基于深度学习的风格迁移综述

王文新

（自动化系，2019011354）

摘要：基于深度学习的神经网络算法在图像处理领域取得了优异的表现。其中，基于卷积神经网络的风格迁移算法是最为引人注目的算法之一。本文分析了几种风格迁移算法的设计思路、灵感来源、发展与创新之处，就其中仍存在的一些问题提出了一些可能的解决方案。

关键词：人工智能；风格迁移；图像处理

Survey of Style Transfer Based On Deep Learning

WANG Wenxin

Department of Automation 2019011354

Abstract

Neural network algorithms, based on deep learning, have showed great performance in image processing area. Of which, the style transfer algorithm, based on convolutional neural network, is one of the most focused algorithms. This paper analyzed a few instances of style transfer, digging out their ideas in design, inspirations and creativity. And tried to give a few possible solutions to the existing problems.

Key words

artificial intelligence; style transfer; image processing

1、引言

近年来，基于深度学习的人工神经网络技术已经渗透入人类社会的各个领域。一方面，深度学习与科技发展紧密结合，其既得益于技术的进步，又推动着科技进一步发展；另一方面，在深度学习技术受到广泛应用的当下，人们的生产生活方式也受到了广泛而深远的影响。而当深度学习被应用于绘画艺术领域，同样取得了令人惊叹的结果，风格迁移、图像生成、以至画作生成等基于深度学习的图像处理算法逐渐成为了人们的又一个关注焦点。

2、风格迁移

2.1、概述

风格迁移是一类近年被发明并具有良好前景与应用价值的图像处理算法。在训练优良、迭代次数充足的情况下，能完成以指定风格重绘图片的任务。前日热度颇高的Prisma手机修图软件便是基于风格迁移算法进行图片的后期处理的。除了这样直接地利用特性产生效果之外，风格迁移还有着多种可能的应用场景，如通过指定多种光照条件作为风格，对图像识别神经网络的训练数据集，也即批量图像进行风格迁移，可以得到近似于多种光照条件下同一物体的图像，再进行训练或许可以达到扩大数据集防止识别网络过拟合的效果，并同时增强图像识别网络对不同光照条件下，识别同一物体的鲁棒性，而后者对诸如人脸识别等精度要求较高的应用场景有着重要意义。

2.2、定义与实例

风格迁移一词一般描述如下过程：给定两张图片A（内容图像）、B（风格图像），生成一张图片，使之具有图片A的内容，同时具有从B提取的风格。生成新图片的过程也可通过直接在图片A的基础上进行修改而完成。这一技术在不破坏内容图像的画面的基础上实现了风格的转换，输出图像呈现出内容图像与指定风格的较好结合。如图1（a）为作为内容图像的摄影，图1（b）为作为风格图像的，日本画家葛饰北斋所作浮世绘版画《神奈川冲浪里》，经过风格迁移后得到了组合摄影内容与画作风格的作品（图1（c））。



图1（a）摄影作品



图1（b）《神奈川冲浪里》（葛饰北斋，19世纪初）



图1（c）风格迁移合成图像

2.3、第一代算法设计

第一代风格迁移算法最先由Gatys等人发明，他们在早年对图像风格提取的研究[2]基础上提出了风格迁移[3]。（事实上，可以看出，在风格迁移这一语境下，原文作者更倾向于将“风格”表述为“特定形状与特定色彩的组合”，也即图像的色彩特征）。

首先，风格（或者说色彩特征）的提取与应用有赖于特殊的损失函数设置。区别于通常作为图像比对依据的L1或L2距离，Gatys等人利用Gram矩阵进行图像间的比对。实验表明，通过计算Gram矩阵进行图像比对具有保留局部特征而淡化该特征在全局的位置的作用。这与风格迁移中“风格图像的色彩特征部分应保留”的要求如出一辙。与此同时，在进行与内容图像的比对时，可选用传统的L1或L2距离，用以强调其内容。

其次，每一次反向传播都对整张图片进行Gram矩阵及L2（或L1）距离的计算，其时间开销是不可承受的，故可通过卷积运算在尽可能保留信息的前提下将图片进行压缩，进而达成快速的Gram矩阵计算与损失函数计算。原文作者选择了VGG-16网络进行这样的卷积压缩运算。

VGG-16是一个深度卷积神经网络，由Karen Simonyan和Andrew Zisserman于2014年提出[1]，其结构如图2，包含了由5个池化层连接的13个卷积层、3个全连接层以及最后的SoftMax输出。经过训练后，可将一张图片输入进行前向传播，通过逐层的卷积提取出图由低级到高级的特征，进而选择部分层所表示的部分特征图即可进行快速的损失函数计算，进而快速进行反向传播。

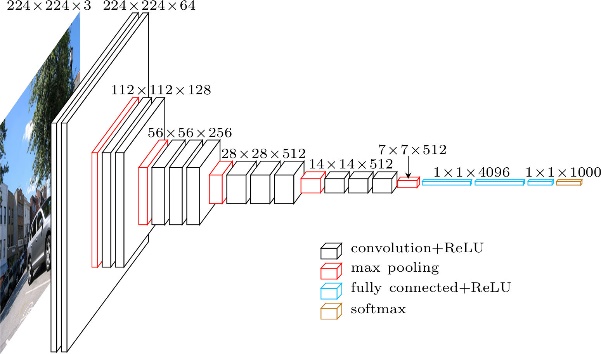


图2 VGG-16的网络结构示意[1]

进而Gatys等人所描述的整个风格迁移算法的流程（图3）大致如下：将内容图像A、风格图像B依次输入VGG-16损失网络中，选择A前馈结果的多层（多级）特征图计算Gram矩阵作为风格图像的数据参照，而选择B较高层的特征图作为内容图像的数据参照。每一次迭代中，将输出图像（开始时为白噪声图像，如图3中“Start with noise”）输入VGG-16网络计算，分别提取与上文一致的特征图，分别计算风格损失和内容损失，将二者的线性加权平均作为总损失函数，进行反向传播，对原输出图像的各个像素进行修改，再次输入VGG-16网络，开始下一次迭代。进而通过多次迭代完成使输出图像继承B的内容的同时具备A的风格的任务。

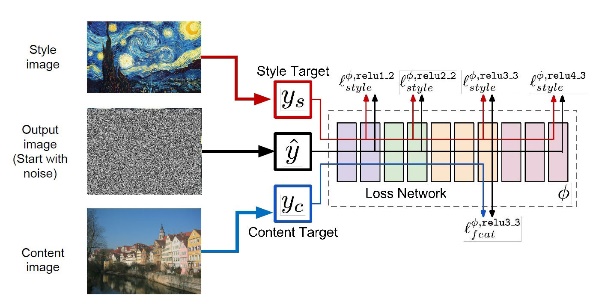


图3[4] Gatys等人的风格迁移算法流程示意图

这一算法的提出是具有开创性意义的，它将深度学习技术应用于图像风格的重塑，为深度学习的应用场景开辟了一个新的领域，也为未来更多有价值的应用的出现奠定了基础。算法中损失函数的设计也成为了风格迁移以至纹理提取研究及应用领域中经典的方案。

但与此同时，该算法也存在着一些问题。

首先，其第一次迭代开始采用白噪声图像，使得内容图像的细节难以再现，潜在地使迭代次数大大增多，收敛缓慢，笔者推测将原内容图片直接作为第一次迭代的输入或许可以提高收敛速率，如图4，笔者在将内容图像直接作为第一次迭代的输入后，仅进行数十次迭代便已经取得了较好的结果。但同时应当注意权重值的调整，避免损失函数不收敛的情况发生。



图4（a）内容图像 摄影作品



图4（b）风格图像 《星月夜》（梵高，1889）



图4（c）经风格迁移后的合成图像

其次，反向传播时，对输入图片的每个像素进行调整，时间开销过大。第三，就应用场景来看，大多数情况下，在使用风格迁移技术时，选择的风格图片相对固定，而内容图片则变化不一，当对多幅不同的内容图片以同样的风格图片为参照进行风格迁移时，就更加暴露出了这一算法的时间复杂度问题，其对同一风格缺失记忆能力。此处的后两个问题在第二代风格迁移技术中得到了良好解决。

2.4、第二代算法设计

第二代风格迁移算法由Johnson等人提出[5]，其继承了第一代风格迁移算法的损失函数设计及计算方案（以VGG-16为损失网络），但在输出图像生成的方式上做出了改进，这一改进基于Johnson等人对图像超分辨率这一领域的研究。

图像超分辨率，一般指从一或多幅低分辨率图像出发，“恢复”出一张高分辨率图像。这里的“恢复”其实是对图像的重新生成。目前图像超分辨率技术一般有三种实现方式：基于插值算法、基于重建、基于机器学习。Johnson等人就第三种实现方式进行了研究。

由Johnson等人提出的基于深度学习的图像超分辨率算法[5]，其核心是通过卷积运算，对图像进行上采样从而获得幅面扩大的图像。后通过逐次与原图像进行L2距离对比获得损失函数，进而完成反向传播，逐步调整上采样卷积网络的权重，使得生成的高分辨率图像逐步趋近原图，完成图像超分辨率化的任务。

这种图像超分辨率技术从某种意义上讲是图像压缩卷积运算的“逆运算”，基于足够的训练，二者能够配合完成对某图像压缩——复原（或者说近乎复原）的任务，这一点为解决第一代风格迁移算法中对原图调整复杂度过大的问题提供了良好的解决方案。

借助这一方法，Johnson等人提出了第二代风格迁移技术（如图5）。固定一幅风格图像，选取内容图像，对二者进行VGG-16特征提取，而将上一次迭代输出图像直接进行下采样卷积压缩，经过残差网络进行风格迁移后，进行上采样卷积超分辨率处理，作为此次迭代输出图像。对输出图像的评估，即特殊的损失函数计算与第一代相似。而这时反向传播将调节生成网络（“风格迁移网络”）的权重值，使得生成的图像逐步贴合“保留内容，提取风格”的要求。

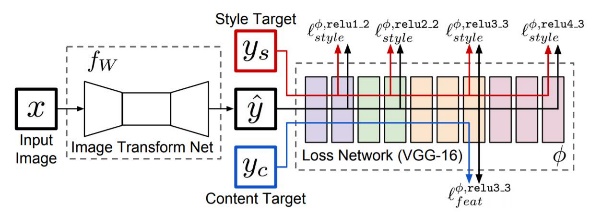


图5[5] Johnson等人的风格迁移算法流程示意图

这一方法相较第一代风格迁移算法的优势主要有以下几个方面：第一，利用卷积网络特性，减小反向传播的规模，减少了时间消耗；第二，针对同一风格图像进行大量内容图像数据的训练后，该生成网络的权值模型可直接应用于以该风格图像为基准的任意内容图像的风格迁移，此时的风格迁移将以前向传播的方式进行，大大加快了风格迁移的速度。

文首示例中的合成图像即笔者通过第二代风格迁移技术制作，与此同时，可能由于采用了“将内容图像直接作为第一次迭代的输入图像”的方法，内容图像的细节保留情况可观。

3、总结

纵观两种风格迁移算法，都基于一定的前置研究才得以开发，足够的基础研究为高层应用中新思路的出现创造了可能。在两种算法的设计中，除创新思维的体现，也不乏对问题的透彻分析与考察，分析现有问题再针对解决，促成了两次具有创造性价值的算法迭代。

风格迁移算法自诞生以来，便吸引了大量关注，除其巨大的应用价值之外，其精妙的设计思路也有重要的参考意义，在走入艺术领域，尤其是绘画领域后，也激起了人们关于人工智能创作版权、主体性、艺术价值等哲学问题的讨论，这里不再过多赘述。总而言之，风格迁移算法必然会成为人工智能领域的一个重要课题。

参考文献

1. Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv: 1409.1556, 2014
2. Leon A. Gatys and Alexander S. Ecker and Matthias Bethge, “A Neural Algorithm of Artistic Style,” arXiv preprint arXiv: 1508.06576, 2015
3. Gatys L A , Ecker A S , Bethge M . Image Style Transfer Using Convolutional Neural Networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
4. <https://zhuanlan.zhihu.com/p/57564626>, 2019-12-11.
5. Justin Johnson and Alexandre Alahi and Li Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” arXiv preprint arXiv: 1603.08155, 2016