

# 基于神经网络的风格迁移算法综述

21821330 吴 协

## 一、简介

20 世纪 90 年代中期，随着计算机图形学的飞速发展，人们已经不再满足于像照片一样真实的绘制效果，开始大力追求具有艺术化效果的、更具含义表达能力的画面，这就是非真实图形学绘制。非真实感图形学绘制的方法有很多，其中，采用机器学习的方法直接生成不同艺术风格的作品近几年非真实感绘制领域的一个重要方向。而基于神经网络的风格迁移算法（Neural Style Transfer，简称 NST）是风格迁移的关键算法，本文着重梳理了近两年 NST 算法的发展脉络，分析讨论了各种算法的特点，对 NST 算法做一个综合性的介绍。

## 二、背景

在神经网络之前，图像风格迁移的程序有一个共同的思路：分析某一种风格的图像，给那一种风格建立一个数学或者统计模型，再改变要做迁移的图像让它能更好的符合建立的模型。这样做出效果还是不错的，但一个很大的缺点：一个程序基本只能做某一种风格或者某一个场景。因此基于传统风格迁移研究的实际应用非常有限。

随着深度学习算法的不断发展，一个深度训练的卷积网络（CNN）取得显著的成功之后，人们就想到，CNN 是否也能应用在图像的风格迁移方面。2016 年，Gatys 等人在 A Neural Algorithm of Artistic Style 这篇文章[1]中提出了基于神经网络的风格迁移算法（Neural Style Transfer，简称 NST），实现了对艺术作品进行不同艺术风格的转换。这篇文章首次将 CNN 引入非真实感图形学的领域，具有里程碑的意义。随后，人们对 NST 算法不断改进，力求达到更好的视觉效果和更好的鲁棒性。

## 三、NST 算法

NST 算法的原理主要基于以下两点：1. 两张图像经过预先训练好的分类网络，若提取出的高维特征(high-level)之间的欧氏距离越小，则这两张图像内容越相似。2. 两张图像经过与训练好的分类网络，若提取出的低维特征(low-level)在数值上基本相等，则这两张图像

越相似。基于这两点，就可以设计合适的损失函数优化网络。Gatys 等人在 A Neural Algorithm of Aritistic Style 中构造了一个多层的卷积网络，通过最小化定义的 content loss 和 style loss 最后生成一个结合了 content 和 style 的图像。

其中，content loss 定义如下：

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2 .$$

style loss 定义如下：

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

$$\mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^L w_l E_l$$

最后，就是采用优化方法来最小化模型 loss：

$$\mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x})$$

#### 四、对NST算法的改进

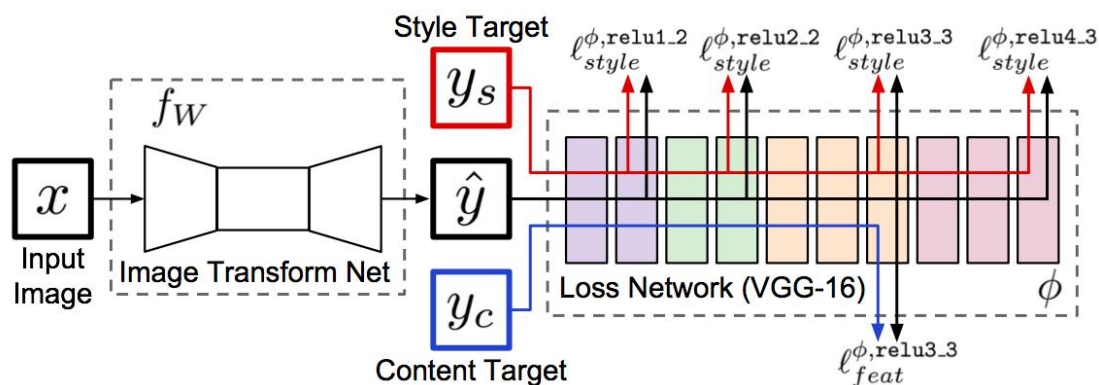
Gastys 等人提出的 NST 算法生成的图片看起来很不错，但是仍存在以下问题：1. 由于每次迁移都要对网络进行训练，速度是非常慢的，无法实现实时迁移。2. 风格迁移效果不好，例如色彩偏差或照片失真等。

针对存在的不足，人们提出了以下二种改进方法：

##### 1. 实时风格迁移(Real-Time Style Transfer)

2016 年 3 月，斯坦福大学的一组研究人员发表了一篇论文[2]，概述了实现实时样式转换的方法。他们能够训练神经网络，将单一风格应用于任何给定的内容图像。该论文题为 Perceptual Losses for Real-Time Style Transfer and Super-Resolution，表明可以训练神经网络将单一样式应用于任何给定的内容图像，因为只进行一次网络的前向计算，速度非常快，可以达到实时的效果。

论文的主要架构是：



该架构主要由两个神经网络构成，第一个为深度残差卷积网络（Image Transform Network），第二个为损失网络（Loss Network），由训练好的 VGG-16 网路构成，用于提取特征。

实现方式：结合两种方法，使用前馈卷积神经网络对 perceptual loss 进行优化并产生图像，图像不是由随机噪声生成，而是由 Input Image 变换而成。

实现细节：1. 使用残差网络下采样 input image，然后再通过上采样插值产生新的 output image。2. 基于欧几里得距离得到 Content Loss，并进行了归一化处理。3. 在 loss 网络的每一层都求出 Gram 矩阵，然后对应层之间计算欧式距离，最后将不同层的欧氏距离相加，得到最后的风格损失。

2. 控制神经风格转移中的感知因素（Controlling Perceptual Factors in Neural Style Transfer）

文章[3]提出了如下的假设：图像风格可以被分解为空间区域、颜色与光照信息和跨空间尺度这三个因素。

(1) 空间因素的控制（spatial control）

通过引入一个新的空间引导矩阵，来控制风格图像的哪一部分用来风格化内容图像对应的图像区域。

目标是控制风格图像的哪个区域用于风格化内容图像中的每个区域。将包含在风格图像中的每个图层的特征图与 R 个引导通道  $T_r$  相乘，并且为风格图像中的每个 R 区域计算一个空间引导的 gram 矩阵。

(2) 颜色因素的控制（colour control）

有两种方法，分别是颜色直方图匹配（colour histogram

matching) 和仅亮度转移 (luminance-only transfer)

在以前的风格转换中,生成图的颜色都会最终变成 style 图的颜色,但是很多时候我们并不希望这样。

颜色直方图匹配采用的方法就是将风格图的颜色转换成内容图的颜色。转换方法采用的是一种线性方法,该转换可以保证风格图与内容图有一致的均值和协方差。或者采用图像类比的方式,也可以得到较好的转换效果。

仅亮度转换就是将 RGB 转换成 YIQ,只在 Y 上进行风格转换,因为 I 和 Q 通道主要是保存了颜色信息。由于人们主观感受到的图像风格大部分是指图像的几何形状、笔触和纹理等特征,这些特征很大程度上是跟图像的明暗变化有关,所以仅对亮度进行转换,可以保留图像原有的颜色不变,又能实现风格化。

### (3) 空间尺度因素的控制 (scale control)

一种在不同空间尺度上独立控制风格化的方法,可以是选择不同比例的单独样式。例如,我们希望将一幅作品的精细笔触与另一幅图像的粗尺度角形几何形状相结合。

由于简单地将不同尺度不同图像的 Gram 矩阵结合起来,就不能对不同的尺度进行独立的控制。所以采用的方法就是创造一个新的样式图片,该样式图片有一幅图片的精细尺度和另一幅图片的粗尺度。粗尺度样式的图像就是使用较低层的 gram 矩阵。

使用粗尺度的样式图像初始化优化过程,并完全省略内容损失,以便粗尺度图像中的细微纹理将被完全替换。因为当尺寸比样式特征更大时,优化会使图像结构保持完整。虽然这是不能保证的,因为它取决于优化器,但我们凭经验发现它对于典型用于神经类型转换的 L-BFGS 方法是有效的。

## 参考文献:

1. Gatys, L. A., Ecker, A. S., and Bethge, M. Image Style Transfer Using Convolutional Neural Networks. In *CVPR,2016*. 018
2. J. Johnson, A. Alahi, and F. Li.,Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *CVPR,2017*.
3. Gatys, L. A., Ecker, A. S., Bethge, M.and Shechtman,E.Controlling Perceptual Factors in Neural Style Transfer,In *CVPR,2017*.