（封面）

摘 要

图像风格迁移技术通过将任意的内容图片和风格样式图片结合起来，形成新的具有风格样式图片的风格和内容图片内容的新图片。以往的艺术家们所遗留下来的作品展现了其个人独具特色的艺术风格，而人们想要借助其独特的艺术风格将不同的场景变换为其独具特色的艺术作品，以往通常通过人工绘画进行描摹。然而深度学习兴起后，通过图像风格迁移技术即可将不同的内容图片加上各种艺术风格。由于艺术风格众多，神经网络很难快速在其中转换。本文将探讨基于快速风格转换的图像风格迁移技术，该技术以AdaIN层为核心，利用自适应实例规范化层对图片样式进行实时的快速样式转换，并通过与其他方法进行比较总结其优缺点并提出改进方向。

关键词 ：深度神经网络；图像风格转换；自适应实例规范化

**Abstract**

The image style transfer technology forms a new picture with style and content picture content by combining arbitrary content pictures and style picture. The works left by the artists in the past show their unique artistic style, and people want to use their unique artistic style to transform different scenes into unique artistic works. describe. However, after the rise of deep learning, various art styles can be added to different content pictures through image style transfer technology. Due to the numerous artistic styles, it is difficult for neural networks to switch among them quickly. This article will discuss the image style transfer technology based on rapid style conversion. This technology uses the AdaIN layer as the core and uses an adaptive instance normalization layer to perform real-time fast style conversion on the picture style. Improve direction.

Keywords: Deep neural network; image style transform; adaptive instance normalization

目 录

关于南开大学本科生毕业论文（设计）的声明

摘要………………………

Abstract…………………………………

目录………………………………………………

第一章 绪论………………………………

1.1 早期图像艺术风格转换……………………………

1.2 基于神经网络的艺术风格转换……………………

1.3 本文中所采用的方法…………………………

第二章 相关方法综述…………………………………

2.1 卷积神经网络………………………………

2.1.1 输入层…………………………

2.1.2 池化层…………………………

2.1.3 全连接层………………………

2.2 批量标准化…………………………

2.3 实例规范化…………………………

2.4 条件实例规范化…………………………

第三章 基于自适应实例规范化的图像风格变迁…………

3.1 实验原理…………………………………

3.2 环境构建与训练…………………………

第四章

第五章

参考文献…………………………

致谢……………………………………………

毕业论文（设计）题目审批表

毕业论文（设计）中期检查表

毕业论文（设计）指导教师评语及打分表

毕业论文（设计）答辩记录及打分表

“查重”检测结果及对检测结果的认定材料

1. 绪论

（一）早期图像艺术风格转换

艺术风格很难通过概念进行界定，对于这一种无法通过语言或公式进行准确界定的事物，想要进行批量生产显得苦难重重。同样如何要把一个图像的风格变成另一种风格更是难以定义的问题。因此对于执行图像风格转换的机器学习方面的程序员来说，这种模糊的定义根本无法通过程序进行合理有用的表达。到底怎么把一个定义不清的东西变成一个准确的可执行程序，是困扰了很多研究图像风格迁移领域的研究者的问题。

在神经网络之前，图像风格迁移程序有一个共同的思路：分析某一种艺术风格的图像，给该艺术风格建立一个数学或者统计模型，再改变要进行风格迁移的内容图像让它能更好的符合建立的模型。这样做出来效果具有一定视觉上的美观。但该方案具有一个很大的缺点：一个程序基本只能做某一种艺术风格或者某一个特殊场景。因此基于传统风格迁移研究的实际应用非常有限。

（二）基于神经网络的艺术风格转换

Gatys等在其论文中表示，深度神经网络（DNN）不仅能够编码图像内容，还能编码图像的样式信息。而且图像的样式和内容在某种程度上是可分离的，即可以在保留图像内容的同时对其样式进行更换。Gatys提出的方法能够组合任意图像的内容和样式，具有十分灵活的性质，但是其需要不断的迭代优化，导致了该方案的生成速度始终较长。为了加速神经风格的转换，而使用训练前馈神经网络的方案，这些神经网络通过单个前向传递执行样式化的目标。绝大多数前馈神经网络方法的局限性在于每一个神经网络都局限于一种样式，而后一些改进的方案要么依然局限于有限的样式，要么比但样式传输方法更慢。

当前的神经样式转换（NST）方法大致可分为两类[2]。一类是Gatys等利用卷积神经网络（CNN）特征激活来重组给定图片的内容以及样式的方法，该方法说明图像样式和内容在某种程度上是可分离的，可任意组合图像的内容及样式，但它依赖于过慢的迭代图像优化过程，使其计算量十分庞大。另一类是Johnson和Ulyanov[3，4]等人提出的一种使用多尺度体系结构作为生成器网络，目标函数类似于Gatys等人的算法，该算法通过基于模型优化的离线图像重建（MOB-IR）来重建风格化图像，以解决速度和计算成本问题。

样式转换的问题可以分为非真实感渲染和针对样式转移的问题。早期的方法包括对线性滤波器响应进行直方图匹配[7]和非参数采样[8]，这些方法通常依赖于低级统计信息，且通常无法获得语义结构。Gatys等首次通过在CNN卷积层中匹配特征统计信息，得到相对更完美的样式转换结果。他们[1]将样式转换描述为纹理合成和内容重建两者结合在一起的优化问题，公式涉及到将附加损失函数置于预先训练的CNN的多层上，其中一些损失函数用于合成样式图像的纹理，另外一些损失函数用于重新构建内容图像。通过反向传播计算梯度，并基于梯度的优化来构建风格化图像。

另一种方法是在内容和样式图像之间使用基于补丁的相似度匹配，其中Li和Wand 在深度特征空间中引入了基于马尔可夫随机场（MRF）的框架构建了基于补丁的损耗函数，其中每个合成补丁都具有必须匹配的最近邻居目标补丁。然后，将这种补丁匹配损失函数与Gatys等人提出的公式相加。虽然这种方法允许使用任意样式的图像，但是这种方法使用的优化框架使其生成样式化图像的速度变的相对缓慢。

Gatys等提出的框架是基于迭代神经网络的慢速优化过程，该过程会迭代更新图像以最大程度地减少内容损失以及损失网络所计算出的样式损失。因为其迭代过慢，在移动端的应用程序上处理太慢而无法大规模实用。常见的解决办法是使用经过培训的前馈神经网络对优化过程进行替换，来最小化同一目标。这些前馈样式传递方法相比基于优化的替代方案快上了不少。Wang等通过多分辨率档案库提高了前馈样式传输的粒度，然而在将每个网络绑定到固定样式上，上述的前馈方法受到限制。为了解决这个问题lem和Dumoulin等人提出可以引入一个单一网络，该单一网络能够编码32种样式。Li等又提出一种前馈架构，该架构可以合成多达300多种纹理和16种训练样式。但是上述两种方法依然无法适应任意样式的转换。

Chen和Schmidt提出一种借助样式交换层可以转发任意样式的前馈方法。该方法由于具有样式交换层而可以传输任意样式。 对于给定内容和样式图像的特征激活，样式交换层将以逐个补丁的方式用最匹配的样式特征替换内容特征。 但是，它们的样式交换层造成了新的计算瓶颈：95％以上的计算花费在512×512输入图像的样式交换上。

样式转换的另一个主要问题是使用哪种样式损失功能。Gatys等人的原始框架。 [16]通过匹配由Gram矩阵捕获的特征激活之间的二阶统计量来匹配样式。现在已经有其他有效的损失函数被提出来，例如MRF损失[30]，对抗性损失[31]，直方图损失[54]，CORAL损失[41]，MMD损失[33]以及通道平均和方差之间的距离[ 33]等。需要注意的是，以上所有损失函数旨在匹配样式图像和合成图像之间的某些特征统计量。

（三）本文采用的方法

在自适应实例规范化层被提出来之前，已经有相关的批量标准化（BN）、实例规范化（IN）和条件实例规范化（CIN）被提出和研究。批量标准化由Ioffe 和Szegedy首次提出并引入实验，该方案的核心是批量标准化层，该层通过标准化特征统计信息大大简化了前馈神经网络的训练，并且可以用于对生成图像的有效建模。Ulyanov发现如果将样式传输网络中的卷积层中所包含的批量标准化层用实例规范化层进行替代，即可实现显著的速度提升。条件实例规范化层由Dumoulin等人提出，该层针对每一种学习样式分别学习一组不同的参数，然后通过样式传输网络处理内容图像，对于不同的内容图片使用相同的样式参数即可将其转换为相同样式，但是在面对大量不同样式时，参数的提取过程过于漫长，使得其在面对大量数据时无法发挥很好的应用。

本文研究的自适应实例规范化是对实例规范化的一种扩展，该方案将接收内容输入和样式输入，并将内容输入的通道平均和方差对齐以匹配样式输入，并且其没有可以学习的仿射参数，而是通过对样式输入进行自适应计算得到仿射参数。自适应实例规范化通过传递特征统计信息在特征空间中执行样式传递，几乎不增加计算成本。

1. 相关方法综述

（一）卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有出色表现。

卷积神经网络由一个或多个卷积层和顶端的全连通层（对应经典的神经网络）组成，同时也包括关联权重和池化层（pooling layer）。这一结构使得卷积神经网络能够利用输入数据的二维结构。与其他深度学习结构相比，卷积神经网络在图像和语言识别方面能够给出更好的结果。这一模型也可以使用反向传播算法进行训练。相比较其他深度、前馈神经网络，卷积神经网络需要考量的参数更少，使之成为一种颇具吸引力的深度学习结构

1.输入层

卷积神经网络的输入层可以处理多维数据，常见地，一维卷积神经网络的输入层接收一维或二维数组，其中一维数组通常为时间或频谱采样；二维数组可能包含多个通道；二维卷积神经网络的输入层接收二维或三维数组；三维卷积神经网络的输入层接收四维数组。由于卷积神经网络在计算机视觉领域应用较广，因此许多研究在介绍其结构时预先假设了三维输入数据，即平面上的二维像素点和RGB通道。与其它神经网络算法类似，由于使用梯度下降算法进行学习，卷积神经网络的输入特征需要进行标准化处理。具体地，在将学习数据输入卷积神经网络前，需在通道或时间/频率维对输入数据进行归一化，若输入数据为像素，也可将分布于[0，225]的原始像素值归一化至[0，1]区间  。输入特征的标准化有利于提升卷积神经网络的学习效率和表现

2.池化层

池化（Pooling）是卷积神经网络中另一个重要的概念，它实际上是一种非线性形式的降采样。有多种不同形式的非线性池化函数，而其中“最大池化（Max pooling）”是最为常见的。它是将输入的图像划分为若干个矩形区域，对每个子区域输出最大值。

直觉上，这种机制能够有效地原因在于，一个特征的精确位置远不及它相对于其他特征的粗略位置重要。池化层会不断地减小数据的空间大小，因此参数的数量和计算量也会下降，这在一定程度上也控制了[过拟合](https://zh.wikipedia.org/wiki/%E8%BF%87%E6%8B%9F%E5%90%88" \o "过拟合)。通常来说，CNN的网络结构中的卷积层之间都会周期性地插入池化层。池化操作提供了另一种形式的平移不变性。因为卷积核是一种特征发现器，我们通过卷积层可以很容易地发现图像中的各种边缘。但是卷积层发现的特征往往过于精确，我们即使高速连拍拍摄一个物体，照片中的物体的边缘像素位置也不大可能完全一致，通过池化层我们可以降低卷积层对边缘的敏感性。

3.全连接层

卷积神经网络中的全连接层等价于传统前馈神经网络中的隐含层。全连接层位于卷积神经网络隐含层的最后部分，并只向其它全连接层传递信号。特征图在全连接层中会失去空间拓扑结构，被展开为向量并通过激励函数 。

按照表征学习观点，卷积神经网络中的卷积层和池化层能够对输入数据进行特征提取，全连接层的作用则是对提取的特征进行非线性组合以得到输出，即全连接层本身不被期望具有特征提取能力，而是试图利用现有的高阶特征完成学习目标。然后将结果进行输出。

（二）批量标准化

传统的深度神经网络在训练时，随着参数的不断更新，中间每一层输入的数据分布往往会和参数更新之前有较大的差异，导致网络要去不断的适应新的数据分布，进而使得训练变得异常困难，但是却只能使用一个很小的学习速率和精调的初始化参数来解决这个问题。而且这个中间层的深度越大时，这种现象就越明显。由于是对层间数据的分析，也即是内部（internal），因此这种现象叫做内部协变量偏移(internal Covariate Shift)。

为了解决这个问题Ioffe 和Szegedy首次将批量标准化层引入实验中，该方案不仅仅对输入层做标准化处理，还要对每一中间层的输入(激活函数前)做标准化处理，使得输出服从均值为0，方差为1的正态分布，从而避免内部协变量偏移的问题。该层通过标准化特征统计信息大大简化了前馈网络的训练。批量标准化层最初旨在对区分性网络的训练进行加速，经过研究发现其可以用于有效的生成图像建模。给定输入批次x∈RN×C×H×W，批量标准化层标准化每个特征通道的均值和标准差：其中γ，β∈RC是从数据中学到的仿射参数。µ（x），σ（x）∈RC是在批次大小和空间尺寸上独立计算的平均值和标准偏差。对于每个功能通道，略有不同：

（三）实例规范化

在原始的前馈样式转换方法中，样式传输网络在每个卷积层之后都加入了一个批量标准化层，Ulyanov等[52]，发现如果使用实例规范化层对批量标准化层进行替换即可实现显著的改善。批量标准化和实例规范化其实本质上是同一个东西，只是实例标准化是作用于单张图片（对单个图片的所有像素求均值和标准差），但是批量标准化作用于一个批次图片（对一个批次里所有的图片的所有像素求均值和标准差）。：

与批量标准化层不同，该公式中的µ(x)和σ(x)针对的是每个维度的跨空间尺寸计算。频道和每个样本：另一个区别是实例规范化层在测试时间不变的情况下进行应用，而批量标准化层通常使用总体统计量对小批量统计量进行替换。

（四）条件实例规范化

Dumoulin等人研究了仿射参数γ和β，并在此基础上提出一个条件实例规范化层（CIN），该层针对每种样式S学习一组不同的参数γs 和 βs

训练期间，样式图像及其索引s是从固定的样式s∈{1，2，...，S}（他们的实验中S = 32）中随机选择的。 然后，通过样式转换网络对内容图像进行处理，处理过程中在条件实例规范化中使用相应的γs和βs层。实验结果发现网络可以使用实例规范化层中相同的卷积参数和不同的仿射参数来生成样式完全不同的图像。

与没有归一化层的网络相比，具有条件实例规范化层的网络需要2FS附加参数，其中F是特征层中特征图的总数。由于附加参数的数量与样式的数量呈线性比例关系，因此扩展该方法用于对大量样式（例如数万种）进行建模不具备实际运用的能力。 而且，在遇到新的样式时，如果不对该的方法的网络进行重新训练就无法适应任意的新样式。

1. 基于自适应实例规范化的图像风格迁移

（一）实验原理

实验表明，深度神经网络的卷积特征统计可以捕获图像的样式[16、30、33]。其中Gatys等， [16]使用二阶统计量作为优化目标。 [33]Li的研究结果表明，匹配许多其他统计信息，包括均值和方差，对于样式转换也同样有效。根据这些研究结果，可以认为实例规范化通过规范化特征统计量（即均值和方差）来执行某一种形式的样式规范化。尽管在一些研究[16，33]中深度神经网络用作图像描述符，但是有理由认为生成器网络的特征统计信息也可以控制生成的结果图像的样式。

由于批量标准化层在执行操作时是规范化一批样本而不是其中单个样本的特征统​计量，因此可以直观地理解为批量标准化将一批样本标准化为以单个样式为中心。但是，对于每个样本可能仍具有不同的样式。因此程序希望将所有图像都转换为相同样式时，这会带来很大的麻烦，就像原始前馈样式转换算法[51]一样。尽管卷积层调整后可以学会补偿某一批样本内的样式差异，但它给训练带来了更多麻烦。另一方面，实例规范化可以将每个样本的样式归一化为目标样式。正是因为网络的其余部分可以专注于内容操作，同时丢弃原始样式信息，所以对模型的训练可以更容易的进行。 正是由于实例规范化的以上优点，条件实例规范化可以成功的运行，其主要的特点在于：不同的仿射参数可以将特征统计量标准化为不同的值，从而将输出图像标准化为不同的样式。

如果实例规范化将输入归一化为仿射参数指定的单个样式，是否可以通过使用自适应仿射变换将其适应于任意给定的样式？ 在这里，我们提出了对IN的简单扩展，我们称之为自适应实例规范化（AdaIN）。自适应实例规范化接收内容输入x和样式输入y，并通过简单的操作对齐x的通道方向均值和方差以匹配y的均方根和方差。 与批量标准化，实例规范化或条件实例规范化不同，自适应实例规范化没有可以学习的仿射参数。 取而代之的是从样式输入中自适应计算仿射参数：

其中，我们只需用σ（y）缩放输入的归一化内容，然后用µ（y）对其进行平移。与IN相似，这些统计信息是跨空间位置计算的。

直观地，让我们考虑一个特征通道，该通道可以检测特定样式的笔触。具有这种笔触的样式图像将为此功能产生较高的平均激活度。 AdaIN产生的输出将为此功能具有相同的高平均激活，同时保留内容图像的空间结构。笔触功能可以通过前馈解码器反转到图像空间，类似于[10]。该功能通道的变化可以对更细微的样式信息进行编码，这些信息也将传输到AdaIN输出和最终输出图像。

（二）环境构建及训练

样式传输网络T将内容图像c和选定的样式图像s作为输入，并合成输出图像，该输出图像将前者和后者的内容重新组合在一起。实验采用一种简单的编码器-解码器结构，其中编码器f固定在预训练的VGG-19的前几个层（最多relu​​4 1）上[48]。首先对特征空间中的内容和样式图像进行编码，然后将两个特征图都馈入自适应实例规范化图层，该图层将输入的内容图像的均值和方差与输入的样式图像的均值和方差对齐，从而生成目标特征图t：

训练一个随机初始化的解码器g以将样式传输网络映射回图像空间，从而生成风格化图像T（c，s）：

解码器主要用于对编码器进行镜像操作，所有池化层均被上采样方法所取代，以减少checker- board效应。同时在目标特征图像和解码器中都使用了反射填充，以避免出现边界伪像。另一个重要的有关体系结构的选择是解码器是应使用实例规范化还是批量标准化。正如上文原理中所述。实例规范化将每个样本归一化为单一样式，而批量标准化则将一批样本归一化为以单一样式。如果我们希望解码器生成截然不同的样式的图像时，两者都不可取。因此，在解码器中不使用归一化层。

实验中使用MS-COCO [36]对网络进行训练，内容图像和大部分绘画数据集来自网络，将WikiArt 作为样式图像。每个数据集包含大约80，000个训练示例。实验中使用adam优化器[26]同时8个内容样式图像作为一个批次进行批处理。在训练过程中，首先将两个图像的最小尺寸调整为512，同时保留宽高比，然后随机裁剪256×256尺寸的区域。由于使用的网络是完全卷积的，因此可以在测试过程中应用于任何尺寸的图像。

参考其他类似实验的设计，我们同样使用预训练的VGG-19计算损失函数来训练解码器。它是内容损失Lc和样式损失Ls与样式损失权重λ的加权组合。含量损失是目标之间的Euclidean距离特征和输出图像的特征。其中使用自适应实例规范化输出t作为内容目标，而不是内容图像的常用特征响应，改变之后发现这会导致收敛速度稍快，也符合反转自适应实例规范化输出t的目标。

由于实验中使用的自适应实例规范化层仅传递样式特征的均值和标准偏差，因此样式损失仅与这些统计信息匹配。尽管常用的Gram矩阵损失可以产生相似的结果，但实验中使用IN统计量进行匹配，因为它在概念上更干净。其中每个φi表示VGG-19中用于计算样式损失的一层。在该实验中，选择使用权重相等的relu1 1，relu2 1，relu3 1，relu4 1层