

基于卷积神经网络的图像风格迁移技术应用

摘要

为了提高作品档次或增加作品趣味性，人们大都希望自己创作的图像作品具有大师级人物图像作品的风格，或相同图像内容具有不同图像风格。人们对美好事物的向往也是 IT 技术人员的奋斗目标，利用信息技术帮助普通人实现图像风格迁移的梦想是件很有意义的事情。因此，设计一个基于卷积神经网络的图像风格快速迁移方法应用。首先先使用 VGG 网络对图像内容和风格特征提取，并且用深度残差网络对提取的内容和风格特征进行重构来生成迁移图像，然后运用生成式对抗网络思想，来迭代训练一个可以快速风格迁移系统模型，这样可以快速高效的对大批量图像进行风格迁移；最后设计一个演示系统，根据我们训练好的不同风格模型，让用户可以通过网页上传自己要进行迁移的图片，然后选择不同的风格，后台加载对应风格的模型进行迁移，最后将迁移后的图像返回到 web 网页，供用户浏览和下载保存。

实验结果显示，该方法能较好的完成图像风格迁移工作，系统生成的迁移图像质量很好的融合内容图像的主体内容和风格图像的风格特征。

关键字：图像风格迁移；VGG 网络；生成式对抗网络；深度残差网络

第 1 章 绪论

1.1 背景、目的及意义

在艺术领域，特别是绘画，影视，画家及导演们通过创造特殊的内容及风格，给人们带来一种前所未有的视觉冲击。人类对美的在追求永无止境，尤其是对自己创作的图像及绘画。然而对于普通大众来说，想要自己的绘画或者图片更具有趣味性只能借助 PS 技术。在艺术家和普通大众两个主体面前，图像风格迁移技术亟待突飞猛进的发展来满足大家对艺术化的追求。而深度学习的兴起，尤其是卷积神经网络的出现，让计算机对图像的识别及提取内容技术上上了一个台阶。2015 年 Gatys 初次将深度学习和神经网络运用到了图像的风格艺术迁移技术上来[6] 然而 Gatys 所运用的基于迭代图像的方法效率低下且操作复杂，对于没有专业知识的普通用户来说很难实现图像风格迁移工作。对于当今快节奏生活的大众，如何解决这效率低下且复杂的问题呢？本系统正好补足效率低下的缺点，并加以迁移系统供用户操作。该系统在后台对模型进行训练，生成已经成熟的模型，在网页上供用户选择。

通过风格迁移技术，可以对图片进行处理，使图片效果展示的更加完美，达到对图像艺术预期的要求。对于此次课题研究，目的是为了让大众摆脱复杂的修图技术，让没有深厚绘画功底的用户轻松地得到自己想要的风格图片。针对这一技术的实现，本系统首先使用 VGG (Visual Geometry Group) 网络对图像内容和风格特征提取，并且用深度残差网络对提取的内容和风格特征进行重构来生成迁移图像，然后运用生成式对抗网络思想，来迭代训练一个可以快速风格迁移系统模型，这样可以快速高效的对大批量图像进行风格迁移；最后加入一个演示系统，这样就能使用户简单快速地对自己喜欢的风格图像进行迁移。

通过卷积神经网络等算法设计实现一个图像风格快速迁移系统。通过这个系统，可以让用户简单快速地对自己的图像进行风格迁移，从而达到用户对图像艺术化的追求。本文设计的快速风格迁移系统，在绘画方面也有很大的作用：可以在绘画作品中填入多元的艺术风格，从而使绘画作品更加完美。在影视作品方面，本设计为影视作品风格迁移打下基础，在今后的学习研究中可以

在本设计的基础上完成影视作品的风格迁移技术。

1.2 相关工作

目前的风格迁移技术，包括物理建模提取纹理、基于图像迭代方法及基于模型迭代三种方法^[14]。

1.2.1 基于物理模型的绘制和纹理的合成方法

基于物理模型的绘制和纹理的合成实现图像风格迁移是一种传统非参数的迁移方法^[14]。主要通过对样本纹理进行拼接和重组来合成新的纹理。在图像处理技术上，把纹理当做一个图片的风格，在后续的图像风格迁移技术上都是对图像的纹理特征进行提取。这种非参数的提取方法也只能提取内容与风格图片底层的纹理，对于颜色、内容复杂的图像合成起来就显得粗糙。

1.2.1 基于图像迭代的方法

基于图像迭代的主要做法是直接在白噪声图像进行迭代优化实现风格迁移，这种做法的优化目标是优化白噪声图像^[6]。这类做法的好处是生成图像的品质高，可控性好，比较容易调节参数，而且不需要大量的训练数据；然而也同样存在一些问题，其计算时间较长，耗费时间。对于大批量的图像进行风格迁移过程中，效率低下，而且对于预训练的图像特征提取模型依赖性大。

1.2.2 基于迭代神经网络模型的方法

基于迭代神经网络模型，以神经网络的前馈方式实现快速风格迁移（fast style transfer），其优化目标是神经网络^[14]。这类做法的优点是，工作速度快，同样可以运用到影像的风格迁移上；同样这类做法的缺点就是：合成图片的品质有待提高。而且在优化神经网络的过程中需要大批的训练数据。

1.3 本文工作

1.3.1 本文主要工作

为解决图像风格迁移技术效率低下的缺点，本设计主要的研究方向是最大程度的减少模型训练次数，因此本系统采用模型迭代方法，结合对抗式生成网络思想，设计一个基于 VGG-19 网络模型的判别器来提取风格图像的艺术特征，用残差神经网络作为生成器来生成迁移图像的模型，从而实现一次训练就可以多次生成迁移图像的快速图像风格迁移。实验结果显示，使用模型迭代方法可以快速生成不同风格的图像，并且能够根据自己喜欢的风格训练图片，从而

达到用户对图像艺术的要求。

1.3.2 论文结构安排

本章将介绍论文中各章节的工作安排。

第一章 绪论。本章主要对图像风格迁移技术背景、目的及意义进行介绍，然后讲述国内外图像迁移技术的现状，最后介绍该项目的主要工作及论文的结构安排。

第二章 任务描述及数据与处理。本章主要介绍本项目的主要工作及数据预处理。首先对本项目所用技术进行介绍，然后提出在完成项目过程中要解决的问题，并找出解决问题的方法。最后介绍在模型迭代的过程中，需要大量的训练数据来训练模型，介绍如何选取训练数据，以及对训练数据的预处理。

第三章 基于卷积神经网络的图像风格迁移技术。主要对本研究中将用的相关技术进行介绍，主要包括卷积神经网络、VGG 网络、深度残差网络、生成式对抗网络。VGG 网络是用牛津大学视觉几何小组预训练的网络模型，该网络在图像的特征提取方面取得了很好的效果。深度残差网络是在 CNN 的基础上加深网络的深度，从而提高网络的性能，用来构建图像生成网络。生成式对抗网络思想，主要用来解决图像迁移效率低下的问题，采用 GANs 思想，从而可以一次训练网络，然后保存模型参数，从而实现快速风格迁移效果。

第四章 图像风格迁移技术应用。本章主要介绍应用系统的结构设计及系统演示结果。对 web 演示系统的结构设计过程详细讲解并展示。

最后是对本文所做工作进行总结。主要对本文的工作和本设计的不足之处进行总结，以及对未来图像风格迁移的发展进行展望。

第 2 章 任务描述与数据预处理

本章节详细描述本文需要完成的任务和对本文建立的模型进行数据预处理的详细过程。

2.1 任务描述

利用信息技术来处理图像来满足人们对图像艺术化的追求是一件非常有意义的工作。本文在图像风格艺术迁移技术的研究与应用的基础上，设计并实现了一个可交互的内容与风格分离的图像风格艺术迁移系统，完成了以下几方面的任务。

- 1) 构建 VGG-19 模型并训练改进。本项目所设计的是图像风格艺术化迁移，通过对内容图片和风格图片的特征进行深度提取，组合成一个新的目标图片。在图像的特征提取方面，VGGNet 模型做的非常好。首先下载 VGG-19 预训练模型 imagenet-vgg-verydeep-19 文件，由于本项目不用于图像分类，将 VGG-19 模型进行部分改进：去掉三个用于图像分类的全连接层，使参数大大减少，提高工作效率。在模型训练过程中，我们对内容特征提取与风格特征提取定义损失函数，计算特征提取时的总损失。
- 2) 基于深度残差网络的图像重构技术。用 VGG-19 模型对图片的内容和风格特征进行提取，用残差神经网络作为生成器来生成迁移图像，并且不断迭代该模型，优化模型权重参数，最后生成一个高质量的迁移图像。
- 3) 演示系统的建立。采用本项目实现的图像风格迁移模型，使用 Django 框架来设计一个应用系统。

2.2 数据预处理

在进行系统模型训练的过程中，使用的是 MS COCO 数据集，在 COCO 数据集中每张图片都是固定的 RGB 三通道的像素值，但是图片的大小形状是不一样的，而且对于 Tensorflow 框架设计的计算图，对于传入的数据为固定相同大小的向量值。所以，首先针对 COCO 数据集内的图片先进行数据预处理，将图片转换为相同大小形状的向量值。

首先借助 python 的第三方工具包 numpy 和 scipy 集成科学数据工具包，使用 scipy 包直接将图片以 RGB 三通道方式读取为像素值数组，然后使用 numpy 工具，将像素值的数组转化为固定大小的 shape (255x255x3)。同时传统的图像数

据预处理过程,需要将图片像素值进行数据归约,就是将图像的像素值归约到 0-1 之间,简称二值化处理,这样做的意义是在训练过程中可以更好的模型损失下降,但是这里不做二值化处理,因为最终是要生成图像,所以要保留图片的原像素值,这样在生成网络输出的特征,可以直接来通过 `scipy` 保存为图像。但是,与此同时损失误差值将会很大,这样对于网络的性能提升有限制性。

第 3 章 基于卷积神经网络的图像风格迁移技术

3.1 基础知识

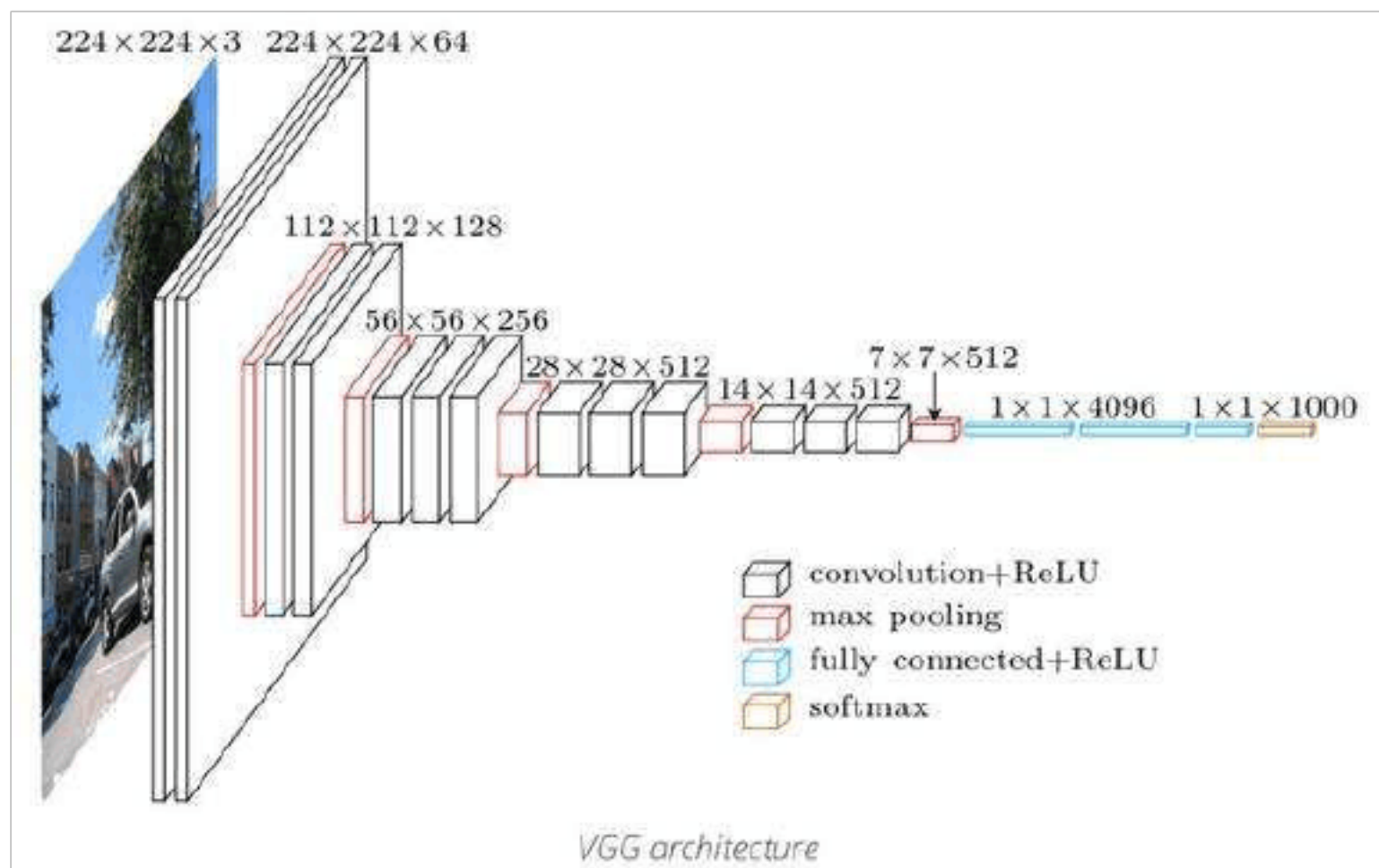
本设计的主要框架，使用对抗式生成网络思想，使用模型迭代，设计一个基于 VGG-19 网络模型的判别器来提取风格图像的艺术特征，并且判断生成图片的迁移质量，用残差神经网络作为生成器来生成迁移图像，并且不断迭代该模型，优化模型权重参数，最后生成一个高质量的迁移图像。这样做的目的就是不需要预先生成白噪图像作为输入，而是直接以内容图像 (content image) 作为输入，用深度残差网络来生成迁移图像，以 VGG 网络为基础的图像判别器，来评价生成图像的质量。接下来重点介绍框架中需要的相关技术准备。

3.1.1 VGG 网络

卷积神经网络 (CNN) 最初设计被用来做图片分类工作，最近也被用来做图片分割，图像风格迁移以及其他计算机视觉的工作；当然，也有被用来做自然语言处理 (NLP) 的例子。卷积神经网络是最能解释深度学习的模型之一，因为可以将它的表达特征可视化来理解它可能学习到的东西。

VGG 是 Oxford 的 Visual Geometry Group 的组提出的。该网络是在 ILSVRC 2014 上的首次被提出，主要工作是证明了增加网络的深度能够在一定程度上影响网络最终的性能。VGG 有两种结构，分别是 VGG16 和 VGG19，两者并没有本质上的区别，只是网络深度不一样，结构如图 3-1。

在计算机图像处理卷积神经网络中，VGG 网络是本人最喜欢使用的模型，并且认为是最好的模型。VGG16 相比 AlexNet 的一个改进是采用连续的几个 3x3 的卷积核代替 AlexNet 中的较大卷积核 (11x11, 7x7, 5x5)。对于给定的感受野 (与输出有关的输入图片的局部大小)，采用堆积的小卷积核是优于采用大的卷积核，因为多层非线性层可以增加网络深度来保证学习更复杂的模式，而且代价还比较小 (参数更少)：验证了通过不断加深网络结构可以提升网络整体性能。但同时，VGG 耗费更多计算资源，并且使用了更多的参数，其中绝大多数的参数都是来自于第一个全连接层，并且，VGG 网络有 3 个全连接层！这样使得模型参数权重达到 550M 左右。

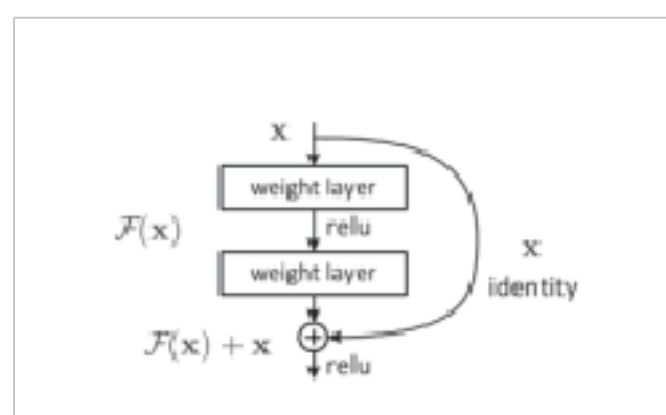


3-1 VGG 网络参数结构图

在本设计中，主要使用 VGG-19 网络来对图片进行特征提取，分别三次使用 VGG-19 网络来提取图片特征，分别是 content image、style image 和 output image（即系统最终迭代输出的迁移图线），然后分解计算 output image 的内容和风格与 content image 和 style image 之间的误差（Loss），然后将这两个误差进行计算反馈给图像生成网络，来优化生成网络的权重参数。

3.1.2 深度残差网络

残差网络是由来自 Microsoft Research 的 4 位学者提出的卷积神经网络，如图 3-2，在 2015 年的 ImageNet 大规模视觉识别竞赛（ILSVRC）中获得了图像分类和物体识别的优胜。残差网络的特点是容易优化，并且能够通过增加相当的深度来提高准确率。其内部的残差块使用了跳跃连接，缓解了在深度神经网络中增加深度带来的梯度消失问题。残差的思想就是去掉相同的主体部分，从而突出微小的变化。



如图 3-2 深度残差网络结构图

本设计使用深度残差网络对提取的内容和风格特征进行重构来生成迁移图像。

3.1.3 生成式对抗网络

生成式对抗网络（GANs, Generative Adversarial Networks）是一种深度学习模型[13] 如图 3-3，是近年来复杂分布上无监督学习最具前景的方法之一。模型通过框架中（至少）两个模块：生成模型（Generative Model）和判别模型（Discriminative Model）的互相博弈学习产生相当好的输出。原始 GAN 理论中，并不要求 G 和 D 都是神经网络，只需要是能拟合相应生成和判别的函数即可。但实用中一般均使用深度神经网络作为 G 和 D。

目前，生成式对抗网络主要的应用前景有两个方向：一个是图像生成，就是本设计用来生成迁移图像，同样还有超分辨任务、语义分割等应用；另一个方向就是用来数据增强，主要是针对一些小数据集数量不做，可以用原数据集来生成图像，增加数据数量。

在 Gatys 的实验中，每次需要将内容图片和风格图片多次迭代训练以达到比较好的风格迁移，效率比较低小。所以使用 GANs 思想，建立一个可以高效生成迁移图像的框架，即一次模型训练，永久图像生成。

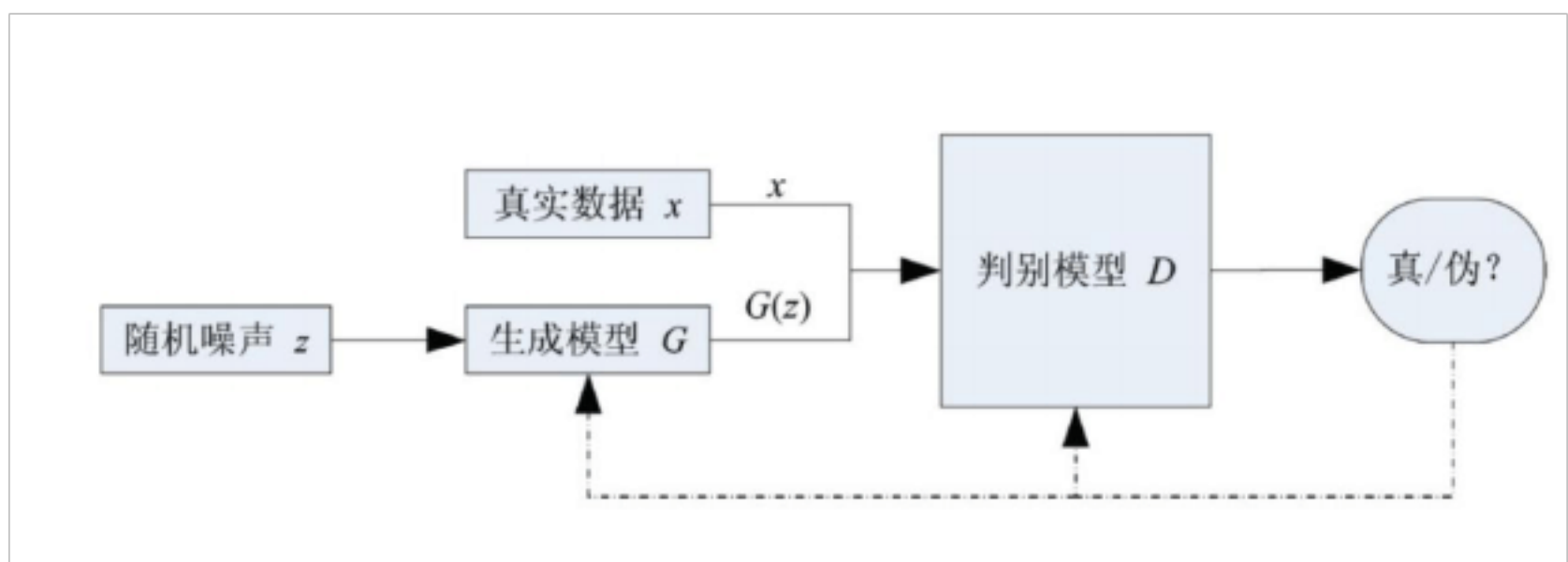


图 3-3 生成式对抗网络结构

本设计使用深度残差网络思想完成对模型的迭代训练，这样就可以得到一个快速的风格迁移系统。

3.2 模型框架设计

3.2.1 模型框架结构

上一章节已经介绍了 VGG 网络、残差网络、生成式对抗网络。所以接下来要设计出能够一次训练过程，并且保存模型，然后可以重复生成对应风格的迁移图像。系统的框架图如图 3-4:

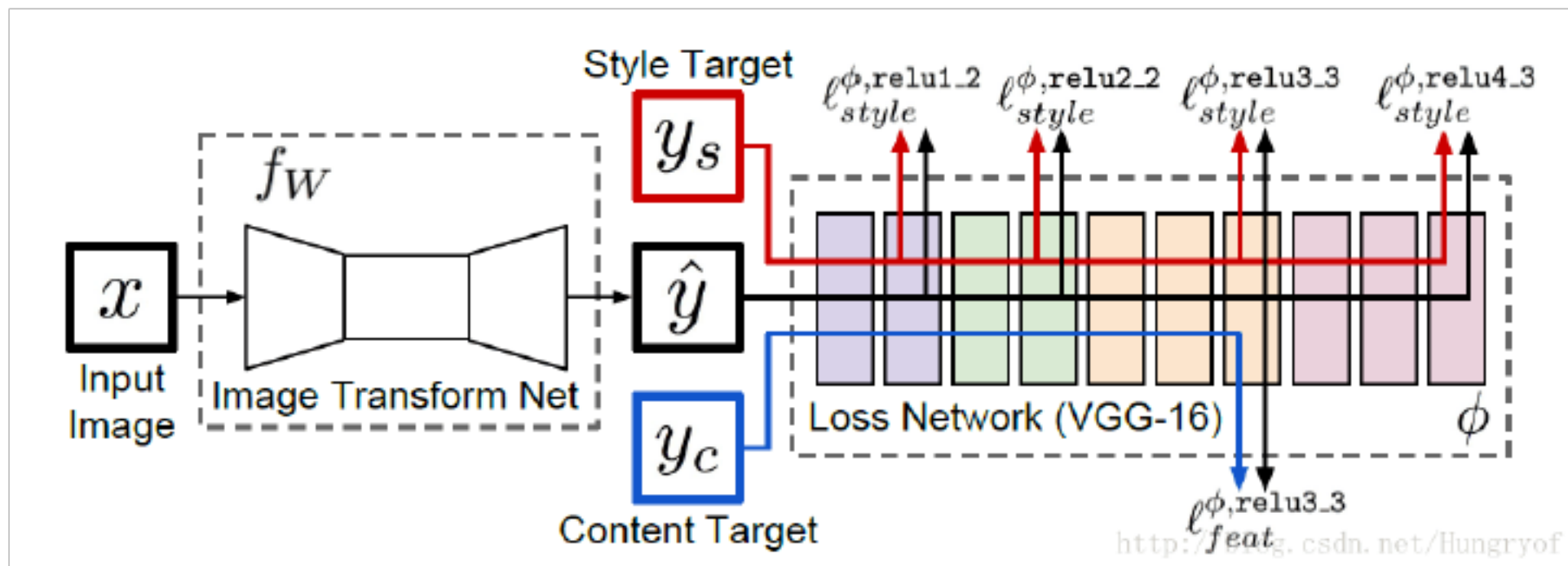


图 3-4 图像风格迁移框架

系统整体思想是生成式对抗网络 (Generative Adversarial Nets, GANs), 可分为两个模块: 图片生成网络 (Image Transform Net) 和损失判别网络 (Loss Network)。

图片生成网络结构即为上边模型框架图的左边部分, 工作流程如下:

x 是内容图片 (content image) 作为输入, 传入项目设计的神经网络中去, 网络的输入层形状和与输入图片形状相同, 然后是三层卷积操作, 也就是下采样过程, 并且使用 ReLU 函数激活, 此时特征图形状变小, 但是通道数增加; 然后是五层的残差网络结构, 这里使用残差网络的左右就是为了增加网络的深度, 以此来提高图片的生成质量。接着是三层的反卷积操作, 就是上采样过程, 并且使用 ReLU 函数激活, 反卷积的作用就是让网络能够在提取到卷积特征的同时, 最终保证输出图像的大小的输入图像保持一致。

损失判别网络结构即为上边模型框架图的右半部分, 工作流程如下:

损失判别网络的结构的作用就是让生成网络输出的图片 output image 经过不断的迭代训练, 在内容上越来越接近 content image, 在风格上越来越接近 style image, 具体实现的过程是将生成网络 output image、content

image 和 style image 同时传入 VGG 网络中，然后分别计算对应的 Content Loss 和 Style Loss。

在分别计算出 Content Loss 和 Style Loss 之后，接下来要计算出系统的整体误差，并且用 α 和 ρ 两个超参数来调节 Content Loss 和 Style Loss 的权重比例，即这两个超参数可以简单地认为是两个调节器来控制的 output image 要继承 content image/style image 的比例。然后将系统整体误差通过 MLP 前馈到生成网络中去，然后生成网络根据这个误差，不断更新迭代每层网络的权重参数。系统的整体误差公式如下：

$$l_{total}(a, x, p) = \alpha * l_{style}(a, x) + \beta * l_{content}(p, x)$$

其中 x 为目标输出图片， a 为风格图片 (style image)， p 为内容图片 (content image)， α 和 ρ 为超参数。

系统在经过迭代训练之后，这样可以得到图像生成网络的权重参数模型，将该参数模型保存下来。在以后的训练过程中，只需要将图像生成网络加载之前保存下来的模型参数，就可以生成出对应该模型的风格图像。

系统框架结构如上所述，与传统的深度学习一样，本系统也需要训练，但是不同的是本系统不需要验证，因为暂时没有任何的公式和算法能够来评价两幅图像的艺术风格相似度。评价系统训练结束后模型的性能优劣，就是根据系统生成的迁移之后的图像，自身来判断输出图像的图片内容与风格与 content image 和 style image 的相近程度。

3.2.2 内容与风格的重构技术原理

(1) 图像内容重构技术

根据 VGG 网络模型图，对 VGG 模型结构能够这样理解：前面部分的卷积层是从内容图像中提取特征，而全连接层把图片的特征转换成了类别概率。其中，VGG 网络中的浅层（如 conv1_1, conv1_2），提取的特征往往是比较简单的（如检测点、线、亮度），VGG 网络中的深层（如 conv5_1、conv5_2），提取的特征往往是比较复杂（如有无人脸或某种特定物体）。VGG 网络的本意是输入图像，提取特征，并输出图像类别。图像风格迁移正好与其相反，输入特征，输出对应这种特征的图片。所以这里将内容图片传入 VGG 网络中去，经过卷积之

后，抛弃网络浅层次中的部分颜色和纹理信息，将网络深层次中大体保留原始图像中物体的形状和位置信息保留。将这些保留的特征用来生成输出图像，如何让输出图像尽可能地和内容图像保持相似呢？本设计中在图片判别网络中加入了图像内容损失函数来解决这个问题。这里在本系统中建立内容损失函数 (Content Loss)，如公式 (3-1)、(3-2)：

$$D_C^L(X, Y) = \|F_{XL} - F_{YL}\|_2^2 = \sum_i (F_{XL}(i) - F_{YL}(i))^2 \quad (3-1)$$

$$F_{XL} \in VGG(X) \quad (3-2)$$

其中， X 为输出目标图片， Y 为输入内容图片； L 为卷积层数，即为VGG网络的深层； F_{XL} 为在VGG网络中第 L 层的特征图。损失函数即为计算输出目标图片和原始内容图片在深层VGG网络中的卷积特征向量的欧几里得距离，如果 $D_C^L(X, Y)$ 越小，则输出图像在内容上就和原始内容图像更加相似，得到的结果就越好。

(2) 图像风格重构技术

Gatys 利用神经网络进行图像艺术风格迁移，把这个过程称之为 neural style，在此之前，Gatys 还做了工作，在发表文章《Texture Synthesis Using Convolutional Neural Networks》[7]中，他们发现如果让神经网络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。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。比如对一张人为主体的图片，传入到神经网络中去，然后提取隐藏层的特征，用协方差进行约束之后，重新生成图片，结果发现人还是人，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，那么如果加上“让生成图的隐藏层特征与内容图片尽量一样，另一方面让生成图的打散特征与艺术风格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。

Gram 矩阵是如何提取到图片的艺术风格？其实可以这样认为，协方差就是一种二阶统计信息，要求输出图的什么信息与风格图相近，肯定不是特征图

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/455230214312011320>