个问题,可以使用转换 Gramian。与 Gram 矩阵直接计算不同特征映射之间的共生性不同,转换 Gramian 是计算一个特征映射 \mathbf{F}^l 与空间转换特征映射 $T(\mathbf{F}^l)$ 的内积。 T 是一个空间转换,相当于计算局部特征与相邻特征的相似性, $T_{x,+\delta}$ 表示特征映射在水平方向上向右平移 δ 个像素得到,即原特征映射移除前 δ 列像素, $T_{x,-\delta}$ 则表示移除后 δ 列像素,可以得到转换 Gramian:

$$G_{x,\delta,ij}^l = \frac{1}{M_l} T_{x,+\delta}(F_{i,:}^l), T_{x,-\delta}(F_{j,:}^l) \quad (4)$$

转换 Gramian 其实是计算位于 (i,j) 处的特征与位于 $(i,j+\delta)$ 或 $(i,j-\delta)$ 处的特征的相关性,所以可以很好地保留特征的排列信息。式(4)只计算了 x 方向上的转换

Gramian, 为了考虑全局, 还需计算 y 方向上的转换 Gramian, 分别用 $\mathbf{G}_{x,\delta}^l$ 与 $\mathbf{G}_{y,\delta}^l$ 表示, 可以得到 Gramian 损失:

$$L_{\text{gramian}} = \frac{1}{2} (\|\mathbf{G}_{x,\delta}^l - \mathbf{A}_{x,\delta}^l\|_F^2 + \|\mathbf{G}_{y,\delta}^l - \mathbf{A}_{y,\delta}^l\|_F^2) \quad (5)$$

其中, $\mathbf{A}_{x,\delta}^l$ 和 $\mathbf{A}_{y,\delta}^l$ 分别表示输出图像在 x, y 方向上的转换 Gramian 矩阵。

3 图像风格转换

本文总的风格损失定义如下:

$$L_{\text{style}} = aL_{\text{hist}} + bL_{\text{gramian}} \quad (6)$$

其中, a, b 分别用来控制两个风格损失函数所占的比重。内容图像使用的是 MSE 距离, 定义如下:

$$L_{\text{content}} = \frac{1}{2} \|F^l - P^l\|_F^2 \quad (7)$$

如果在损失函数中加入总变差正则化项, 对于生成的图像效果没有特别大的影响, 因此为了减少计算量, 就不考虑了。总的损失函数如下:

$$L_{\text{total}} = \alpha L_{\text{content}} + \beta L_{\text{style}} \quad (8)$$

本文实验用的是已经训练好的 VGG19 模型, 采用的池化方法是均值池化。初始化的输出图像为原始内容图像。输入的内容图像需要进行预处理, 即减去均值 RGB, 图像内容损失层为 conv4_2, 直方图损失层为 conv1_1, conv4_1。计算的时候发现直方图损失过大, 所以将其除以 100, 以消除量纲影响。转换 Gramian 损失层为 pool1、pool2、pool3、pool4。平移像素个数 δ 固定为 4, 每个损失函数中, 各层所占比例是平均的, 风格与内容在生成图像中所占比重分别为 100 和 1, 直方图损失与转换 Gramian 损失比重分别为 1 和 0.5。

4 实验结果与分析

如图 1 所示, 第 1 列第 2 列分别是运用直方图损失与转换 Gramian 矩阵所得到的结果图, 第 3 列为两者的结合所得到的结果图。很明显, 第 3 列图像的质量比前两列好了很多。



图 1 不同损失函数实验结果对比

Gatys 等提出的图像风格转换方法大概需要迭代 1 000 次才能产生艺术效果较好的图像。但是用直方图损失结合

转换 Gramian 后, 只需要迭代 5 次即可, 如图 2 所示, 分别是两幅图迭代 5 次的所得到的图像(上下两行结果图的内容图与风格图如图 3 所示)。可以发现, 即使是迭代了 1 次的结果图, 也可以表示图像风格转换的结果, 虽然风格化不是很明显, 观察迭代 3 次、4 次与 5 次的结果, 可以发现这 3 幅图的差异很小。所以对于某些风格图像, 只需要迭代 2~3 次就可以了, 而迭代不同次数所得到的结果图, 可以对应不同风格比重的效果图。

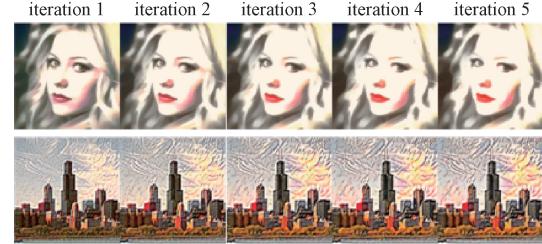


图 2 不同迭代次数所得结果对比图

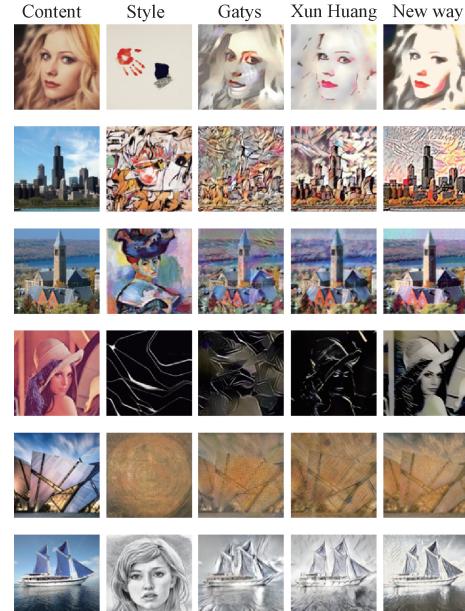


图 3 不同方法实验结果对比

图 3 为本文结果与其他论文结果的对比。第 1 列为内容图像, 第 2 列为风格图像, 第 3 列为 Gatys 等人的结果, 第 4 列为 Huang 等的实验结果, 第 5 列为本文实验结果。

对比图 3 的后 3 列, 可以发现, 本文实验所得到的图像在质量上有了较大的提高, 观察图 3 中的第 2 行, Gatys 等的实验生成的图像有非常多的线条扭曲现象, Huang 等的实验生成的图像也有少许的线条扭曲, 而本文的实验生成的图像明显没有线条扭曲的现象, 第 4 行也能证明这一点。另外第 4 行 Huang 等的实验生成的图像虽然没有线条扭曲, 但是图像的颜色看上去很糟糕, 而本文实验所产生的结果却避免了这点, 生成的图像看上去很美观。

图4所示为各种内容风格的结果对比,第1列和第1行分别为风格图和内容图。可以发现,生成的结果图像质量都不错,所以基于新损失函数的图像风格转换方法还是比较通用的。



图4 各种内容风格结果对比图

第1列的第3、第4个风格图像相对于其他几张风格图像,纹理效果比较明显,但是相对于图4中的风格图,纹理却并不明显。部分结果图的细节图,如图5所示,都能看到明显的对应纹理细节。所以这两种损失函数的结合可以很好地捕捉内容图的纹理特征,即使是较少的纹理。



图5 结果图局部细节图

5 结 论

为了解决图像风格转换中出现的线条扭曲现象,提出将直方图损失和转换Gramian矩阵相结合的方法。实验结果表明,这个方法不仅使生成的图像没有曲线扭曲现象,还减少了图像的迭代次数,可以以较快的速度得到结果图像。但是对比现今速度最快的图像风格转换,这个新方法在增加了损失函数的同时,也增加了模型的复杂度,如果加入正则惩罚项来控制也许能更快些。另一方面,这个方法没有考虑对生成的图像添加颜色因子来控制生成合理颜色的图像。

参考文献

- [1] 邓盈盈,唐帆,董未名.图像艺术风格化的研究现状[J].南京信息工程大学学报(自然科学版),2017,9(6):593-598.
- [2] GATYS L A, ECKER A S, BETHGE M. Texture synthesis using convolutional neural networks[J]. Febs Letters, 2015, 70(1):51-55.
- [3] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]. Computer Vision and Pattern Recognition, IEEE, 2016:2414-2423.
- [4] LI C, WAND M. Combining Markov random fields and convolutional neural networks for image synthesis [C]. Computer Vision and Pattern Recognition, IEEE, 2016:2479-2486.
- [5] LI C, WAND M. Precomputed real-time texture synthesis with Markovian generative adversarial networks [C]. European Conference on Computer Vision, 2016:702-716.
- [6] ULYANOV D, LEBEDEV V, VEDALDI A, et al. Texture networks: Feed-forward synthesis of textures and stylized images [C]. International Conference on Machine Learning, 2016:1349-1357.
- [7] LI C, WAND M. Perceptual losses for real-time style transfer and super-resolution[C]. European Conference on Computer Vision, 2016:694-711.
- [8] LI M, ZUO W, ZHANG D. Deep identity-aware transfer of facial attributes [J]. Computer Science, 2016, arXiv:1610.05586.
- [9] CHEN T Q, SCHMIDT M. Fast patch-based style transfer of arbitrary style[J]. Computer Science, 2016, arXiv:1612.04337.
- [10] DUMOULIN V, SHLENS J, KUDLUR M. A learned representation for artistic style[J]. Computer Science, 2016, arXiv:1610.07629.
- [11] BERGER G, MEMISEVIC R. Incorporating long-range consistency in CNN-based texture generation[J]. Computer Science, 2016, arXiv:1606.01286.
- [12] RISSE E, WILMOT P, BARNES C. Stable and controllable neural texture synthesis and style transfer using histogram losses[J]. Computer Science, 2017, arXiv:1701.08893.
- [13] LUAN F, PARIS S, SHECHTMAN E, et al. Deep photo style transfer[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:6997-7005.
- [14] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]. IEEE International Conference on Computer Vision, 2017:1510-1519.
- [15] LIAO J, YAO Y, YUAN L, et al. Visual attribute transfer through deep image analogy [J]. ACM Transactions on Graphics, 2017, 36(4): 1-15.

作者简介

钱燕芳,硕士研究生,主要研究方向为图像处理。
E-mail:1275756082@qq.com

王敏,副教授,主要研究方向为机器学习。
E-mail:119998413@qq.com