（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，该技术以自适应实例规范化层为核心，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

[一、 绪论 3](#_Toc30377)

[（一）早期图像艺术风格转换 3](#_Toc22826)

[（二）基于神经网络的艺术风格转换 4](#_Toc24909)

[（三）本文采用的方法 7](#_Toc20320)

[二、 相关方法综述 7](#_Toc29028)

[（一）卷积神经网络 7](#_Toc1941)

[（二）批量标准化 11](#_Toc19832)

[（三）实例规范化 13](#_Toc10970)

[（四）条件实例规范化 13](#_Toc27717)

[三、 基于自适应实例规范化的图像风格迁移 14](#_Toc10045)

[（一）实验原理 14](#_Toc5337)

[（二）环境构建及训练 16](#_Toc23530)

[四、 实验结果对比 17](#_Toc21691)

[五、 实验结论 18](#_Toc18218)

[参考文献 20](#_Toc17515)

# 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困难重重。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。

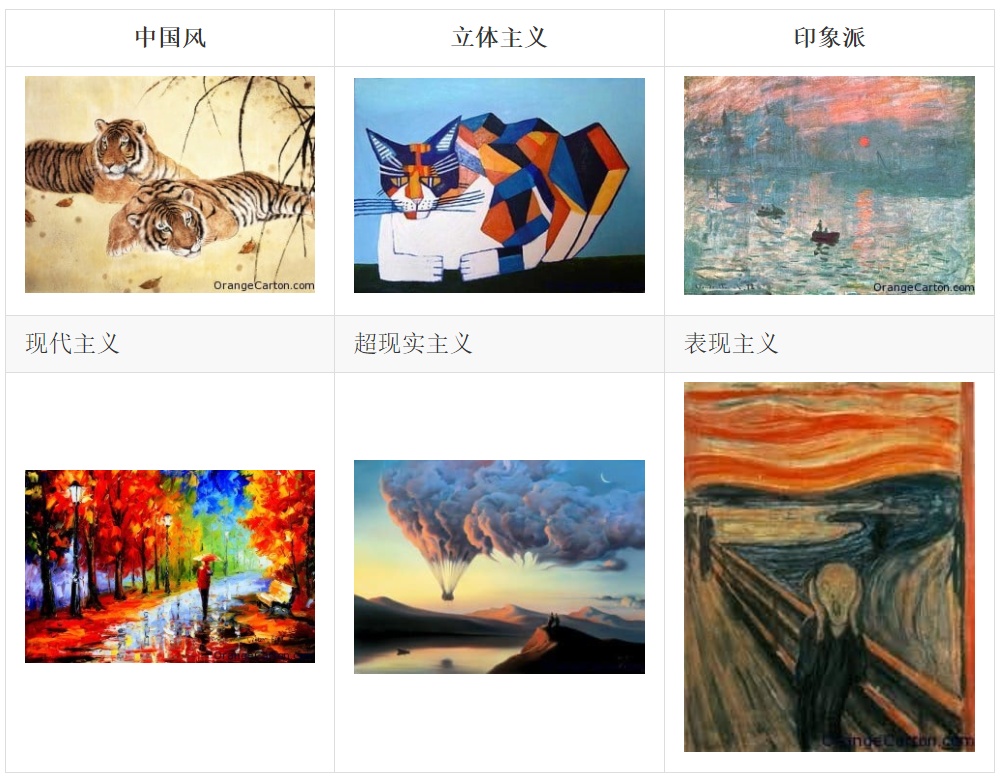


图1.1：各种常见的艺术风格

## （二）基于神经网络的艺术风格转换

Gatys[1]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[5]和非参数采样[6]，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[7]，其中Li和Wand [8]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[10]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[11]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。

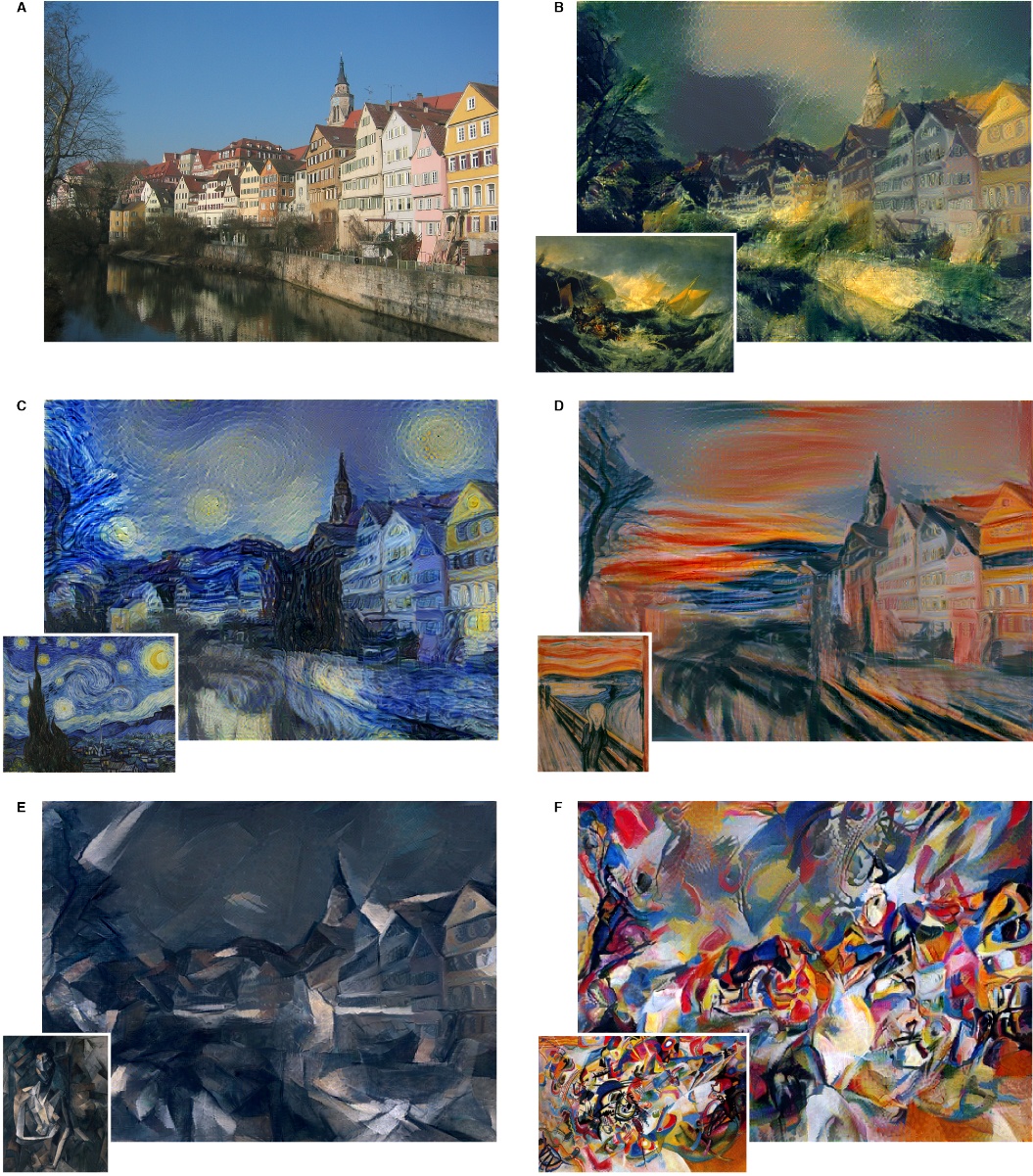


图1.2 早期样式转换图片

Chen和Schmidt[12]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。 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，对抗性损失，直方图损失，CORAL损失，MMD损失以及通道平均和方差之间的距离等。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。

# 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。卷积神经网络相较于传统的图像处理算法的优点之一在于避免了对图像复杂的前期预处理过程，尤其是人工参与图像预处理过程，卷积神经网络可以直接输入原始图像进行一系列工作。卷积神经网络成功的原因之一在于其所采用的局部连接和权值共享的方式：一方面减少了权值的数量使得网络易于优化；另一方面降低了模型的复杂度，也就是减小了过拟合的风险

该优点在网络的输入是图像时表现的更为明显，使得图像可以直接作为网络的输入，避免了传统识别算法中复杂的特征提取和数据重建的过程，在二维图像的处理过程中有很大的优势，如网络能够自行抽取图像的特征包括颜色、纹理、形状及图像的拓扑结构，在处理二维图像的问题上，特别是识别位移、缩放及其他形式扭曲不变性的应用上具有良好的鲁棒性和运算效率等。

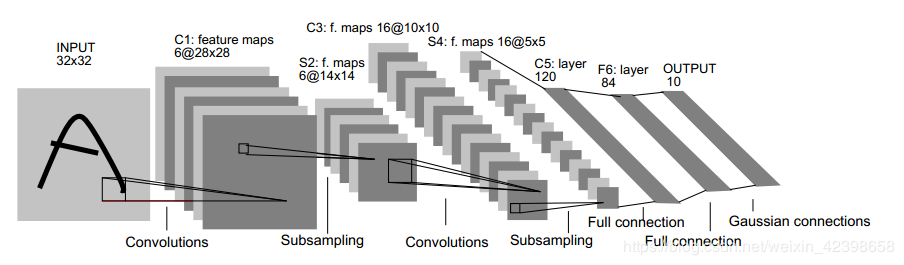


图2.1.1卷积神经网络结构图

卷积神经网络由一个或多个卷积层和顶端的全连接层组成，同时也包括ReLU激励层和池化层（pooling layer）等。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。

### 1.数据输入层

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。该层要做的处理主要是对原始图像数据进行预处理，其中包括：

1. 去均值：把输入数据各个维度都中心化为0，如下图所示，其目的就是把样本的中心拉回到坐标系原点上。
2. 归一化：幅度归一化到同样的范围，如下所示，即减少各维度数据取值范围的差异而带来的干扰，比如，我们有两个维度的特征A和B，A范围是0到10，而B范围是0到10000，如果直接使用这两个特征是有问题的，好的做法就是归一化，即A和B的数据都变为0到1的范围。
3. PCA/白化：用PCA降维；白化是对数据各个特征轴上的幅度归一化

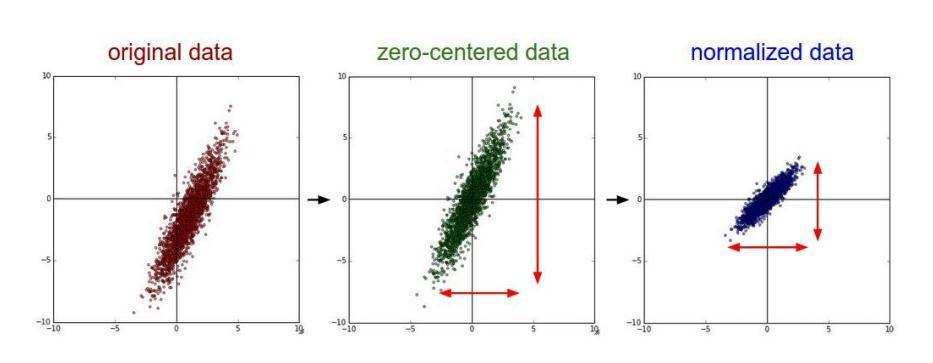


图2.1.2去均值效果图

### 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。

这种机制具有如此有效性的原理在于，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。池化层的具体作用为：

①特征不变性，也就是我们在图像处理中经常提到的特征的尺度不变性，池化操作就是图像的resize，平时一张狗的图像被缩小了一倍我们还能认出这是一张狗的照片，这说明这张图像中仍保留着狗最重要的特征，我们一看就能判断图像中画的是一只狗，图像压缩时去掉的信息只是一些无关紧要的信息，而留下的信息则是具有尺度不变性的特征，是最能表达图像的特征。

②特征降维，我们知道一幅图像含有的信息是很大的，特征也很多，但是有些信息对于我们做图像任务时没有太多用途或者有重复，我们可以把这类冗余信息去除，把最重要的特征抽取出来，这也是池化操作的一大作用。

1. 在一定程度上防止过拟合，更方便优化。

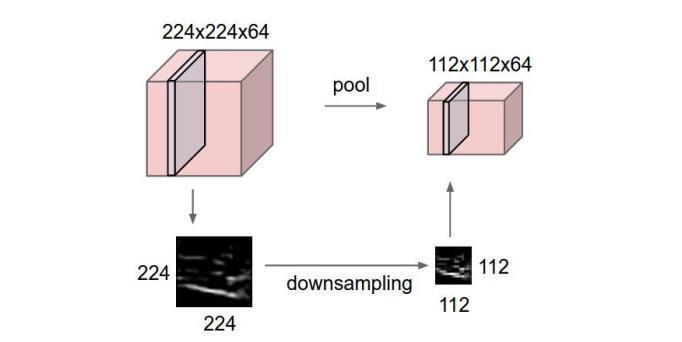


图2.1.2池化层原理图

### 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。

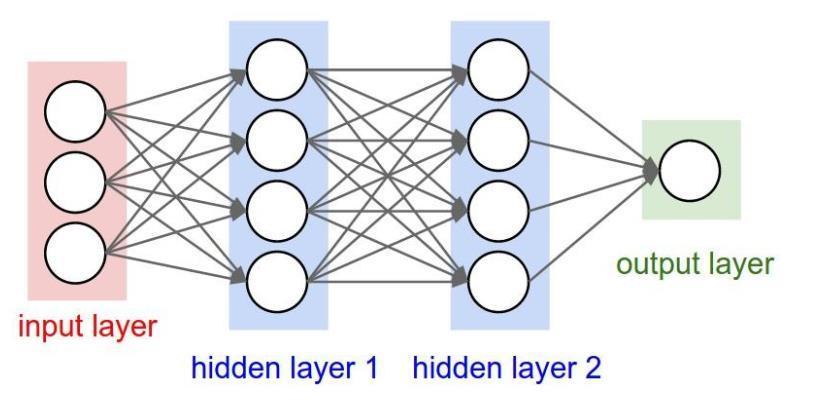


图2.1.3全连接层示例

## （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，批量标准化层标准化每个特征通道的均值和标准差：

 (1)

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：

 (2)

 (3)

BN 的具体流程如下图所示：首先，它将隐藏层的输出结果（如，第一层：，状态值）在 batch 上进行标准化；然后，经过缩放(scale)和平移(shift)处理，最后经过 RELU 激活函数得到送入下一层（就像在数据预处理中将 x 进行标准化后送入神经网络的第一层一样）。

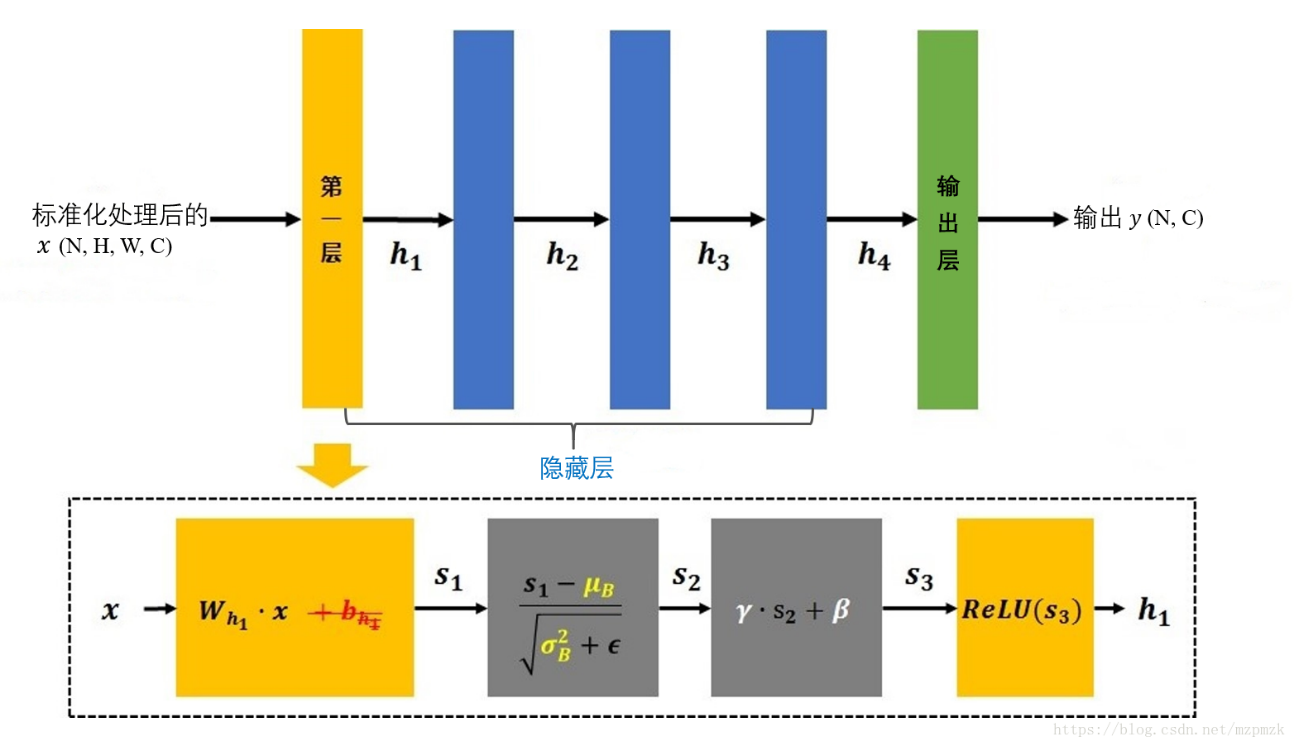


图2.2批量标准化操作

①：γ 和β： γ和 β都是可学习参数，分别用作对标准化后的值进行缩放(scale)和平移(shift)，提高网络的表达能力(不加此参数的话，中间层的输出将被限制在标准正态分布下)。当参数分别初始化为 和 时，标准化的变量被还原为原来的值。随着网络的训练，网络会学到最合适的 γ和 β，最终中间层的输出将服从均值为 β方差为 的正态分布

②：ϵϵ： 为了避免方差为 0 而加入的微小正数(例如：0.001)

③：： 减去第一层 batch 个数据的均值 后， 偏置项 b的作用会被抵消掉，所以在使用 BN 的层中没必要加入偏置项 b (红色删除线)，其平移的功能将由 β来代替。

## （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）。：

 (4)

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：

 (5)

 (6)

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和，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和 具有如下关系：

 (7)

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。

# 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。其中Gatys等， 使用二阶统计量作为优化目标。 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：

 (8)

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。

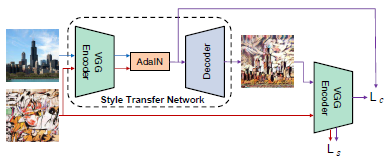


图3.1自适应实例规范化操作图

## （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：

 (9)

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）：

 (10)

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:

 (11)

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:

 (12)

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

 (13)

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。

# 实验结果对比

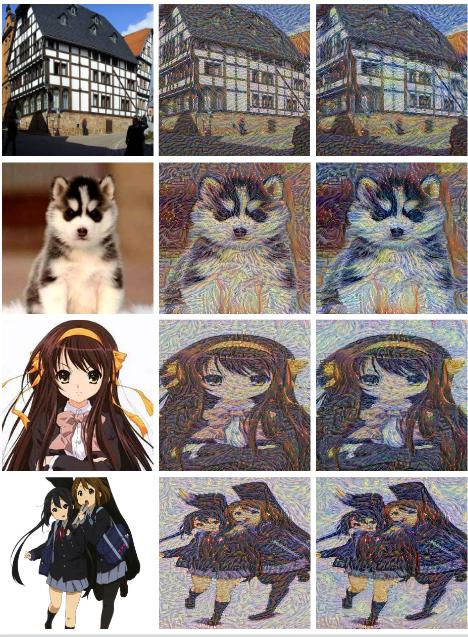


图4.1从左到右分别为内容图片、自适应实例规范化图片、对照组



图4.2 样式图片

为了更加直观的体现自适应实例规范化进行风格转换的速度和对转换的结果图片进行对比，该小节将会将该方法与另一种使用VGG19网络且基于优化的方法进行比较。通过在默认配置下运行两种方案的代码来获得比较方法的结果。同时为了控制变量，对所有测试图像的尺寸均设定为512×512。  
   在图中，显示了两种方法对不同样式图片和不同内容图片所产生的样式转换结果。值得注意的是，在本文种所研究的自适应实例规范化的模型训练过程中从未观察到所有测试样式图像。而基于优化的样式转方案的结果是通过将一种网络拟合到每种测试样式而获得的。即使这样，自适应实例规范化所转换出的风格化图像的质量依然可以与基于优化的方案媲美。

在更多的一些情况下，自适应实例规范化方法所生成的图片质量应该比优化方案的图片质量稍差一点。因为图像风格转换算法一定程度上符合等价交换原理，即想要获得某些方面的提升必然要相应的用其他方面的优势去进行交换。所有算法都要在速度，灵活性和质量之间进行三项权衡。

对于大多数比较的图像，自适应实例规范化方法会更加完整的传递样式。大多数图像风格转换算法都致力于将每个内容补丁与最接近的样式补丁进行匹配。但是，一旦大多数内容补丁与少数无法完整代表目标样式的样式补丁进行匹配，即存在一定的概率会将目标样式的样式提取错误，则样式传输将失败。因此，自适应实例规范化使用的匹配全局特征统计量是一个更加一般化的经过数学期望计算的样式补丁。

# 实验结论

在本次实验中，研究说明了一个简单的自适应实例规范化（AdaIN）层，该层实现了实时的任意样式转换。 该方法与基于特征统计的先前的神经样式转换方法之间的概念具有一定的差异。 Gatys等，采用优化过程来操纵像素值以匹配特征统计，在其实验中，前馈神经网络取代了优化过程。但是，前馈神经网络依然是使用通过修改像素值以间接匹配特征统计信息的方案，并经过一定训练来实现。自适应实例规范化采用了一种非常不同的方法，即一次将特征空间中的统计信息直接对齐，然后将特征反转回到像素空间。

同时为了提升速度和灵活性，导致所生成的样式转换图片会有一定的质量损失，在实验中由于内容图片和样式图片过少无法完全体现出图片质量的区别，即具有一定的偶然性。同时由于自适应实例规范化基于更大的VGG网络，在处理高分辨率图片时处理时间会较长，通过对体系结构的改进可以优化该问题，使得使用范围更加宽广。

考虑到该方法在结构上的简单性，有理由认为该方案仍有很大的改进空间。在未来的工作中，可以通过探索更高级的网络架构来进行改进，例如残差架构或具有来自编码器的附加跳过连接的架构。同时还应当研究更复杂的训练方案来改进图片的质量，例如增量训练。此外自适应实例规范化层仅对齐最基本的特征统​​计信息（均值和方差），用相关对齐或直方图匹配代替实例规范化层可以通过传递高阶统计信息来进一步提高质量。

# 参考文献

[1] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016. [1](#_bookmark0), [2](#_bookmark1), [3](#_bookmark2), [5](#_bookmark8), [6](#_bookmark15), [7](#_bookmark18), [8](#_bookmark23)

[2] J. E. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg, “State of the ‘art’: A taxonomy of artistic stylization techniques for images and video,” IEEE transactions on visualization and computer graphics, vol. 19, no. 5, pp. 866–885, 2013.

[3] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for realtime style transfer and super-resolution,” in European Conferenceon Computer Vision, 2016, pp. 694–711.

[4] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, “Texture networks: Feed-forward synthesis of textures and stylized images,” in International Conference on Machine Learning, 2016, pp. 1349–1357.

[5] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/ synthesis. In *SIGGRAPH*, 1995.

[6] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, 2001

[7] M. Elad and P. Milanfar. Style-transfer via texture-synthesis. *arXiv preprint arXiv:1609.03057*, 2016.

[8] C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *CVPR*, 2016. [1](#_bookmark0), [2](#_bookmark1), [3](#_bookmark2)

[9] V. Dumoulin, J. Shlens, and M. Kudlur. A learned represen- tation for artistic style. In *ICLR*, 2017. [1](#_bookmark0), [2](#_bookmark1), [3](#_bookmark2), [5](#_bookmark8), [6](#_bookmark15), [7](#_bookmark18)

[10] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang. Diversified texture synthesis with feed-forward networks. In *CVPR*, 2017. [1](#_bookmark0), [2](#_bookmark1), [8](#_bookmark23)

[11] T. Q. Chen and M. Schmidt. Fast patch-based style transfer of arbitrary style.*arXiv preprint arXiv:1612.04337*, 2016. [1](#_bookmark0), [2](#_bookmark1), [4](#_bookmark5), [5](#_bookmark8), [6](#_bookmark15), [7](#_bookmark18)