基于卷积神经网络的图像风格迁移技术文献综述

作者：艾孜尔江·艾尔斯兰
北京工业大学软件学院

摘　要：2015 年之前图像风格建模技术多采用人工建模模拟图像风格，深度学习以其快速提取大量数据特征的独特优势在图像风格迁移技术方面广泛应用。本文简介了图像风格迁移的历史，并详细介绍了基于卷积神经网络的图像风格迁移技术的两种著名算法，对其发展和应用进行叙述，并概述发展前沿。
关键词：图像风格迁移；卷积神经网络；深度学习；纹理合成；计算机视觉

0 引言

图像风格迁移（Neural Style Transfer）是计算机视觉领域的新课题，输入具有内容特征的图片 A 和具有风格特征的图片 B，然后结合生成具有 A 图片内容和 B 图片风格的图片 C，达到图像风格迁移的结果。该技术通常运用在艺术风格的学习和图片生成领域。下图将黄公望的《富山春居图》作为风格输入，长城实景照片作为内容输入，最终生成兼具两者特征的水墨长城[1]。

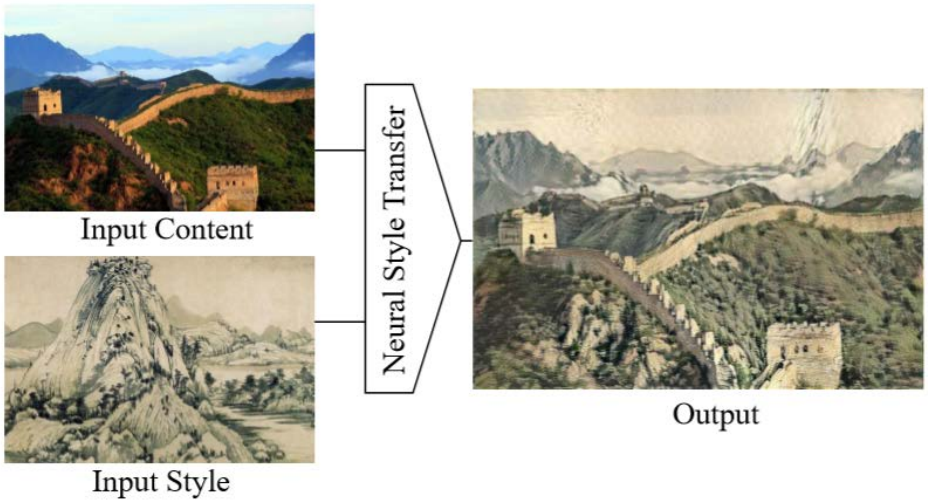

图 1 文献[1]的图像风格迁移效果

早期图片纹理生成主要是手工建模，比如基于统计分布的参数化纹理建模方法（Parametric Texture Modelling with Summary Statistics）或者非参数化纹理建模方法（Non-parametric Texture Modelling with MRFs），手工建模耗时耗力，依赖于建模者的经验和机器性能，并且具有一定的局限性。在深度学习兴起的时候，Gatys 开创性地将卷积神经网络运用到人工艺术图像合成方面，改变原有的技术环境，并引起了学术界和工业界的关注。 本文将从基于卷积神经网络的图像风格迁移技术的优势，研究方法，研究现状以及研究趋势展开叙述。

1 卷积神经网络在图像风格迁移方面的优势

卷积神经网络（Convolutional Neural Networks, CNN）是深度学习（deep learning）的代表性算法之一。卷积神经网络由卷积层，池化层，全连接层三个部分组成，卷积层由多个卷积核用以识别不同的特征，池化层进行下采样，避免数据的过度拟合，降低最终数据量，全连接层输出结果。卷积神经网络通过卷积核提取局部特征，逐层分级进行认知的过程与人类视觉形成过程相近，而且以用来处理自动大量数据，能够识别原有特征，在图像风格迁移中发挥了重要作用。

2 基于卷积神经网络的图像风格迁移的算法

图像风格迁移算法可以分为两大类，一种是基于图像迭代的风格迁移方法，另一种是基于模型迭代的风格迁移方法。前者生成的图现象质量优秀美观，但是生成速度较慢，后者生成图像速度快捷，但模型的灵活性下降。

2.1 基于图像迭代的图像风格迁移方法

Catys 提出的 Neural Style Transfer (NST)[2]方法是神经网络图像风格迁移的基础算法，该方法使用了 19 层 VGG 网络的 16 个卷积层和 5 个池化层，每一层都定义了一个非线性滤波器组，滤波器组中有拥有 N_l 个过滤器，生成 N_l 个大小为 M_l 的特征图，层中响应存储在 $F_l \in R^{N_l \times M_l}$ 中，同时定义两个特征值 P^l_{ij} 和 F^l_{ij} 分别表示第 l 层 i 个滤波器在位置 j 的表示。$\vec{p}$和$\vec{x}$分别为初始图象和初始化的白噪声图像，P^l 和 F^l 分别为原图像的内容特征和白噪声图像的内容特征。

（1）内容损失

为了可视化不同层编码的图像信息，由一张白噪声图出发，已找到和原始图像的特征相匹配的另一图像，实现内容重建，特征值之间的平方误差公式为：

$$\mathcal{L}_{content}(\vec{p},\vec{x},l)=\frac{1}{2}\sum_{i,j}\left(F^l_{ij}-P^l_{ij}\right)^2 . \tag{1}$$

求导得到：

$$\frac{\partial \mathcal{L}_{content}}{\partial F^l_{ij}}=\begin{cases}\left(F^l-P^l\right)_{ij} & \text{if } F^l_{ij}>0\\ 0 & \text{if } F^l_{ij}<0\,.\end{cases} \tag{2}$$

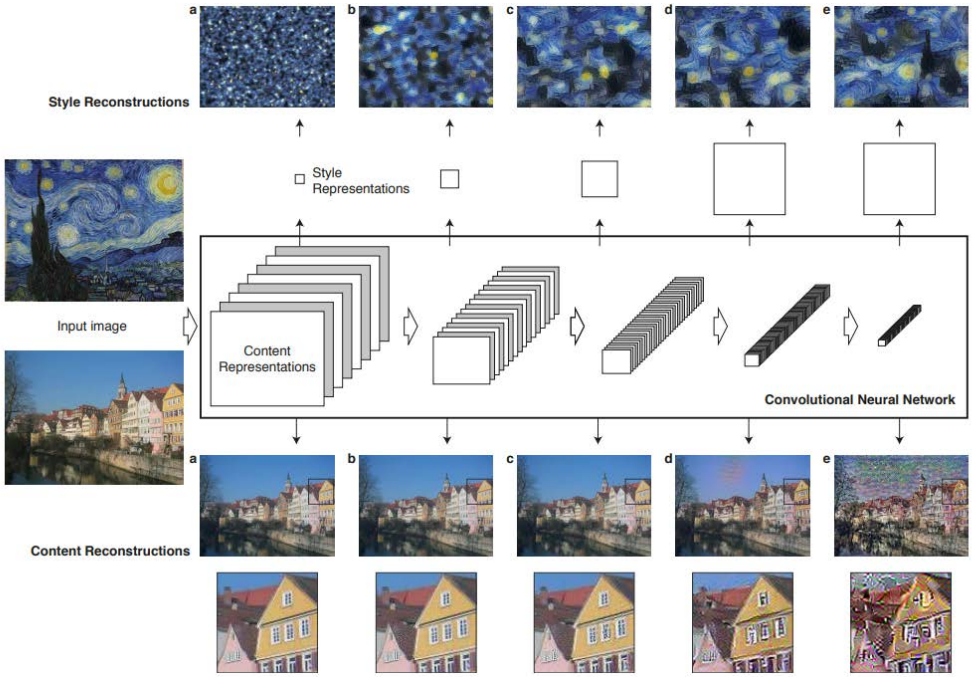

图 2 文献[2]不同层级卷积神经网络结果

Figure2 中结果中给出了一个很重要的结论：随着处理层级越来越高，可以发现在网路较高层中，丢失了详细的像素信息，但是保留了风格特征，网络较低层中较为良好地表达原有的像素信息。

（2）风格损失

Gram 矩阵可以构建在网络任何层之间的滤波器响应上，能计算不同卷积层之间的相关特征性，保证各个特征图之间的相关程度，最终获取其纹理信息。每个层特征图的 Gram 矩阵为：

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}. \tag{3}$$

第 l 层的损失函数为：

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

所有层的损失为：

$$\mathcal{L}_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l \tag{5}$$

E_l 相对于 l 层中的激活的导数是：

$$\frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_l^2 M_l^2} \left((F^l)^{\mathrm{T}} \left(G^l - A^l \right) \right)_{ji} & \text{if } F_{ij}^l > 0 \\ 0 & \text{if } F_{ij}^l < 0 \,. \end{cases} \tag{6}$$

（3）风格迁移

将不同图像的内容特征和风格特征结合在一起，就是风格迁移。其中有$\vec{a}$为风格图像，$\vec{p}$为内容图像，$\vec{x}$为最终生成的图像。

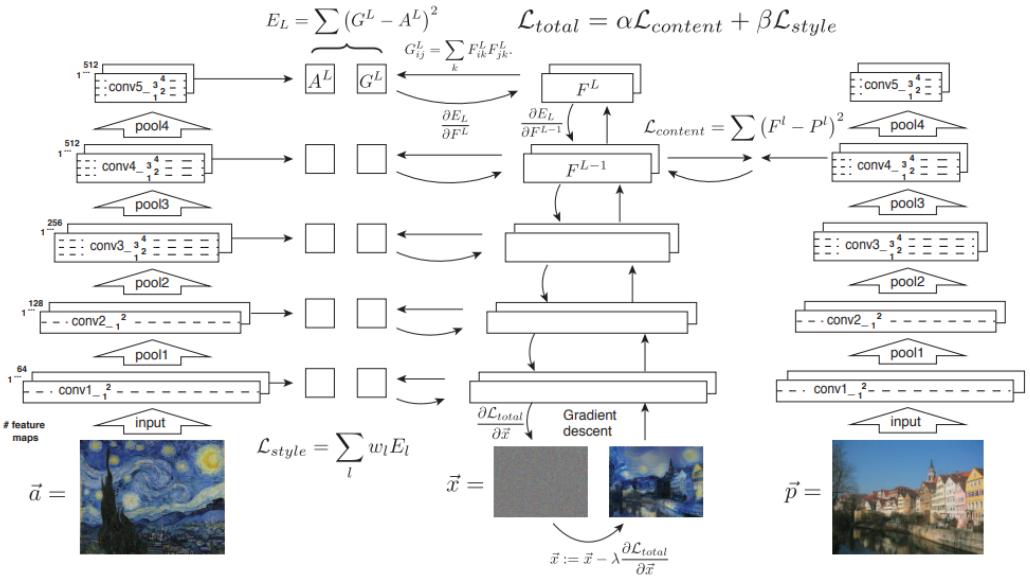

图 3 文献[3]风格转换流程

通过最小化随机图像内容和绘画风格的距离来实现风格迁移。其中 α 和 β 用来调节风格特征和内容特征的损失函数的权重，最小化损失函数是:

$$\mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x}) \tag{7}$$

图 4 文献[3]不同风格特征和内容特征的损失函数的权重的影响

根据采用的风格损失函数不同，这一方法还能细分为，基于最大平均值差异（MMD）的方法，基于马尔科夫随机场（MRF）的方法

2.2 基于模型迭代的图像风格迁移方法

基于图像迭代的图像风格迁移方法效率比较低，基于模型迭代的方法（又称快速图像风格迁移）[4]改进了这一点。图像风格转换通常通过逐像素损失生成，高质量的图像可以通过建立感知损失函数，图像通过使损失函数最小化来生成，结合两者优势通过训练一个用于图像转换任务的前馈网络，使用感知损失函数，从预训练好的网络中提取高级特征。达到图像风格化和超分辨率重建的需求。

快速风格迁移的网络结构包括两个部分，一个是图像转换网络，另外一个是损失网络。图中左边为转换网络，右边为损失网络。

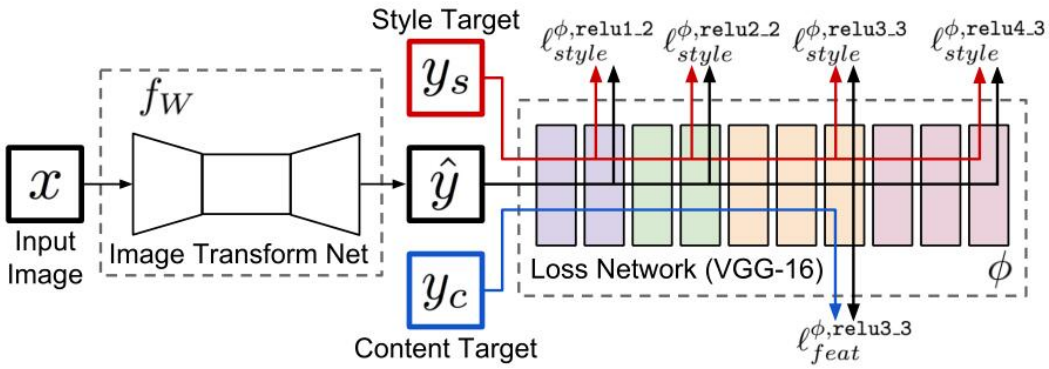

图 5 文献[4]快速网络风格迁移的网络结构

图片转换网络是一个深度残差网络，参数是权重 W，它把输入的图片 x 通过映射 $y=f_w(x)$ 转换成输出图片 y，每一个损失函数计算一个标量值 $l_i(y,y_i)$，衡量输出的 y 和目标图像 y_i 之间的差距。图像转换网络通过 SGD 训练，获取总损失函数：

$$W^* = \arg\min_W \mathbf{E}_{x,\{y_i\}} \left[\sum_{i=1} \lambda_i \ell_i(f_W(x), y_i) \right] \tag{1}$$

(1) 内容损失

使用 VGG 算法提取内容特征，而非逐像素计算。

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \tag{2}$$

(2) 风格损失

定义 Gram 矩阵获取图像的颜色和细节等风格特征。

$$\ell_{style}^{\phi,j}(\hat{y}, y) = \|G_j^{\phi}(\hat{y}) - G_j^{\phi}(y)\|_F^2. \tag{3}$$

(3) 像素损失

像素损失是输出图和目标图之间标准化的差距，它只能被使用在有完全确定目标的图像匹配上。

$$Lpixel(y,y) = \|y^{\wedge}-y\|_2^2/CHW \tag{4}$$

(4) 风格迁移

网络层次越高的生成图像保留越多的风格特征丢失空间结构，为了从多层网络中进行风格重建，根据 NST 算法， 多层网络损失总和应当有下列形式。

$$Loss_{total} = \gamma_1 l_{feat} + \gamma_2 l_{style} \tag{5}$$

综合像素损失，获取最终结果 λ 开头的都是衡量不同损失权重的参数，y 为最终生成的图像。

$$\hat{y} = \arg\min_y \lambda_c \ell_{feat}^{\phi,j}(y, y_c) + \lambda_s \ell_{style}^{\phi,J}(y, y_s) + \lambda_{TV} \ell_{TV}(y) \tag{6}$$

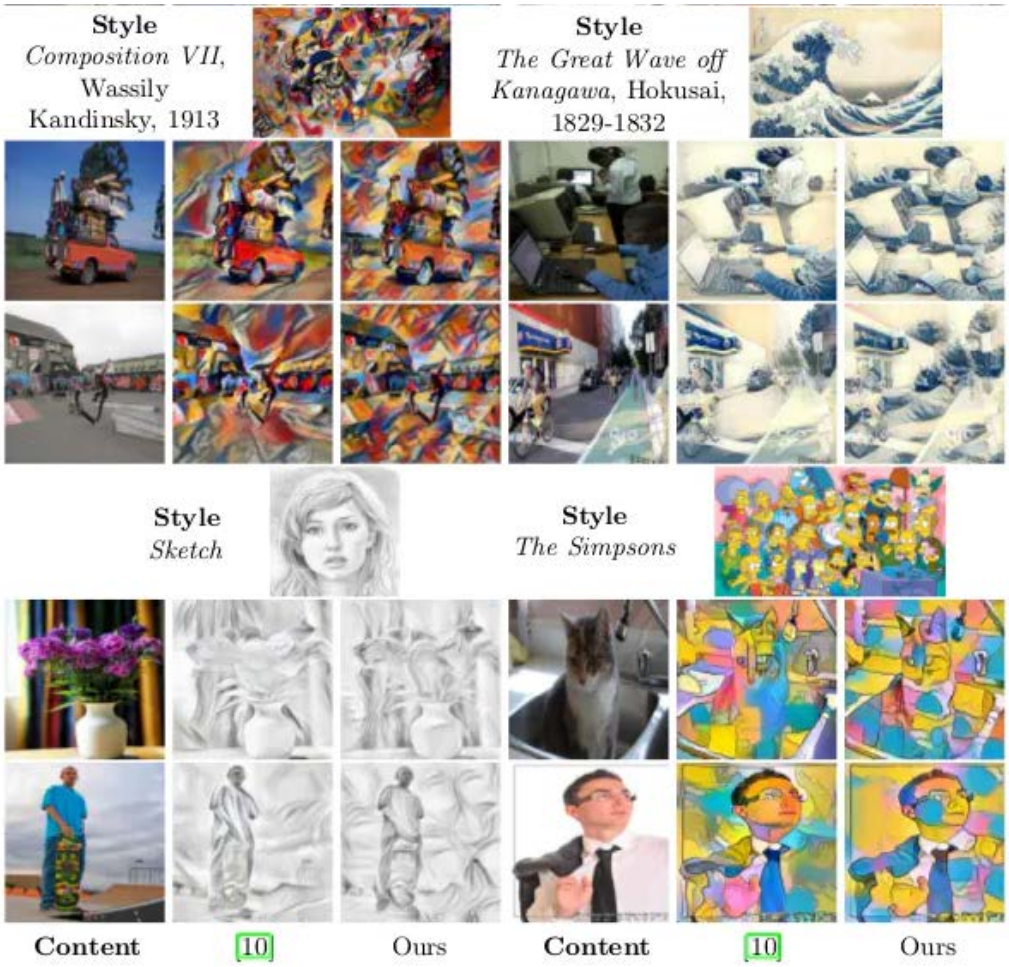

Fig. 6. Example results of style transfer using our image transformation networks. Our results are qualitatively similar to Gatys *et al* [10] but are much faster to generate (see Table 1). All generated images are 256 × 256 pixels.

图 6 文献[4]快速风格迁移算法生成结果

快速风格迁移算法大大转换了风格转换的速度，但是损失了模型的灵活性，一个网络只能变换同一种风格。

3. 改进和拓展

自从 NST 算法出现以来，也有一些研究致力于通过控制感知因素（例如笔触大小控制，空间样式控制和颜色控制）来改进当前的 NST 算法，因此，各种后续研究旨在将一般的 NST 算法扩展到这些特定类型的图像，甚至将其扩展到超出艺术图像样式（例如音频样式）的范围[1]。

3.1 基于笔触的图像风格迁移算法[1]

Jing 等人提出了一种行程可控的 PSPM 算法。他们算法的核心组成部分是 Stroke Pyramid 模块，该模块通过自适应接收场学习不同的笔画大小。在不折衷质量和速度的情况下，该算法利用单一模型来实现灵活的连续笔画大小控制，同时保持笔画一致性，并进一步实现空间笔画大小控制以产生新的艺术效果。 Gatys 等人在提出的算法，该算法是从粗到

细的样式化程序。这个想法是利用一个包含多个子网络的多模型。每个子网络都接收前一个子网络的上采样风格化结果作为输入，并用细笔画对其进行再次风格化。

3.2 基于深度的图像风格迁移算法[1]

当前 NST 算法的另一个局限性在于它们不考虑图像中包含的深度信息。为了解决这个局限性，P. L. Rosin 等人提出了一种深度保留的 NST 算法。该方法基于增加了深度损失函数，用于测量内容图像和风格化图像之间的深度差。图像深度是通过应用单图像深度估计算法获得的。为了加快语义风格化过程，Lu 等人 [68] 提出在特征空间而不是像素空间中优化目标。他们首先通过预先训练的 VGG 编码器转发内容和样式图像，以获得相应的内容和样式特征。之后，使用内容和样式的分割蒙版将获得的内容和样式特征划分为不同的区域。然后在每个对应区域内分别重建样式化图像的特征。最后，将不同区域中的重构特征组合并转发到解码器中，以获得语义风格化结果。

3.3 基于属性的风格迁移算法[1]

图像属性通常指的是图像的颜色，纹理等。以前，图像属性的传递是通过有监督的方式通过图像类比完成的， Liao 提出了一个深层图像类比来研究 CNN features 领域中的图像类比。他们的算法基于补丁匹配技术，实现了一个弱监督图像类比，即他们的算法只需要一对源图像和目标图像而不是大型训练集。

4. 风格迁移技术应用

随着算法的不断改进， 图像风格迁移逐渐出现在生活场景和商业场景中，目前主要的运用方向是以下三种:

（1）设计及艺术创作辅助

艺术家创作的图像具有个人，而且丰富多样，很难模仿。图像风格迁移可以作为画家和设计师的创作工具，便于其更方便地创作特定风格的作品，或者完成相应的 CAD 设计服装设计。对于各国字体设计师来说，将每个字体进行绘制载入字库是一件需要付出劳动力的事情，风格迁移技术在这个应用领域也能得到莫大裨益。

（2）图片处理色彩

Prisma 是一款大众皆知的滤镜 app，它能够轻松将照片转变为具有艺术风格的图画，深受好评，并具有一定的商业价值。普通人能够在一些基于深度学习的图像风格迁移软件的技术支持下创造出属于自己特色的风格照片。

（3）视频处理

影视特效和动画合成方面涉及到大量图像处理的工作，不仅有一定技术要求，而且人工和硬件成本都很高，图像风格迁移技术可用于将真人视频自动样式化为特定样式，简化制作流程，降低成本，并且可以在动画制作和社交上获得好评。

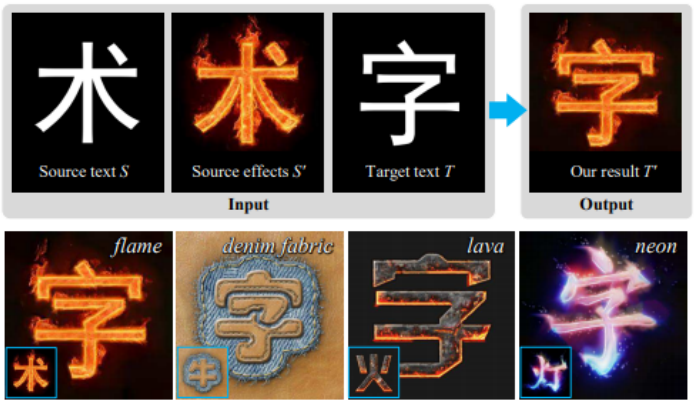

Figure 1. Overview: Our method takes as input the source text image S, its counterpart stylized image S' and the target text image T, then automatically generates the target stylized image T' with the special effects as in S'.

图 7 文献[5]字体风格迁移训练结果

4. 总结

深度学习算法几乎在所有计算机视觉相关领域发挥着作用，甚至在一些问题上有着不可替代的地位，图像风格迁移技术可以将不同图片的图像内容和图像风格结合在一起，在生活和艺术创造领域有着广阔的使用空间，同时这也会面临着审美价值的评估。

在技术层面，基于卷积神经网络的图像风格迁移技术面临着一些指标上的评估与考验：不同大小的单个图像生成时间；单个模型的训练时间；内容图像的平均损失；训练期间的损失变化；样式可伸缩性等等，同时还能拓展应用空间。

总而言之，这是一个新兴的课题，获得了学术界和工业界广泛关注和技术探索，并取得的一定成功，在未来还有着不断取得突破的前景。

[1] Yongcheng Jing, Yezhou Yang, Member, ect.Neural Style Transfer -A Review. IEEE,2019

[2] Gatys.Image Style Transfer Using Convolutional Neural Networks, 2015

[3] Gatys L A, Ecker A S, Bethge M,ect. A neural algorithm of artistic style [J]. arXiv preprint arXiv. 2015: 1508. 06576.

[4] Justin Johnson, Alexandre Alahi, Li Fei-Fei,ect. Perceptual Losses for Real-Time Style Transfer and Super-Resolution arXiv:1603.08155v1 [cs.CV] 27 Mar 20

[5] Samaneh Azadi1* , Matthew Fisher2 , Vladimir Kim2. Johnson, etc. Multi-Content GAN for Few-Shot Font Style Transfer arXiv:1712.00516v1 [cs.CV] 1 Dec 2017

[6] Y. Jing, Y. Liu, Y. Yang, Z. Feng, Y. Yu, D. Tao, and M. Song, “Stroke controllable fast style transfer with adaptive receptive fields,” in ECCV, 2018.