编 号

本科生毕业设计

题目：基于 AdaIN 的快速风格迁移算法研究

人工智能与计算机 学院

计算机科学与技术 专业

学　　号 1033190609

学生姓名 刘鸿嘉

指导教师 李辉　讲师

二〇二三 年 五 月

设计总说明

近年来，计算机视觉领域取得了重大进展，特别是在图像和视频处理领域。该领域最令人激动和最具挑战性的任务之一是风格迁移，其目的是在保留内容的同时将输入图像风格迁移成参考图像风格。风格迁移在艺术、设计、广告和娱乐等各个领域都有大量应用，吸引了学术界和工业界的大量关注。然而，大多数现有的风格迁移算法存在三个主要的限制：速度慢、质量低以及模型灵活性差。速度慢是由于优化过程的高计算成本，这涉及到为输入图像或视频中的每个像素或特征解决一个复杂的优化问题。低质量模型灵活性差是由于风格表示缺乏灵活性和多样性，它通常被限制在一套固定的预定义风格或纹理中。

为了解决这些局限性，提升风格迁移算法的迁移速度与迁移多样性，本文拟使用基于自适应实例归一化(Adaptive Instance Normalization, AdaIN)和引导滤波(Guided Filter)的风格迁移模型 GuideStyle 开发一款应用于任意风格的快速风格迁移系统。该文启发于常见的绘画过程，在绘画过程中通常先绘制草稿图，以确定画作的基础结构，之后再对画面进行细节修改，以达到更优美的画面质量。故本系统首先通过草稿网络绘制全局样式并生成低分辨率图像。然后，将草稿图像与引导滤波残差图输入修订网络以增强图像的局部细节，生成高分辨率图像。并且，更高的图像细节可以通过堆叠多层引导滤波金字塔的网络轻松生成。最后的结果图像是通过聚合多层的修订网络输出得到的。

本文快速风格迁移系统主要由三部分组成。基于混合 Python 编程语言和 Qt 库的跨平台 GUI 库 PyQt5 开发前端界面实现与目标用户的交互功能；基于 AdaIN 和引导滤波的快速风格迁移模型进行对输入图片特征的提取、处理与迁移作为本文的核心部分；基于在丰富且适用性强的数据集上的实验结果，进行模型的迁移效果在不同条件下的分析和最终迁移效果的结果映射。

为了验证设计的有效性，本文使用 2 个通用的评价指标用于评价该快速风格迁移系统，分别为峰值信噪比(Peak Signal-to-Noise Ratio，PSNR)、结构相似(Structural similarity，SSIM)。该系统实验数据集采用含超过 10000 张内容图片与风格图片数据集，由于对风格图片与内容图片组合无限制，因此风格与内容的组合接近无限，数据丰富性极高，可很好地适用于本文系统的研究。

结果表明，本文设计的基于自适应实例归一化和引导滤波的风格迁移模型，在 SSIM、PSNR 等性能指标上有较好的表现。风格化图像在保留内容信息的同时，其整体风格模式被正确迁移。通过与传统自适应实例归一化风格迁移模型相比，尽管整体的风格均被成功迁移了，但是在局部特征分布上传统模型仍有明显不足，本文的风格化图像具有更高的图像质量和图像细节。同时 GuideStyle 网络具有较高的迁移速度，易于工业上使用，因此预期本文的快速风格迁移系统能够在实际场景中取得较好的效果。

关键词：自适应实例归一化；风格迁移；引导滤波；GuideStyle

ABSTRACT

In recent years, significant progress has been made in the field of computer vision, particularly in the area of image and video processing. One of the most exciting and challenging tasks in this field is style transfer, which aims to convert the style of an input image to the style of a reference image while preserving its content. However, most existing style transfer algorithms suffer from two major limitations: slow speed and low quality. The slow speed is due to the high computational cost of the optimization process, which involves solving a complex optimization problem for each pixel or feature in the input image or video. The low quality is due to the lack of variety in the style representation, which is usually limited to a fixed set of predefined styles or textures.

To address these limitations and to improve the speed and variety of style transfer algorithms, this paper proposes to develop a fast style transfer net, termed as GuideStyle, based on Adaptive Instance Normalization (AdaIN) and Guided Filter for arbitrary styles. The proposed method is inspired by the common painting process, where we usually draw a sketch first to determine the structure of the base of the painting, and then modify the details to achieve a more beautiful image quality. Therefore, firstly, the global style and a low-resolution image are generated by a draft net. Then the initial image and the residual map from the guide filter are fed into the revision net to enhance the local details of the image and generate a high-resolution image. Image details can be easily generated by stacking the network of multi-layer guide filter pyramids. The final stylized image is obtained by aggregating the outputs of the multi-layer revision network.

The fast style transfer system consists of three main parts. PyQt5, a cross-platform GUI library based on a mixture of Python programming language and Qt library, develops the front-end interface to realize the interaction function with the target user; the fast style transfer model based on AdaIN and guide filter for the extraction, processing and transfer of input image features as the core of this system; the analysis of the transfer effect of the model under different conditions and the mapping of the stylized image are performed, based on the experimental results on a rich and applicable dataset.

Two common evaluation metrics are used to evaluate the fast style transfer system, namely peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The experimental dataset contains over 10,000 content images and style images. Since there is no restriction on the combination of style images and content images, the combination of style and content is almost infinite, and the data richness is very high, which is well suited for studying this system.

The results show that the style transfer model based on AdaIN and “guide filter” algorithm performs well in terms of performance metrics, and its holistic style patterns are correctly transferred while preserving the content information. Compared with the existing style transfer model based on AdaIN, our method can preserve more natural color information and generate better stylized image. Meanwhile, our method has high migration speed, which is easy to use in industry. It is expected that the fast style transfer system in this thesis can achieve better results in practical scenarios.

Keywords: AdaIN, style transfer, guided filtering, GuideStyle

目 录

第 1 章 绪论

1.1 研究背景与意义

近年来，计算机视觉领域取得了重大进展，特别是在图像和视频处理领域。该领域最令人激动和最具挑战性的任务之一是风格迁移，其目的是将输入图像或视频的风格迁移成参考图像或视频的风格同时，保留其结构内容。风格迁移在艺术、设计、广告和娱乐等各个领域都有大量应用，并吸引了学术界和工业界的大量关注。

然而，大多数现有的风格迁移算法存在两个主要的限制：迁移速度慢和质量低。速度慢是由于优化过程的高计算成本，这涉及到为输入图像或视频中的每个像素或特征解决一个复杂的优化问题。低质量是由于风格表示缺乏灵活性和多样性，它通常被限制在一套固定的预定义风格或纹理中。

为了解决这些局限性，研究人员提出了各种快速和高质量的风格迁移算法，其中基于AdaIN[1]的风格迁移算法显示了很好的效果。AdaIN是自适应实例归一化(Adaptive Instance Normalization)的缩写，它是一种归一化技术，可以调整特征图的平均值和方差，使其与风格图像相匹配，从而实现灵活多样的风格迁移。基于AdaIN的算法可以在现代GPU上实现实时或接近实时的速度，同时保持高感知质量和保真度。AdaIN已被用于各种最先进的风格迁移算法，如任意风格迁移(AST)和多风格迁移(MST)。

然而，尽管有这些优势，基于AdaIN的风格迁移算法仍然有一些需要解决的限制和挑战。一些需要改进的潜在领域是：

速度： 尽管基于AdaIN的算法可以在现代GPU上达到实时或接近实时的速度，但与其他图像处理任务相比，它们仍然需要大量的计算资源和时间。因此，提高算法的速度和效率是一个至关重要的挑战，特别是对于需要实时或交互式风格迁移的应用。

质量： 尽管基于AdaIN的算法可以产生高质量和有视觉吸引力的结果，但它们仍然受到一些伪影和扭曲的影响，特别是在处理复杂或多样化的风格时。因此，提高算法的质量和保真度是另一个重要的挑战，特别是对于那些需要高质量或逼真的风格迁移的应用。

灵活性： 尽管基于AdaIN的算法可以转移广泛的风格和纹理，但它们在风格表现的多样性和丰富性方面仍有一些限制。因此，提高算法的灵活性和多样性是第三个挑战，特别是对于那些需要个性化或定制化风格迁移的应用。

1.2 国内外研究现状

风格迁移的历史可以追溯到计算机图形学和图像处理的早期，当时研究人员探索了纹理合成、图像过滤和艺术渲染的各种技术。然而，现代风格迁移的时代是随着深度学习的出现而开始的，特别是卷积神经网络(CNN)，它使人们能够从大规模的数据集中自动学习风格和内容的表示。

在过去的十年中，国内外都出现了风格迁移算法的研究热潮。其中一些值得注意的里程碑和贡献是：

(1) 神经风格迁移(Neural Style Transfer, NST)[2]： NST是Gatys等人(2015)的一项开创性工作，他们开启了神经风格迁移的时代。在Gatys等人的文章中，他们提出了

一张图片中艺术家的风格可以使用预训练的DCNN(Deep Convoltional Neural Networks)提取特征之间的相关性获得。NST激发了许多后续工作，探索原始算法的不同变化和扩展。

(2) 快速风格迁移(FST)[3]： FST是Johnson等人(2016)的一项工作，介绍了一种基于前馈CNN的快速和可扩展的风格迁移算法。FST可以在现代GPU上达到实时或接近实时的速度，并且可以应用于图像和视频。

(3) 自适应实例归一化(AdaIN)：AdaIN是Huang和Belongie(2017)的一项工作，介绍了一种可以自适应地调整特征图的平均值和方差，以匹配风格图像的平均值和方差，从而实现灵活多样的风格迁移的归一化方法。同时如图1-1所示，AdaIN模型网络使用了较为简单的自编码器架构，并使用预训练的VGG网络作为模型的编码器。AdaIN已被用于各种最先进的风格迁移算法，如任意风格迁移(AST)和多风格迁移(MST)。

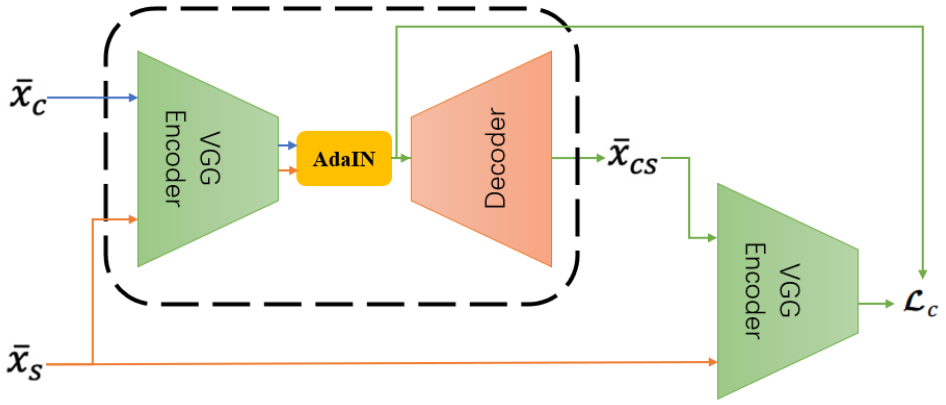


图1-1 AdaIN网络模型

(4) 超分辨率风格迁移：Wang[28]等人使用模型压缩方法，即协同蒸馏(collaborative distillation)来减少VGG-19的卷积核数量，来使其可以在拥有较小显存的 GPU上渲染超分辨率图像。尽管内存消耗显著的减少，修剪后的模型生成超分辨率图像的速度仍然不够理想。此外，较大程度的压缩会极大的损失生成图像的质量。另外的解决方案是直接设计一个轻量级模型。Sanakoyeu等人[29]设计了用来迁移一个特定风格样例或高分辨率(如1024*1024像素)的小型前馈网络。然而他们的模型并不能适用于任意风格迁移。

(5) 风格生成对抗网络(StyleGAN)[30]：StyleGAN是由NVIDIA公司Tero Karras等人在2019年提出的。他们的开创性工作不仅可以生成高质量和高仿真度的伪造图像，同时其生成模型能够对特征新信息具有更深层的理解，从而使其能够对生成的伪造图像进行语义级的控制和调整。StyleGAN是ProGAN[31]图像生成器的升级版本，其重点关注生成器网络(G)。StyleGAN可以对生成伪造人脸的风格进行多类型调整，包括人脸的表情、朝向以及发型等，甚至包对人脸纹理细节上的调整，如肤色、光照等。

1.3 主要研究内容

本论文围绕基于AdaIN的快速风格迁移算法展开研究，通过对比实验对传统图像风迁

移模型结构进行优化提升。在本文最终的风格迁移模型GuideStyle中，我们引入内容感知特征重组(Content-aware reassembly of features, CARAFE)来获得更大的感受野以更好的捕获模型信息，并增加残差网络块以获得更深的网络深度。并引入多种损失优化项(如变差分损失、rEMD损失，Identity损失等)来更好的逼近人类主观评估特质，即获得更高质量的风格化图像。同时，本文的模型使用了基础的U-net网络架构，在网络中间层引入多层AdaIN跳跃连接结构，使风格特征与内容图片在多维度下更充分的结合。此外，在每次AdaIN操作之前，本文引入随机噪声，来获得更大的网络自由度，以更加充分的表现风格图像的纹理信息，其中噪声的增加权重是可学习的。

具有创新性的是，本文在提出本文的基础风格迁移模型外，介绍了基于引导滤波的修正网络。其由一个最简单的自动编码器结构组成，通过控制卷积层数量来保证网络仅具有一个较小的感受野，目的在于对草稿网络的生成图像进行细节补全并获得更高的图像分辨率。此外，本文的修正网络可以通过上采样金字塔结构进行堆叠以生成更高分辨率的图像。由于网络结构简单，其可以大幅缩减模型训练时间和硬件资源的消耗。

最后，通过计算结果图与输入图像间的结构相似性(Structural Similarity, SSIM)[6]、峰值信噪比(Peak signal-to-noise ratio，PSNR)等评估参数评价模型效果，并进一步采取措施提升模型识别准确率，最终得出研究结论。

本文的主要研究内容如下：

(1) 对开发基于AdaIN的图像快速风格迁移系统的经济可行性和技术可行性进行深入研究。

(2) 选取恰当的内容与风格数据集(COCO、WikiArt)，并根据实验特点，选择合适的数据预处理方式与数据增强方法。

(3) 选取风格迁移历史上多种模型进行研究，并对模型效果进行比较。

(4) 调整模型训练策略，对比研究不同训练策略下模型的识别效果。

(5) 综合上述研究方式，得出最终模型对比与模型改进的结论，实现较好风格迁移图像。

1.4 可行性研究

1.4.1 经济可行性

移动互联网技术催生了各种智能产品，其中移动应用(APPs)在人们的工作和日常生活中已经无处不在。目前的移动应用市场中，众多类型的APP和产品已经趋于饱和，导致用户留存率低。而随着视觉传播时代的到来，用户对丰富多彩的视觉内容的要求越来越高，他们渴望对自己拍摄的图片进行即时美化、编辑、分享、标记和渲染。照片美化已经逐渐成为一种流行的爱好，70%的来自中国的手机拍照使用者每周都会在她们的手机上(如使用QQ空间、微信朋友圈等等)分享照片。此外，外国的互联网用户在脸书上分享社交照片的数量甚至是推特的11倍。同时，越来越多的女性用户使用手机拍摄自己的日常照片并上传到互联网上分享给朋友。因此，不断更新手机应用以及功能以适应手机使用者日益增长的生活需要正变得越发迫切，如对手机摄像功能的改善与增强。

在智能手机时代，美容应用程序进一步扩大了女性的自拍需求。这也反映出人们在

分享照片前需要修饰，这一点至关重要。自2014年以来，照片美化应用程序在iOS和Android平台上的安装和活动指数呈现上升趋势。根据对第三方照片处理软件使用情况的调查，2016年，照片美化软件的使用率达到81.6%，比2015年高出3.9个百分点。这说明人们对照片美化软件的认识和使用有了很大提高。根据市场分析，可以得出结论，基于风格迁移的项目开发迎合了市场的需求。因此，项目的收益超过了成本，在经济上是可行的。

1.4.2 技术可行性

本项目的主要模块，即快速风格迁移网络。鉴于目前工业界基于神经网络模型的开发与应用变得日益成熟，因此在本项目中，本文融合了多种神经网络模型以达到最优的风格迁移效果。

在代码实现方面，本项目的使用Pytorch框架来编写模型的主干部分。此外，基于python的pytorch框架具有与GPU的多类型接口，使得用户可以很方便的调用电脑的GPU资源来加速模型的运算，并由于python语言的泛用性，其可以很方便的与多领域的扩展工具进行协同开发。在本项目中，使用了一张NIVIDIA的RTX3090显卡来处理模型的运算，其具有10490个CUDA核心以及24GB的显存容量，在训练过程中，本文对512分辨率的图像进行训练，单次训练批大小(batch_size)可容纳四张图像，符合训练需求。同时，该项目使用了Python的图像处理库(Image library)以更方便的对图像进行放缩、裁切等操作，并使用了python端的OpenCV图像处理库来实现对图像的滤波、色彩转变等操作。

在服务端技术方面，python领域下的PyQT非常成熟且易用，契合本项目的设计目标，也易于与图像风格迁移模块进行数据交换。

以上各个模块的架构、技术和操作的便捷与多样，保证了本项目实现的技术可行性。

第 2 章 图像风格迁移的理论基础

2.1 图像风格迁移的理论基础

2.1.1 自动编码器 (Autoencoder, AE)

自编码器的初始设计目标是将输入的高维特征向量x通过编码器E编码为低维特征向量$z = E(x)$，编码的标准是尽可能多的保留特征信息，因此需要再训练一个解码器D，通过低维特征向量z来重构高维特征x，即$x \approx D(E(x))$，其优化目标为：

$$E, D = arg\min_{E,D} \mathbb{E}_{x\sim D}\left[\left\|x - D(E(x))\right\|^2\right] \tag{1}$$

本文使用的encoder-decoder即为一种经典的AE。如图2-1所示，在训练过程中增加一些扰动，其便可以变为去噪自编码器(DAE)。

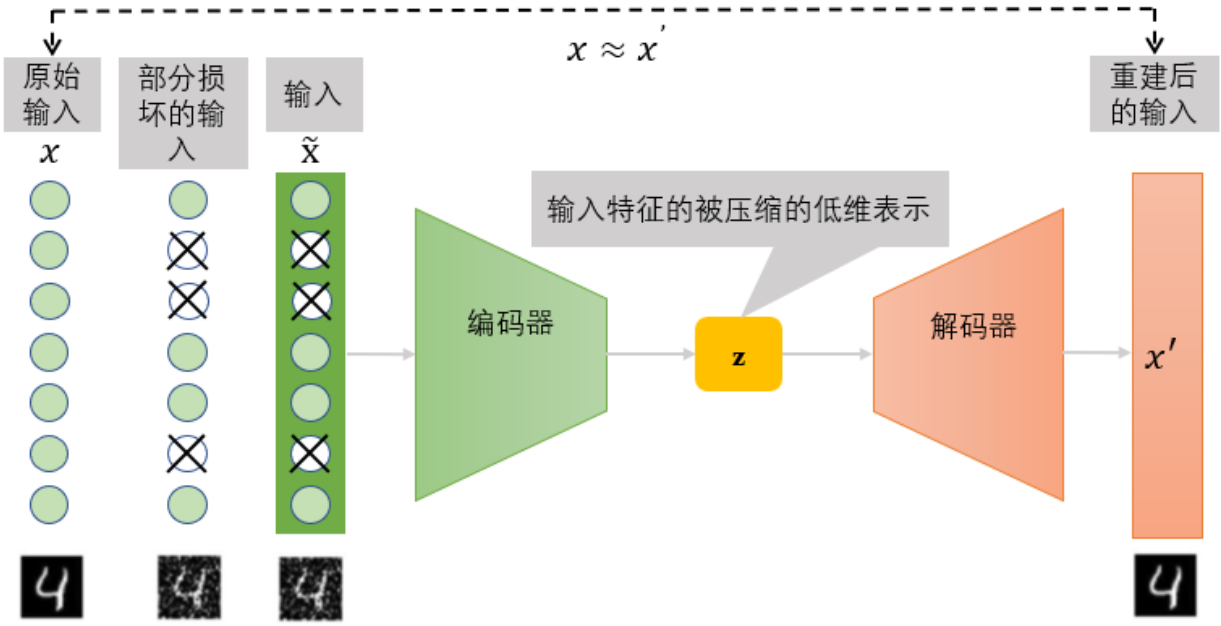


图2-1 DAE模型结构

其中，编码器网络将输入数据转化为低维表示，这个特征维度会远小于原始数据维度，通过这样的操作便达到了数据压缩和降维的目的，同时，低特征维向量往往为中间隐含特征(latent representation)；而解码器将这种压缩后的低维表示转化回原始数据。通时该方法希望解码器的输出数据即重建后数据与编码器的输入数据是近乎一致的，其模型的损失优化项通常使用均方差(MSE)损失。

此外，由于经编码器压缩后的低维特征表示能够被解码器重建，因此可以使用自动编码器网络的编码器部分对高维数据进行压缩，来得到数据的隐含特征。目前的主流方法为先在大规模含有标注的训练集上对网络模型进行预训练操作，之后将预训练参数导入模型，在其之上训练具体的任务。在本文中，我们使用预训练的VGG-19网络的卷积层部分作为模型的编码器，在模型训练的过程中，其权重是被冻结的。

2.1.2 VGG 网络(Visual Geometry Group)

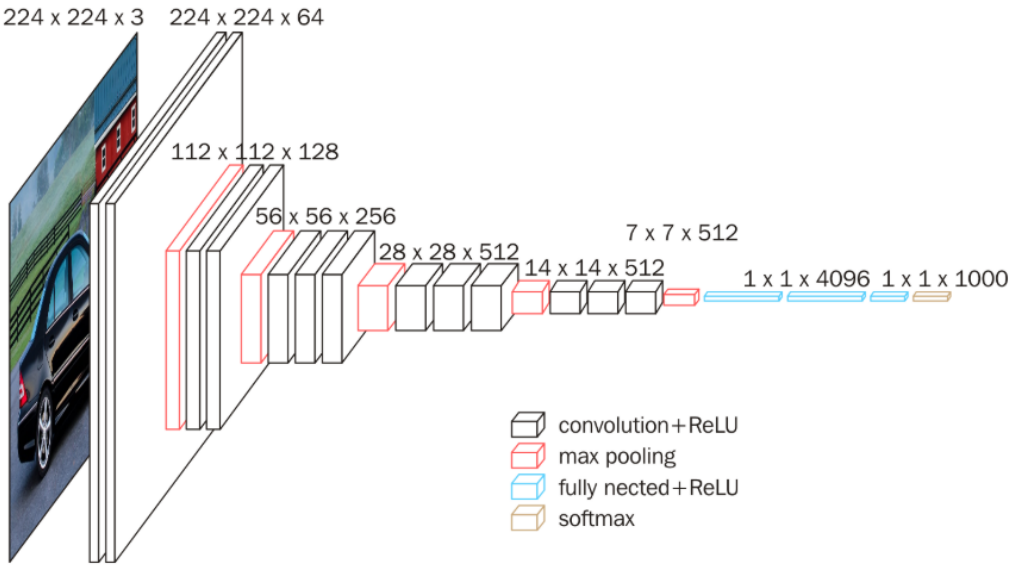


图2-2 VGG-19网络结构

如图2-2，VGG-19[6]网络是一种卷积神经网络(CNN)架构，由牛津大学的视觉几何小组在2014年推出。在AlexNet模型被提出之后，许多学者在模型卷积核大小和图像维度上进行研究以获得更好的识别准确度。而VGG的作者们却对模型的深度进行研究，事实证明他们是正确的。

VGG网络的组成是非常格式化的，其基本上均使用了3*3的小卷积核以及2*2的最大池化层来对输入图像进行特征提取。其中，最大池化层在保留图像的重要特征的同时，大大减少了特征图的像素数量，以加快网络的计算速度。之后VGG使用一个或多个全连接层来实现图像的分类操作，这些层将最后一个卷积层的扁平化输出作为输入，并产生一个类别概率的矢量作为输出。但由于全连接层通常伴随着很大的参数量，因此VGG网络所占用的存储空间以及其所需要的参数计算量是较为巨大的。

该文作者的开创性工作表明，一个5*5的卷积核与两个3*3的卷积核具有相同的感受野，因此在一定程度上它们是等价的。但是由于每个3*3的卷积核后都存在一个ReLU激活函数，因此使用更大的卷积核带来的结果是网络可使用的激活函数次数更少，在通常情况下具有更深网络结构的VGG模型显示出了更优越的表现，其模型的非线性拟合能力更强。此外，使用更小的卷积核堆叠带来了更少的参数数量，例如假设2个具有3*3卷积核的卷积层且输入输出维度均为C个通道，那么其参数数量为$2*(3*3*C*C)=18C^2$，而同样效果的一个具有5*5卷积核的卷积层的参数数量为$5*5*C*C=25C^2$。

VGG网络通过反向传播和随机梯度下降法进行训练，其损失优化项通常使用交叉熵损失来衡量的类概率和实际的类标签之间的差异。

2.2 归一化方法

2.2.1. 批量归一化(Batch Normalization, BN)

批量归一化(BN)由Ioffe和Szegedy[8]提出，如今其已经成为现代深度学习中不可或缺的的一部分，他被应用在许多著名的深度学习模型结构中。批量归一化通常作为全连接或卷积层模块的一部分，通过归一化特征统计以帮助在模型训练时稳定网络。此外，其

在生成性图像建模中也被发现很有效[8]。给定一个输入批次$x \in R^{N\times C\times H\times W}$，BN将每个单独的特征通道的平均值和标准差归一化：

$$BN(x) = \gamma\left(\frac{x-\mu(x)}{\sigma(x)}\right) + \beta \tag{2}$$

其中，γ、$\beta \in R^C$是从数据中学习的仿射参数；$\mu(x)$、$\sigma(x) \in R^C$是平均数和标准差，在每个特征通道的批次大小和空间维度上独立地计算：

$$\mu_c(x) = \frac{1}{NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{nchw} \tag{3}$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(x_{nchw} - \mu_c(x)\right)^2 + \epsilon} \tag{4}$$

BN在训练过程中使用小批量的统计数据，与在推理过程中使用的统计数据不同，其增加了在训练和推理过程中的差异。此外，批量再归一化[9]作为批量归一化算法的延申，其通过在训练过程中使用全局统计数据，能够在一定程度上解决该问题。作为BN的另一个有趣的应用，Li等人[11]发现BN可以通过重新计算目标域中的流行统计数据来缓解域的转移。最近，人们提出了几种替代性的归一化方案，以将BN的有效性扩展到递归架构[12][13][14][15][16][17]。

2.2.2. 实例归一化(Instance Normalization, IN)

在最初的风格迁移方法[18]中，风格迁移网络在每个卷积层之后包含一个BN层。此外，Ulyanov等人[19]发现，不进行网络结构的更改，仅仅通过将BN层取代为IN层就可以显著提高风格迁移效果：

$$IN(x) = \gamma\left(\frac{x-\mu(x)}{\sigma(x)}\right) + \beta \tag{5}$$

与BN层不同，这里的$\mu(x)$和$\sigma(x)$是对每个通道和每个样本独立进行跨空间维度的计算：

$$\mu_{nc}(x) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{nchw} \tag{6}$$

$$\sigma_{nc}(x) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(x_{nchw} - \mu_{nc}(x)\right)^2 + \epsilon} \tag{7}$$

另一个区别是，IN层在测试时的应用是不变的，而BN层通常是用群体统计来取代小批量统计。

2.2.3. 条件性实例归一化(Conditional Instance Normalization, CIN)

Dumoulin等人[20]提出了条件实例归一化(CIN)层，其不是学习单一的仿射参数γ和β，该层为每种风格s学习不同的参数γ^s和β^s：

$$CIN(x;s) = \gamma^s\left(\frac{\mathrm{x}-\mu(\mathrm{x})}{\sigma(\mathrm{x})}\right) + \beta^s \tag{8}$$

在不包含归一化层网络的基础上，包含CIN层的网络增加了$2FS$的额外参数，其中，

F代表网络中特征图的总数量[20]。并且，额外参数的数量随着样式的增加而增长，将上述方法应用于对更多种的样式进行建模将面临着巨量的额外参数，且不能在不更新模型的基础上适应新风格。

2.2.4. 解释实例归一化(Interpreting Instance Normalization, IIN)

尽管(条件)实例归一化获得了相当大的成果，但它们在风格迁移方面十分有效的原因仍然难以明了。Ulyanov等人[19]将IN的成功归功于它对内容图像的对比度的不变性。然而，IN是在特征空间发生的，因此它应该比像素空间的简单对比度归一化有更深远的影响。并且，IN中的仿射参数可以完全改变输出图像的风格。

众所周知，DNN的卷积特征统计可以捕捉图像的风格[2][3][5]。Gatys等人[2]将二阶统计量作为优化提升的目标，Li等人[5]发现匹配通道均值和方差等其他统计量，同样有利于风格迁移效果的提升。综合以上观察，本论文利用实例归一化将图像的特征统计即均值和方差进行归一化，从而得到风格的归一化形式。虽然DNN于[2][5]中充当了图像描述符，但本文认为生成器网络的特征统计也可以控制生成图像的风格。

2.2.5. 自适应实例归一化(Adaptive Instance Normalization, AdaIN)

由于IN将输入归一化为由仿射参数指定的单一风格，那么使用自适应仿射变换使其适应任意给定的风格很自然地成为接下来的研究方向。

本论文提出将IN扩展为自适应实例归一化(AdaIN)，通过内容输入x和风格输入y，并匹配调整x的通道的平均值和方差和y的均值和方差。并且，AdaIN没有可学习的仿射参数，而是根据风格输入自适应地计算仿射参数：

$$AdaIN(x,y)=\sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right)+\mu(y) \tag{9}$$

其中，本文只是用$\sigma(y)$对归一化的内容输入进行缩放，并用$\mu(y)$对其进行移动。与IN类似，这些统计数字是跨空间位置计算的。

AdaIN产生的输出将具有与风格图像相同的高平均激活的特征，同时保留了内容图像的空间结构。与[21]类似，笔触特征可以通过前馈解码器倒置到图像空间。这个特征通道的方差可以编码更微妙的风格信息，这些信息也被转移到AdaIN的输出和最终输出的图像中。

简而言之，AdaIN在特征空间中进行风格迁移，通过转移特征统计，特别是通道平均数和方差。本文的AdaIN层扮演着与[22]中提出的风格互换层类似的角色。由于AdaIN拥有着与IN近似的计算复杂度，因此在风格迁移模型中使用AdaIN来替代IN方法并不具有计算量上的增加。

2.3 引导滤波(Guided Filter)

引导滤波[1]通过参考一个引导图像的内容来生成结果图像，其中引导图像可以是输入图像本身，也可以是任何其他的图像。相比于双边滤波(Bilateral Filter)，引导滤波在图像边缘附近拥有更好的表现。引导滤波具有相当快的计算速度，并通过选取特定的引导图像，其能够获取更加丰富灵活的操作。此外，如图2-3所示，可以通过调整参数ϵ和滤波窗口尺寸w_k的数值来控制输出图像的滤波效果，以得到适当的引导残差图细节信息来输入

修正网络。

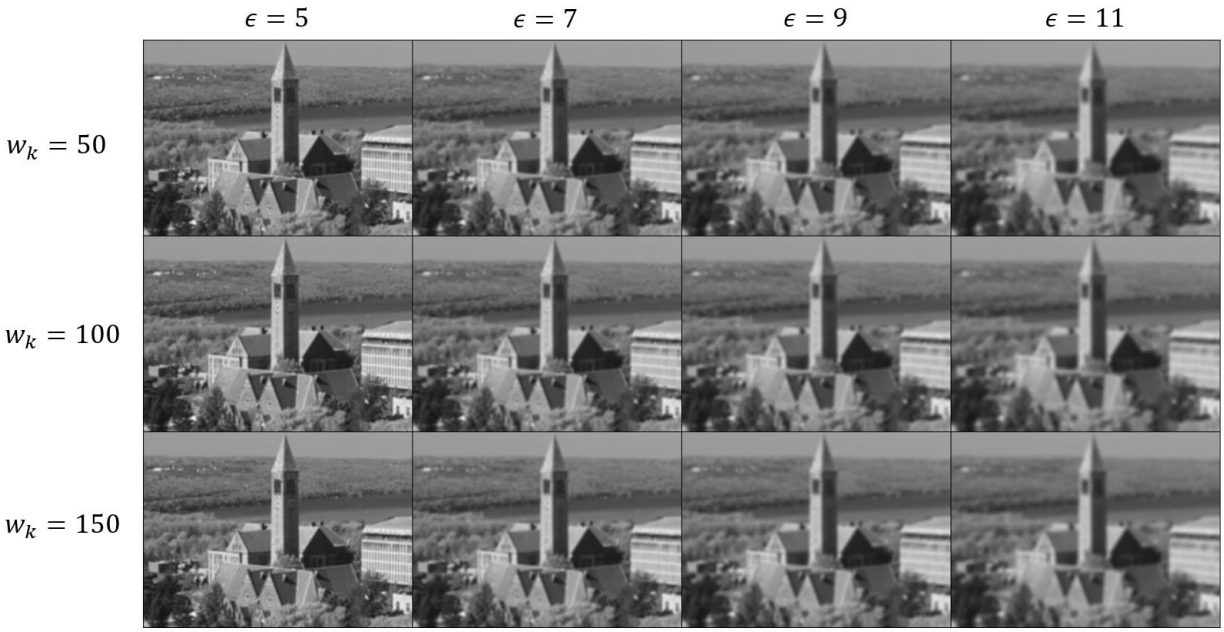


图2-3 引导滤波

总之，引导滤波在图像细节增强、图像降噪、羽化和去雾等应用中具有广泛的应用前景。它能够通过灵活调整参数，平衡去除图像噪声和保留细节之间的权衡关系。它的灵活性和效果使其成为图像处理领域中的重要工具之一。

2.3.1 算法概述

引导滤波首先定义一个一般的线性平移变量滤波过程，其包含一张引导图I，一张输入图p，以及一张输出图q。在本文中图I和图p是相同的内容图片，而在通常引导滤波方法中他们可以由使用者按照需求灵活指定。

在像素i处单的滤波输出可以表示为一个加权均值：

$$q_i = \sum_j W_{ij}(I) p_j \tag{10}$$

其中i, j为像素索引。卷积核W_{ij}是引导图I的函数，以及一个独立的变量p_j。这个卷积核与p是线性相关的。

引导滤波认为引导图I和输出图p可以被看作一个局部线性模型。假设q是一个以像素k为中心的窗口w_k上对I的线性变换：

$$q_i = a_k I_i + b_k, \forall i \in w_k \tag{11}$$

其中(a_k, b_k)是假定在w_k中为常数的一些线性系数。引导滤波使用一个边长为r的正方形窗口。由∇q=a∇I，该局部的线性模型确保只有当图像I拥有边缘信息时图像q才表现出边缘信息。此外，在抠图，图像超分辨率，以及图像去雾中，该模型被证明是有效的。

为了定义这些线性系数，对于公式(10)去最小化图像q和滤波输入图像p的差异。在窗口中最小化如下的损失函数：

$$E(a_k, b_k) = \sum_{i \in w_k} ((a_k I_i + b_k - p_i)^2 + \epsilon a_k^2) \tag{12}$$

其中ϵ是一个正则化参数去防止a_k变得过大。对公式(3)的解决被提出通过线性回归：

$$a_k = \frac{\frac{1}{|w|}\sum_{i\in w_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \epsilon} \tag{13}$$

$$b_k = \bar{p}_k - a_k \mu_k \tag{14}$$

其中μ_k和σ_k^2是在窗口w_k中图像I的均值和方差，$|w|$是在窗口w_k中的像素数量，$\bar{p}_k = \frac{1}{|w|}\sum_{i\in w_k} p_i$是图像p在窗口$w_k$中的均值。

之后，应用线性模型对于整张图片的所有窗口进行操作。然而，任意像素点i是包含在可能包含像素i的所有窗口中的，因此在公式(10)中q_i的值是不同的，当该像素处于不同的窗口中时。一个简单的策略是对所有可能q_i的值求平均。因此，在对所有窗口计算(a_k, b_k)后，计算滤波输出图像通过如下公式：

$$q_i = \frac{1}{|w|}\sum_{k:i\in w_k} (a_k I_i + b_k) \tag{15}$$

$$= \overline{a}_i I_i + \overline{b}_i \tag{16}$$

其中$\overline{a}_i = \frac{1}{|w|}\sum_{k\in w_i} a_k, \overline{b}_i = \frac{1}{|w|}\sum_{k\in w_i} b_k$。

经过这样的修改，∇q不再只是∇I的缩放，因为线性系数$(\bar{a}_i, \bar{b}_i)$在空间上变化。但是因为$(\bar{a}_i, \bar{b}_i)$是平均值的输出，他们的梯度应该小的多在图像I的强边缘附近。在这种情况下，仍然有$\nabla q \approx \bar{a}\nabla I$，这意味着I图像突然的强度改变能够极大程度的在图像q中保留。

通过公式(12)、(13)、(14)，指出图像I,p,q之间的相关性，它们以图像滤波(9)的形式存在。事实上，公式(12)中的a_k可以被重写为p的权重和形式：$a_k = \sum_j A_{kj}(I)p_j$,$A_{ij}$是仅取决于图像I的权重。同理，由公式(13)、(15)可得：$b_k = \sum_j B_{kj}(I)p_j$, $q_k = \sum_j W_{kj}(I)p_j$。核权重可以明确表示为：

$$W_{ij} = \frac{1}{|w|^2}\sum_{k:(i,j)\in w_k}\left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \epsilon}\right) \tag{17}$$

进一步得计算表明$\sum_j W_{ij}(I) = 1$。因此，不需要对权重进行归一化操作。

2.4 内容感知特征重组(Content-aware reassembly of features, CARAFE)

特征上采样是许多现代卷积神经网络中的关键操作(如特征金字塔)，其在语义分割和目标检测等任务中具有重要地位。其中包括一种通用的、高效的轻量级运算符内容感知特征重组(CARAFE)上采样方法。其具有如下优势：

(1) 更大的感受野。相较于传统的利用亚像素邻域的上采样方法(如双线性插值)无法捕获密集预测任务所需的丰富语义信息，本方法可以在更大的感受野内提取上下文的特征信息。

(2) 内容感知处理。传统的反卷积操作在整个图像上应用相同的内核，而忽略底层内容的变化，这限制了反卷积运算响应局部变化的能力。CARAFE与传统的反卷积操作不同，其对不同的样本使用不同的内核，即启用特定于实例的内容感知操作，从而即时生成自适应内核。

(3) 高速计算和轻量级。相较于相同感受野下反卷积操作使用的大内核带来的高计算量，CARAFE的计算开销很小，可以很方便的集成到现代神经网络架构中。

CARAFE由两个步骤组成。第一步是按照每个目标位置的内容预测一个重组内核，第二步是用预测的内核来进行特征重组。给定大小为$C \times H \times W$ 的特征图 X 和上采样率 σ(假设 σ 为整数)，CARAFE 将生成大小为 $C \times \sigma H \times \sigma W$ 的新特征图 X'。对于输出特征向量 X' 的任意目标位置 $l' = (i', j')$，在输入特征向量X 处都有对应的源位置 $l = (i, j)$，其中$i = \lfloor i'/\sigma \rfloor$, $j = \lfloor j'/\sigma \rfloor$。这里将 $N(X_l, k)$ 表示为以位置l 为中心的X的$k \times k$子区域，即X_l的邻域。

在第一步中，内核预测模块 ψ 根据X_l 的邻居为每个位置l' 预测位置合适的内核 $W_{l'}$，如等式 (1)所示。 重组步骤公式化为 Eqn.(2)，其中 φ 是内容感知重组模块，它将X_l 的邻居与内核 $W_{l'}$ 重组：

$$W_{l'} = \psi\big(N(X_l, k_{encoder})\big). \tag{18}$$

$$X_{l'} = \emptyset\left(N(X_l, k_u p), W_{l'}\right). \tag{19}$$

2.5 本章小结

在本章中本文介绍了图像风格迁移的理论基础，包括模型框架以及模型核心的两个算法自适应实例归一化(AdaIN)与引导滤波。并描述了该算法的内容与实现细节。此外本文介绍了该模型使用到的诸多经典算法以及网络结构，如各种归一化方法、VGG网络以及基于内容感知重组的上采样方法等。

在下一章，本文将详细介绍本文模型的具体实现方式(基于自适应实例归一化的草稿网络、基于引导滤波的修正网络)，并提供模型损失优化项的计算方法以及结果参数。

第 3 章 基于 AdaIN 的快速风格迁移方法

在这个章节，我们将介绍本文的核心网络模型。本文使用自适应实例归一化(AdaIN)作为风格迁移和特征融合的核心算法，并在其基础上提出我们的快速风格迁移方法GuideStyle。我们的GuideStyle模型启发于常见的绘画过程，其由两部分组成，草稿网络(Drafting Network)用来在低分辨率下迁移全局风格信息，修订网络(Revision Network)用来在高分辨率下填充局部内容、风格信息。下面，本文将介绍GuideStyle快速风格迁移模型的具体实现细节。

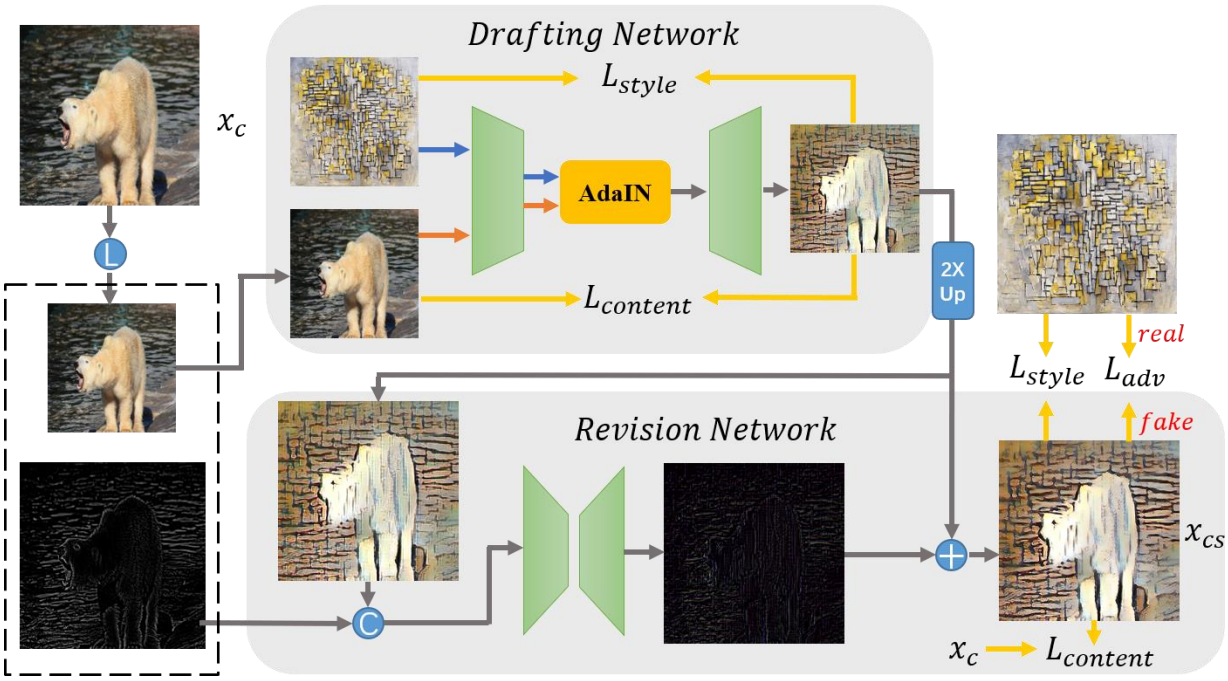


图3-1 GuideStyle网络结构概要

为了便于理解且由于训练设备所限，本文仅使用单层修订网络来修订图像细节，如图3-1所示。此外，基于简单的拉普拉斯上采样方式，对于更多层的修订网络的追加是容易的。

3.1 网络结构

本文的GuideStyle网络将一张任意内容图像$x_c \in R^{H_c \times W_c}$和一张任意风格图像$x_s \in R^{H_s \times W_s}$作为输入，并最终合成一张风格化图像$X_{cs}$。如图3-1所示，在前向过程中，$\bar{x}_c$是$x_c$的一倍下采样图像。$r_c$为保留内容图像$x_c$高频信息的残差图，$r_c = x_c - G(x_c, x_c)$，其中G为引导滤波函数，$G(x_c, x_c)$是使用$x_c$作为引导图对$x_c$进行引导滤波后的平滑图像。同时，风格图像$x_s$也进行了一倍下采样操作生成低分辨率图像$\bar{x}_s$。

在第一阶段，草稿网络分别对低分辨率风格图像$\bar{x}_s$和内容图像$\bar{x}_c$进行编码(编码器是预训练的)，之后通过风格特征在多尺度下调制内容特征向量，并逐级输入解码器网络生成风格化结果图像$\bar{x}_{cs} \in R^{H_c/2 \times W_c/2}$。在第二阶段，修订网络首先对低分辨率风格化图像$\bar{x}_{cs}$进行一倍上采样为$x_{cs}{}' \in R^{H_c \times W_c}$。之后，对$x_{cs}{}'$与$r_c$在C维进行连接(concatenates)操作$\{x_{cs}{}', r_c\}$。最后，将$\{x_{cs}{}', r_c\}$输入修订网络来生成一个风格化的细节修复图像$x_{cs} \in R^{H_c \times W_c}$。

其中，在一个金字塔式上采样过程中，修订操作可以使用多次：$x_{cs} = A(\bar{x}_{cs}, r_{cs})$。

下面，本文将详细的介绍草稿网络和修订网络的实现细节。

3.2 基于 AdaIN 的草稿网络

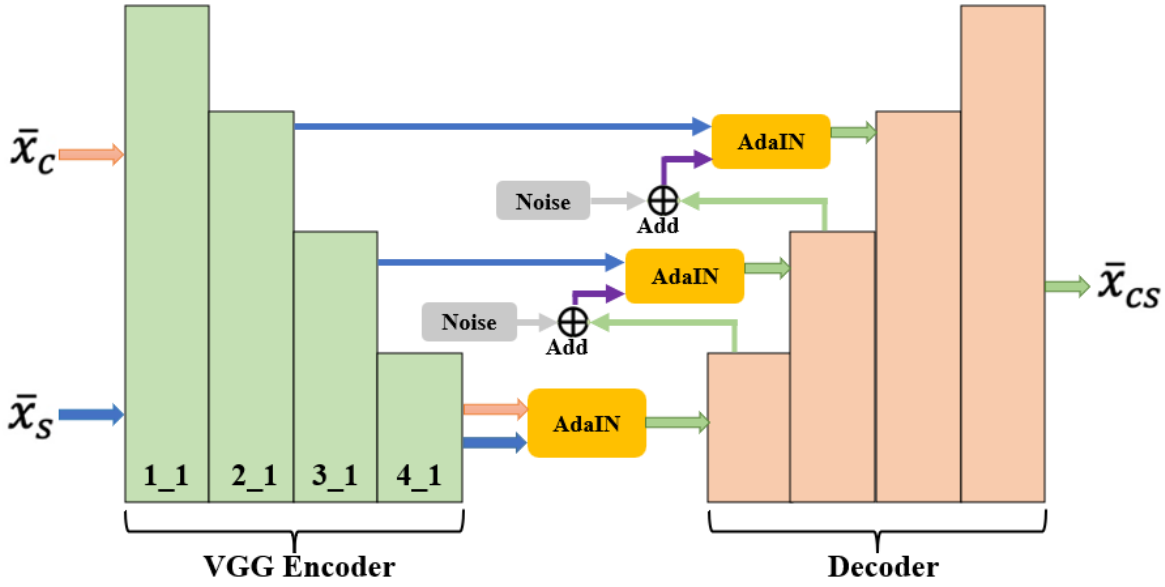


图3-2 草稿网络结构概要

(1) 解码器是一个预训练的VGG-19网络，在训练中其权重是锁定的。给定$\bar{x}_c$与$\bar{x}_s$，VGG编码器分别在2_1,3_1,4_1层进行多尺度特征提取。

(2) 本文使用AdaIN方法首先对4_1层的风格和内容特征进行调制。其后，对2_1,3_1层引入与特征向量相同尺寸的随机噪声，并使用AdaIN方法对加噪后的特征向量和风格图像进行进一步调制。

(3) 最后，通过跳跃连接将调制后的特征向量传入解码器对应的间隔尺度中。使用从低到高多层的基于AdaIN的跳跃结构来帮助保存内容结构同时融合风格特征，其对于低分辨率图像来说尤其有效。

如图3-2所示，草稿网络采用了一个比较简单的Unet + 多尺度AdaIN的结构，其将一张任意的内容图像c和一张任意的风格图像s分别输入解码器网络，并通过多尺度AdaIN调制成一个将前者的内容和后者的风格重新组合的风格化图像。草稿网络的首要目的是最大限度的在整体上迁移全局的图像风格。本文根据在低分辨率输入图像下同样卷积核尺寸和深度的卷积网络拥有相对于高分辨率输入图像更大的感受野，因此理论上图像的全局特征在低分辨率下能够被更好的识别和迁移。此外，根据本文实验可以说明上述理论在实际操作中是可行的同时在相同实验环境下，训练低分辨率图像意味着可以使用更大的batch_size，使得网络能够更快速的收敛到较优区间。

由于该网络目标为实现任意风格下对任意内容的风格迁移，因此为了更好的结合风格特征和内容特征，本文采用一个简单的编码器-解码器架构，并使用多尺度AdaIN方法对多维度特征进行融合来获得一个较优的风格迁移效果。通过实验发现，在通常情况下对于局部细节较多的即平滑度较差的内容图像，网络反而实现了较优风格迁移效果。故本文受此启发，在每次进行AdaIN操作之前加入了随机噪声，其中噪声是对图片逐像素添

加的，为网络提供了更加丰富的灵活性。同时增加随机噪声的程度即噪声权重是可学习的，由网络在训练过程中通过学习来自主决定，其初始权重为0。

此外，因为草稿网络为全卷积结构，理论上其可以接受任意尺寸的输入图像，同时，通过实验我们发现对于高分辨率图像，其仍拥有较优的风格迁移质量。但是，由于网络使用DCNN带来的高复杂性，网络在高分辨率图像的风格迁移具有速度不够理想的缺陷，故在不蒸馏网络参数或缩减网络层数的情况下，修正网络对于加速超分辨率图像风格迁移仍是必要的。

3.2.1 解码器详述

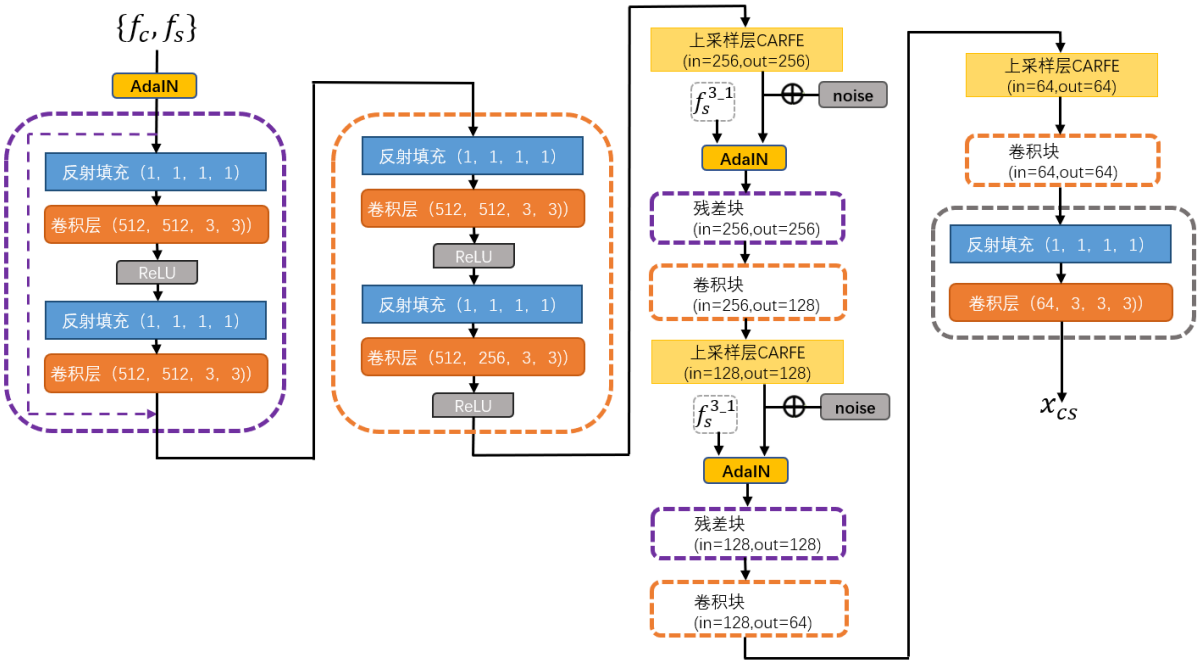


图3-3 解码器网络

如图3-3所示，解码器主要反映了编码器的情况，同时为了便于跳跃连接的需要，解码器的尺度结构与编码器是一一对应的。在使用解码器架构的同时，为了使用更深的网络结构以提取更复杂的语义信息，本文在每个卷积块前添加了残差块(ResNet Block)[24]以防止DCNN网络退化现象的发生。此外本文的网络设计与VGG网络相似，其组成是非常模块化的，每个残差块由两个反射填充层、两个卷积层与一个激活函数ReLU层组成，卷积块与残差块类似，仅在残差操作部分将残差运算更改为一个ReLU激活函数运算。并且在卷积核大小的选取上，本文依旧依据VGG网络的经验，均使用了3*3的卷积核。

同时，所有的池化层被CARAFE[24]层所取代，以减少棋盘效应。相较于AdaIN论文中使用的传统的利用亚像素邻域的上采样方法，即最近邻上采样方法，CARAFE拥有更大的感知域，能够更好的利用到特征图的语义信息。同时，CARAFE对不同的样本使用不同的内核，即时生成自适应内核，其具有更高的灵活性。此外，CARAFE在兼具上述优点的同时，仍然保持轻量化特质，使得网络可以达到快速风格迁移。

本文在编码器和解码器中均使用反射填充以避免边界伪影。有关解码器的另一个重要的架构抉择是是否应该使用实例、批量或无归一化层。在这里，本文尊崇AdaIN论文中

所提及的，当希望解码器生成风格迥异的图像时，实例归一化(IN)和批量归一化(BN)这两种方法都是不可取的，因为其更多的将样本归一化为单风格中心。因此，本文在解码器中不使用归一化层。

3.3 修订网络

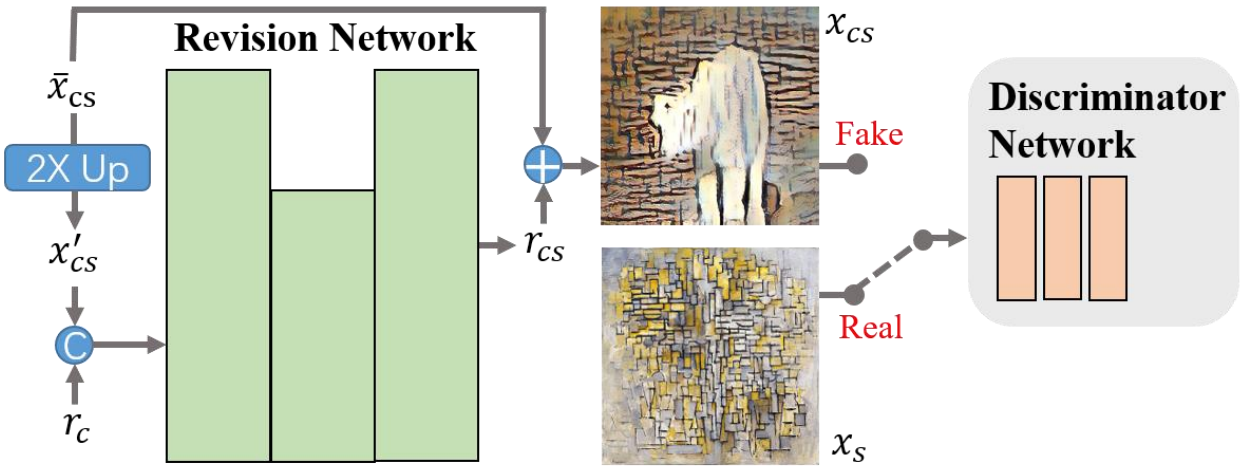


图3-4 修订网络概要

修订网络的目标是修订粗糙的风格化图像，通过结合粗糙风格化图像$\bar{x}_{cs}$和引导滤1波操作后的内容残差图r_c来生成一张残差细节图r_{cs}。最终的风格化图像为r_{cs}和上采样后的$\bar{x}_{cs}$图像的加和，这样做的目的是确保$\bar{x}_{cs}$图像的全局风格特征被保留。同时，通过残差细节学习局部风格细节而非整体内容对于轻量化的修订网络更加容易。

如图3-4所示，修订网络被设计为一个简单的但有效的编码器-解码器结构，其仅仅具有一层上采样和一层下采样。更进一步的，本文引入了一个补丁判别器网络帮助修订网络在对抗性学习下进行更精细的纹理块捕捉。

本文的补丁判别器网络D模仿SinGAN[1]的结构进行设计，其中网络D仅含有5个卷积层和32个隐藏通道数。本文选择定义一个相对浅的网络D是为了：(1)预防过拟合(2)控制感受野确保D仅仅能捕获局部纹理信息。

3.4 模型训练

在该项目中，本文使用了COCO和WikiArt数据集作为训练集。COCO是一个包含超过33万张图像的大规模图像数据集，其一共具高达250万个物体实例以及80个不同的物体类别。WikiArt是一个集合了超过20万张不同风格和流派的艺术作品的高质量图像数据集，包括古典、现代和当代艺术。

这两个数据集在风格迁移算法中的使用主要是为了提供丰富和多样化的样式和纹理，以及增强模型的泛化能力和鲁棒性。具体来说，COCO提供了丰富的物体和场景图像，可以用于学习和捕捉图像的内容特征；而WikiArt提供了多样化和个性化的艺术作品图像，可以用于学习和捕捉图像的风格特征和纹理信息。

每个数据集大约包含80，000个训练实例。本文使用Adam优化器[26]和8个内容-风格图像对的批量大小。在训练过程中，本文首先将输入图像的像素尺寸缩放为512*512像素分辨率，同时预留输入图像的原始长宽比，之后本文对缩放后图像随机裁切尺寸为256×

256像素面积的区域。此外，由于本文的模型是完全卷积的，因此在训练、测试过程中，它的输入图像可以是任意尺寸的。

在训练过程中，草稿网络损失L_{Draft}和修正网络损失L_{rev}分别为：

$$L_{Draft} = l_p + \alpha \cdot l_m + \beta \cdot \mathrm{l_r} + \gamma \cdot \mathrm{l_i} + \delta \cdot l_{tv} \tag{1}$$

$$L_{Rev} = L_{adv} + \beta \cdot \mathrm{L_{base}} \tag{2}$$

草稿网络和修订网络均使用内容损失和风格损失来优化。因此本文先介绍风格损失和内容损失，之后再介绍两个网络的具体训练过程。

3.4.1 风格损失

本文使用了较为普遍的均方差损失作为风格损失。首先，当一张图片输入编码器(VGG19)网络时，会得到一系列特征向量$F = [F^{1_1}, F^{2_1}, F^{3_1}, F^{4_1}, F^{5_1}]$, 其中均方差损失计算公式为：

$$\mathrm{l_m} = \|\mu(\mathrm{F_s}) - \mu(\mathrm{F_{cs}})\|_2 + \|\sigma(\mathrm{F_s}) - \sigma(\mathrm{F_{cs}})\|_2 \tag{3}$$

其中μ和σ分别为特征向量的均值和方差运算。

3.4.2 内容损失

对于内容损失本文采用归一化后的感知损失，输入为$F_c \in R^{h_c w_c \times c}$和$F_{cs} \in R^{h_{cs} w_{cs} \times c}$。其中，由于$x_c$和$x_{cs}$是同尺寸的，因此 $h_{cs} = h_c, w_{cs} = w_c$。感知损失的定义如下：

$$l_p = \|\mathrm{norm}(F_c) - \mathrm{norm}(F_{cs})\|_2 \tag{4}$$

其中，这里的norm表示F的通道归一化。

3.4.3 草稿网络的损失

由实验表明，总变分数值可以用来评价图像的受噪声污染程度，其定义如下：

$$J_{T_0}(\mathrm{u_0}) = \int_{D_u} \sqrt{u_x^2 + u_y^2}\, dxdy \tag{5}$$

其中，$\mathrm{u}_x = \frac{\partial u}{\partial x}$, $\mathrm{u}_y = \frac{\partial u}{\partial y}$，$D_u$为定义域。

因此为了增强生成图像的空间平滑性，本文引入离散化的全变分损失(The total variation loss)，其计算公式如下：

$$l_{tv} = \sum_{i,j} \left(\left(x_{i,j-1} - x_{i,j}\right)^2 + \left(x_{i+1,j} - x_{i,j}\right)^2 \right)^{\frac{\beta}{2}} \tag{6}$$

同时，为了更好的提取风格图片的风格信息，本文引入rEMD损失去评估风格图片x_s和风格化图片x_{cs} 在特征分布上的距离。便于描述，如下叙述本文省略了特征向量的索引层上标。假设$F_s \in R^{h_s w_s \times c}$, $F_{cs} \in R^{h_{cs} w_{cs} \times c}$是$x_s$和$x_{cs}$的特征向量，它们的rEMD损失可以被计算为：

$$l_r = \max\left(\frac{1}{h_s w_s} \sum_{i=1}^{h_s w_s} \min_j C_{ij}, \frac{1}{h_{cs} w_{cs}} \sum_{j=1}^{h_{cs} w_{cs}} \min_i C_{ij} \right) \tag{7}$$

其中余弦距离项C_{ij}被定义为：

$$C_{ij} = 1 - \frac{F_{s,i} \cdot \mathrm{F_{cs,j}}}{\|F_{s,i}\| \|F_{cs,j}\|} \tag{8}$$

此外，为了防止网络自主修改图像色调，使生成图片整体颜色出现变化，本文引入Identity损失，其计算过程为：

$$l_{i1} = \|\mathrm{norm}(\bar{x}_{\mathrm{cc}}) - \mathrm{norm}(\bar{x}_{\mathrm{c}})\|_2 \tag{9}$$

其中，$\bar{x}_{\mathrm{cc}}$为将风格图片设置为与内容图片相同，网络输出的风格化图像。

$$l_{i2} = \|\mathrm{norm}(F_c) - \mathrm{norm}(F_{cc})\|_2 \tag{10}$$

其中，F_{cc}为将$\bar{x}_{\mathrm{cc}}$输入编码器后的特征输出，$F_{cc} \in R^{h_{cc}w_{cc}\times c}$。

最后，Identity损失被定义为：

$$l_i = l_{i1} + 0.02 * \mathrm{l_{i2}} \tag{11}$$

在草稿网络的训练阶段，低分辨率图像$\bar{x}_c$和$\bar{x}_s$不仅被当作网络的输入，还分别被用来评估风格损失和内容损失。起草网络的整体训练目标函数定义为:

$$L_{Draft} = l_p + \alpha \cdot l_m + \beta \cdot \mathrm{l_r} + \gamma \cdot \mathrm{l_i} + \delta \cdot l_{tv} \tag{12}$$

其中λ,α和β是权重参数。本文控制内容和风格的损失通过权重α，β。此外，$\mathrm{l_r}$被定义在3_1和4_1层，同时，l_m、l_p和l_{i2}被定义在1_1至5_1层。

另外，尽管加入全变分损失会牺牲一定的图像清晰度和图像细节，但考虑到草稿网络的主要功能在于生成图像整体结构，因此本文认为图像的空间平滑性是优先考量的，在加入全变分损失后，其可以显著减弱图像的边缘伪影。

3.4.4 修订网络的损失

在修订网络的训练阶段，草稿网络的权重被固定，训练损失被设置在x_{cs}上。由于修订网络仅用来对图像局部细节进行修改，并可以引用金字塔结构进行扩展来生成超高分辨率图像，因此本文并没有使用rEMD作为风格损失，其目的在于减少在高分辨率图像训练时带来的高昂显存成本。为了更好的学习局部细节纹理，除了基础的内容和风格损失$L_{base} = l_p + \alpha \cdot l_m$，本文引入了判别器网络D，并通过对抗性损失来训练修订网络。整体优化目标被设置为：

$$\min_{Rev} L_{base} + \beta \cdot \min_{Rev} \max_{D} L_{adv}(Rev, D) \tag{13}$$

其中Rev代表修订网络，D代表判别器网络，权重β被用来控制基础风格迁移损失和判别器损失之间的平衡。L_{adv}是标准的对抗性损失。

3.4.5 优化和微调

对于两个网络的训练，本文均使用Adam优化器，学习率初值为1e-4，同时本文根据模型训练轮数对学习率进行衰减，以达到更好的收束效果，学习率衰减值为5e-5。此外，本文设置每轮训练的批大小(batch size)为4。对于草稿网络，训练过程包含50000次迭代，损失权重$\alpha = 3$，$\beta = 3$，$\gamma = 4$，$\delta = 0.4$，其训练细节如图3-6所示。

本文可以通过修改修订网络的风格权重和内容权重来修改图像细节侧重，并且由于网络的感受野限制，因此本文可以通过对多层引导金字塔修订网络使用不同的训练权重来达到更好的效果。如图3-5所示，修正网络的训练过程包含15000轮迭代，损失权重$\alpha = 2$，$\beta = 1$，保持了风格与内容的统一。

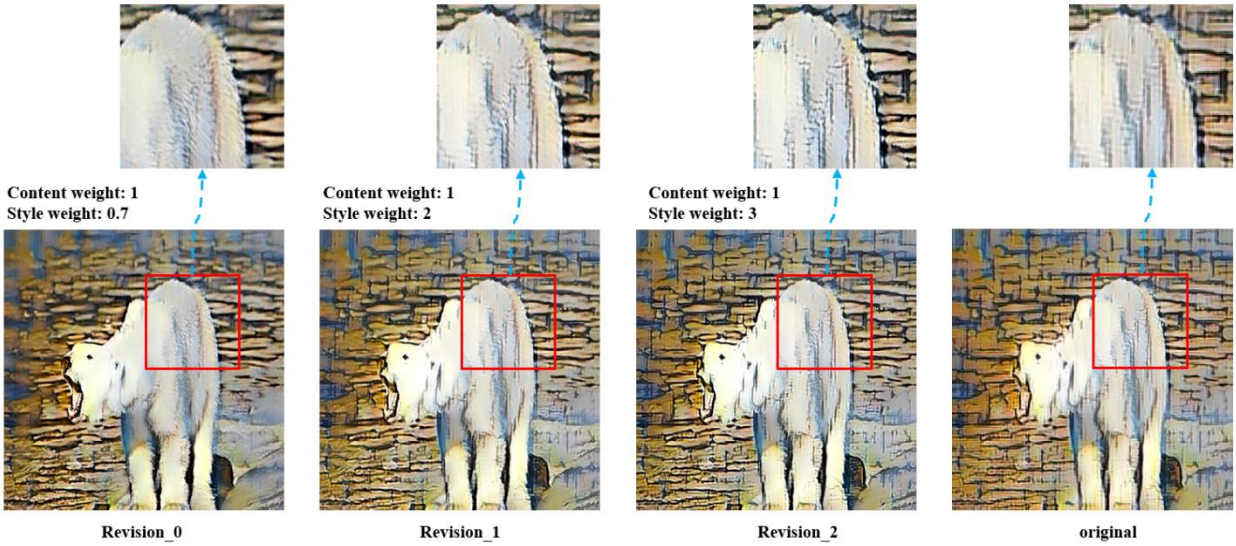
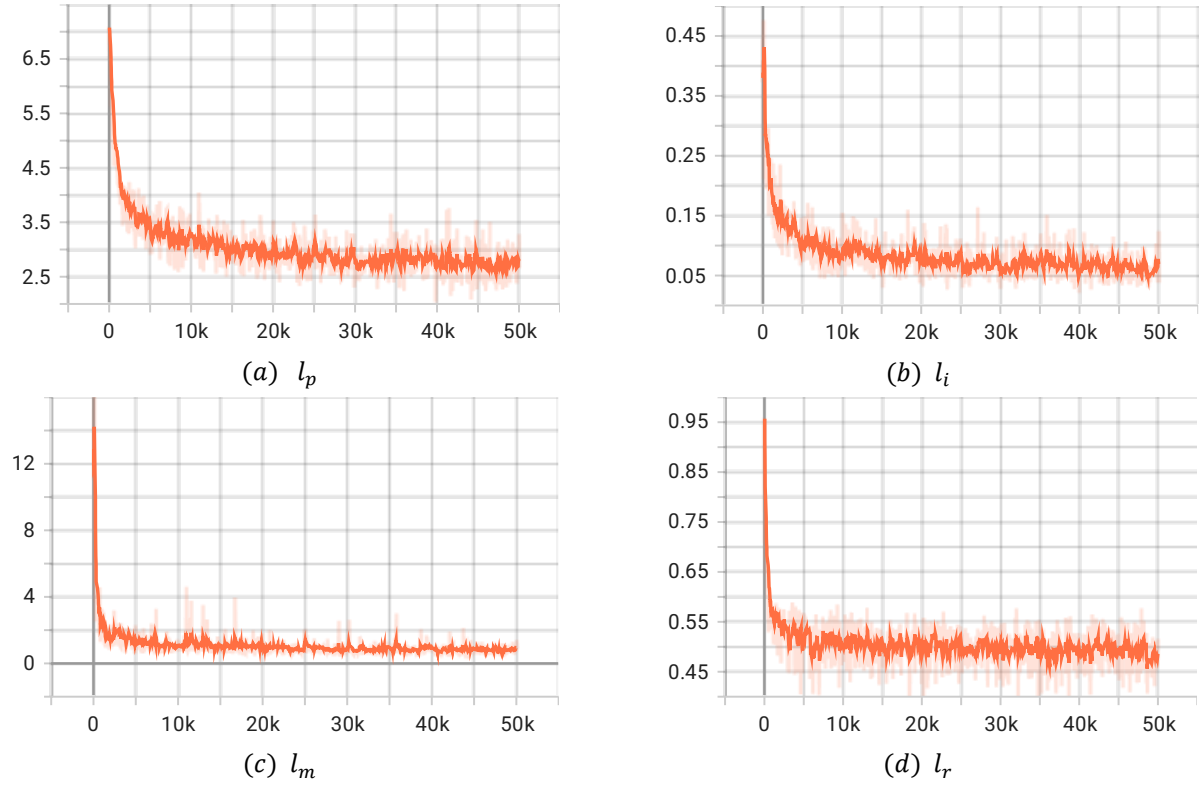


图3-5 修订网络微调



(a) l_p (b) l_i

(c) l_m (d) l_r

图3-6 草稿网络损失

3.5 本章小结

在本章中介绍了基于AdaIN与引导滤波的风格迁移系统GuideStyle的详细实现过程，以及具体介绍了草稿网络与修订网络的网络结构，并说明了设计思路与灵感来源。此外，本章系统介绍了本文模型GuideStyle的训练细节，包含风格损失、内容损失、Identity损失、rEMD损失等等。并通过调整总体优化目标中的各项超参数，以达到对风格、内容的权衡，最终达到了良好的风格迁移效果。

第 4 章 实验结果与分析

4.1 风格迁移图像性能评价标准

4.1.1 峰值信噪比(Peak signal-to-noise ratio，PSNR)

峰值信噪比用来计算原始图像与压缩图像之间的质量差异，其值越高意味着压缩后图像的质量越好。峰值信噪比在图像压缩领域中受到广泛应用，其能很好的评估压缩后图像的信号重建质量。在峰值信噪比的计算过程中使用了均方差(Mean Square Error，MSE)作为基础衡量标准，其通过计算两图像的累积平方误差来衡量图像质量差异，值越低，误差就越小。均方差定义如公式(1)所示：

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{1}$$

其中， I和K为两张m×n尺寸的单色图像，I为K的噪声近似。

峰值信噪比定义如公式(2)所示：

$$PSNR = 10log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{2}$$

其中，MAX_I^2为图像中单像素的颜色数值表示的最大值，若每个采样点用 8 位二进制表示，则$MAX_I^2 = 255$ 。

4.1.2 结构相似性(Structural similarity，SSIM)

SSIM(结构相似指数)是一种常用于评估图像相似度和感知质量的指标。通过分析图像的对比度、结构特征和亮度，SSIM能够捕捉到人眼感知中重要的视觉差异。在本项目，SSIM可以用来衡量风格化图像与风格图像之间的相似程度，以评估模型的风格迁移水平和迁移质量。

亮度是指图像的整体明暗程度，而对比度则描述了图像中不同区域之间的明暗差异。结构则反映了图像中物体的形状、纹理等特征。SSIM通过综合考虑这些因素，提供了一个全面的图像相似度度量。

在风格迁移任务中，我们希望保留原始图像的结构信息，同时将其与另一个图像的风格进行融合。SSIM可以作为一个指导，帮助我们评估生成图像与目标图像之间的相似度，并优化风格迁移算法以提高感知质量。SSIM值的范围介于0到1之间，值越大表明图像越相似。SSIM公式如(3)所示。

$$SSIM = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{3}$$

其中，$l(x,y)$、$c(x,y)$、$s(x,y)$分别代表两幅图像进行亮度、对比度、以及结构的比较，公式如(4)-(6)所示：

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{4}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{5}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{6}$$

μ_x为x的均值、μ_y为y的均值、σ_x^2为x的方差、σ_y^2为y的方差、σ_{xy}为x和y的协方差，c1、

c2、c3均为常数。

将上述公式代入，可得：

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{7}$$

在图像整体SSIM的计算过程中，在图片中选取合适的N×M的窗口，并逐步滑动窗口直至覆盖整张图像，分别计算每次窗口内的SSIM数值，最后对全部的部分SSIM数值做均值计算得到图像整体SSIM数值。其运算过程如图4-1所示。

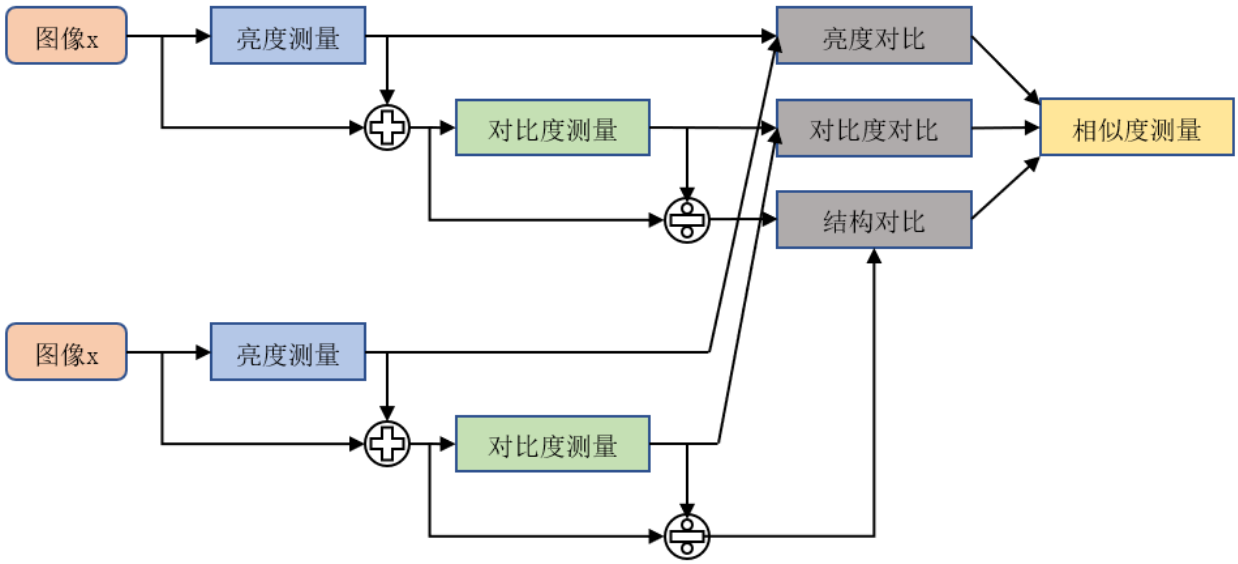


图4-1 SSIM运算过程图

4.2 与经典算法比较

Model	PSNR	SSIM
GuideStyle (Ours)	**61.6902**	**0.6070**
Drafting Network	61.2499	0.4682
AdaIN[1]	60.0263	0.3494

表 1 与内容图像对比

Model	PSNR	SSIM
GuideStyle (Ours)	59.1580	0.2854
Drafting Network	59.3043	**0.2998**
AdaIN[1]	**59.6222**	0.2611

表 2 与风格图像对比

如表 1、表 2所示，可以看出在风格参数相差不大的情况下，本文的模型对于内容信息的保留具有显著提高，在图像性能评价标准PSNR和SSIM中，都显示出较优的水准。同时，在草稿网络中，其在低分辨率下对图像风格的保留同样具有很高的水准，在与风格图片的对比中，其SSIM数值0.2998为三个网络模型的最高值，并在对内容信息的保持中，PSNR数值高达61.2499，与SSIM评价均高于AdaIN模型的数值。

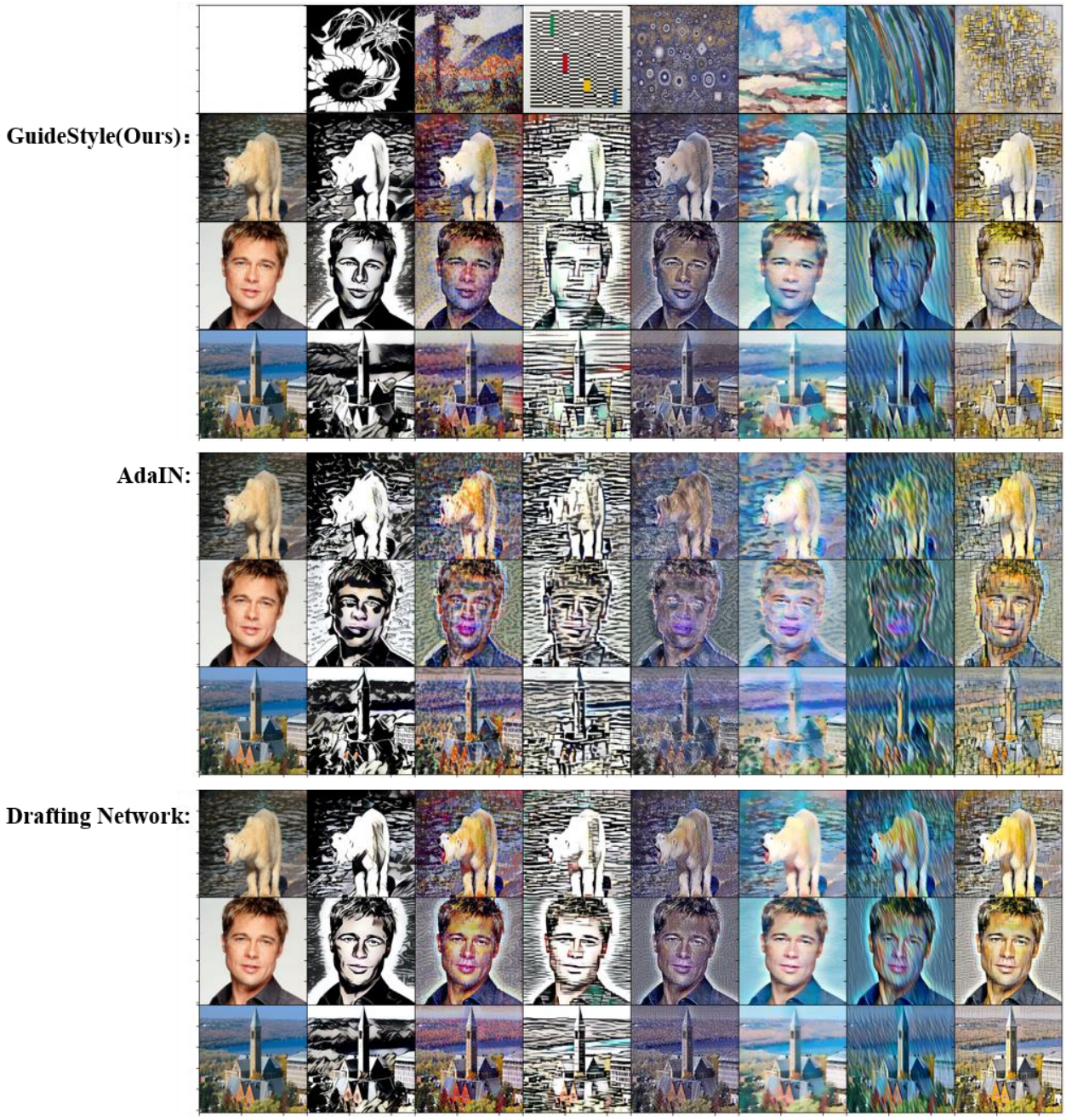


图4-2 本文的方法与AdaIN模型生成图片的质量比较(512*512px)

如图4-2所示，本文比较了GuideStyle、AdaIN、Drafting Network三个AST模型的图像风格迁移效果。其中，GuideStyle在风格迁移和内容保留上均显示出了最优的效果。

从示例中可以看出，GuideStyle生成的图像具有更加全局的风格纹理与结构，其极大程度的迁移了风格图像的笔触分布信息，如在第三张风格图像中本文的模型对于红、蓝、黄色块的区域分布感知更加优秀，其能较明显的在风格化图像中表现。相反在AdaIN模型与512分辨率下的草稿网络中，区域风格信息并没有被较好的展现出来，其风格化图像中仅具有复杂黑色网格纹理，而缺少特定区域的结构信息。

此外，与AdaIN模型进行对比，我们可以看出尽管Drafting Network的风格纹理迁移的更加局部，但其在保留内容图片信息的同时，迁移的风格化纹理特征仍然十分清晰整洁。而AdaIN模型生成的风格化图像不仅极大的扭曲了内容图像的内容信息，其迁移的风格纹理也具有很大的边缘伪影以及扭曲模糊等现象。同时，AdaIN模型的风格化图像与原风格

图像色彩的色相与明度并不高度统一，其大部分风格迁移后图像颜色更加厚重，且会出现炫光等现象。而本文的模型通过引入Identity损失使得风格化图像在颜色还原上具有很好的表现。

在迁移速度上，对于512分辨率的图像，AdaIN与Drafting Network模型的迁移速度均在0.015s左右，而我们的GuideStyle可以以接近实时的速度进行图像风格迁移。但值得注意的是，GuideStyle模型虽然在迁移速度与整体迁移质量上显著高于其他两个模型，但其风格化的图片的色差深度表现却略逊色于Drafting Network模型的结果，本文认为其与修正网络结构较为简单有关，因此今后对修正网络进行进一步研究仍需进行。

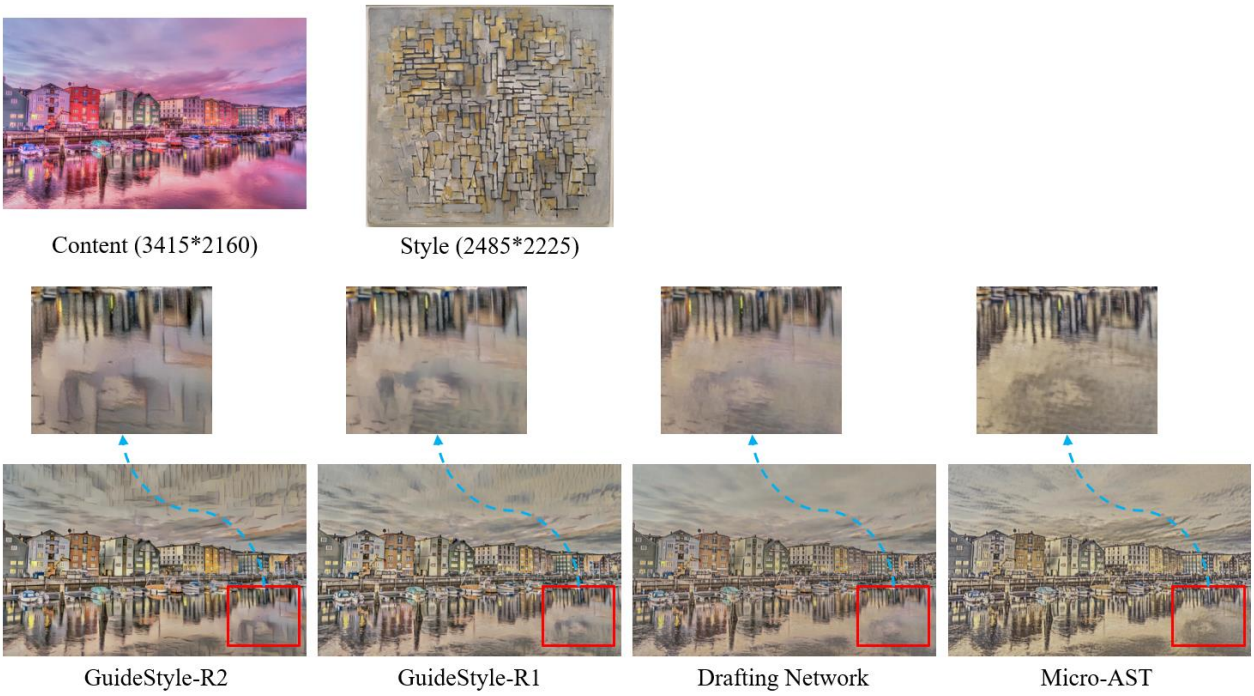


图4-3 超分辨率图像风格迁移

Model	运行时间	GPU存储占用	GFLOPS
GuideStyle-R2	0.2290	5.2269	409
GuideStyle-R1	0.3650	6.2606	1636
Drafting Network	0.9205	13.9292	6543
Micro-AST[33]	0.1775	3.2553	286

表 3 模型迁移速度与资源消耗

其中，GuideStyle-R2为具有两层上采样金字塔结构修订网络(Revision Network)的GuideStyle模型，同理，GuideStyle-R1为具有单层上采样金字塔结构修订网络(Revision Network)的GuideStyle模型。

在超分辨率图像风格迁移上，如图4-3和表 3所示，对于相同的内容图像，GuideStyle-R2模型拥有与目前行业内较新的超分辨模型Micro-AST[32]近似的风格迁移速度和GPU存储占用率。并且通过与GuideStyle-R1模型对比，可以看出堆叠更多的修正网络能够显著的减少网络计算速度和资源消耗。此外，我们GuideStyle-R2模型在拥有高速度的同时，通过设置多层修正网络来降低草稿网络输入图像分辨率，使得草稿网络能够获得更大的相对感受野以捕获风格图像的全局纹理信息。由实验结果证明，该种方法是有

效且适用的。

因此，综上所述，本文提出GuideStyle方法与传统风格迁移模型相比拥有较大改进，其可以高速的生成高质量风格化图像，并能够应用于多种现实场景。

4.3 与未引入噪声的草稿网络效果比较

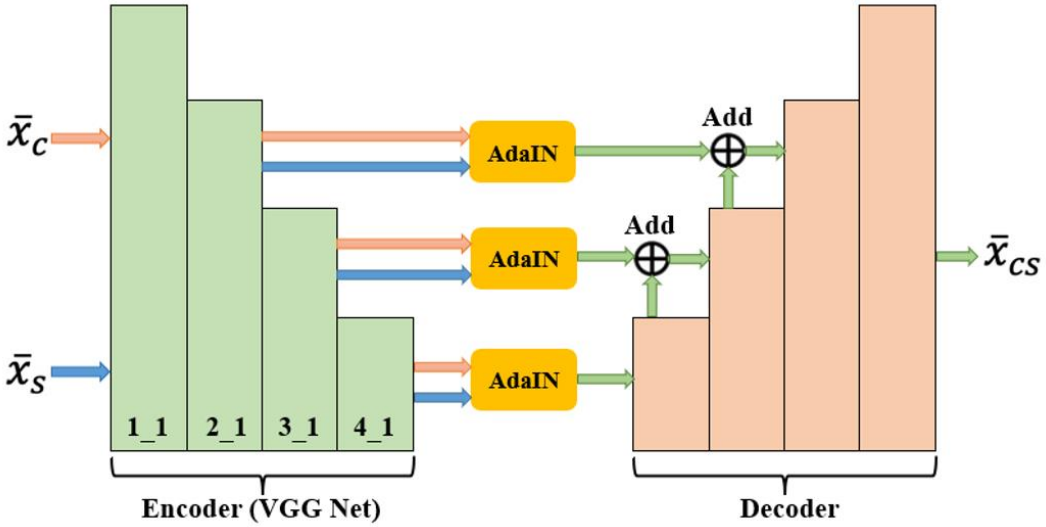


图4-4 未引入噪声草稿网络结构

图4-5 与未引入噪声的草稿网络效果比较

在项目初始阶段下，本文借鉴了[32]中的解码器结构，如图4-4所示，在该论文中

AdaIN是作为跳跃结构在中间层以残差的方式引入到解码器网络中的，与传统AdaIN模型的解码方式相比其虽然也能做到较好的融合风格信息(在章节4.3中提及)，但对于任意风格迁移来说，该网络对于不同风格的迁移仍缺少一定的灵活性。

因此，本文受到传统自动编码器的启发，在跳跃结构部分引入了噪声图(在章节3.2中提及)，以获得对不同风格的风格迁移更高的网络灵活性。如图4-5所示，本文看到在加入噪声图之后的网络，其在更好的保留图像内容信息的同时，能够进行更高质量的图像风格迁移，且对于不同的风格图像，对风格图像纹理的迁移效果更加显著，更加接近于人类思维中的理想风格化图像。

4.4 消融实验

如图4-6所示，与加入Identity损失的模型的风格化图像相比，未加入Identity损失的模型的风格化图像的色彩深度明显较弱。此外，在素描风格迁移图像中，未加入Identity损失的模型色调相较于风格图像更加偏冷，且色彩明度较差。

观察未引入TV损失的模型的风格化图像可以看出其有明显边缘伪影，同时通过细节对比，我们可以看出虽然引入TV损失获得了更高的的图像平滑度，但其图像分辨率并没有因此明显下降。

故根据消融实验，本文认为加入TV损失和Identity损失对生成风格迁移图像具有明显改善效果。

Full model: **no Identity loss:** **no TV loss:**

图4-6 消融实验

4.5 系统使用说明书及效果分析

(1) 点击“选择风格图片”可以从本地文件目录选取特定的内容图像：

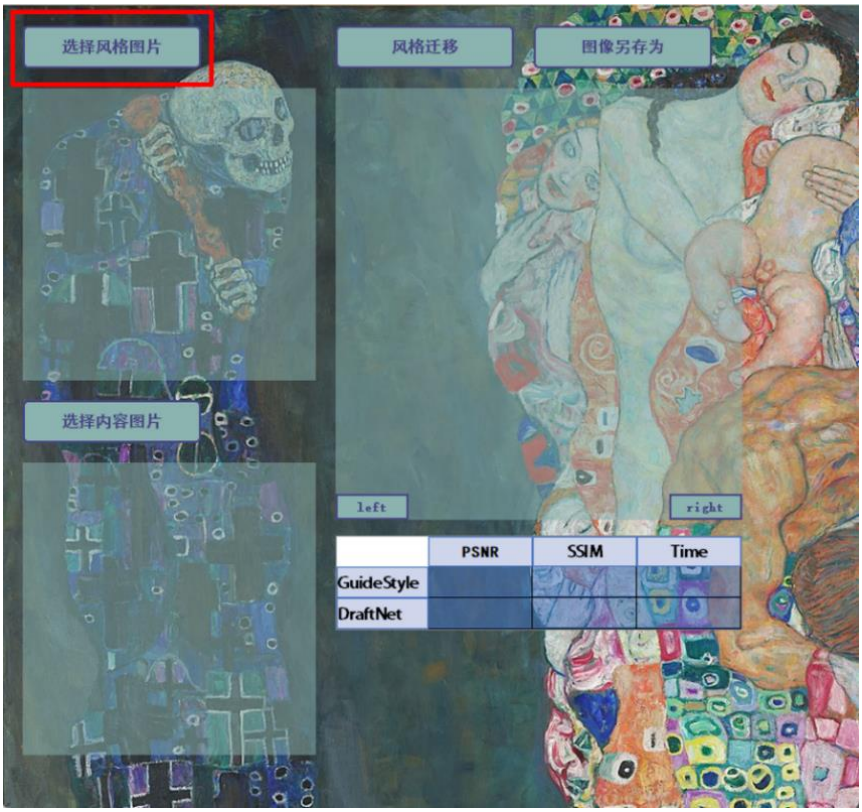


图 4-7 选择风格图像

随后风格图片会显示在界面窗口内。

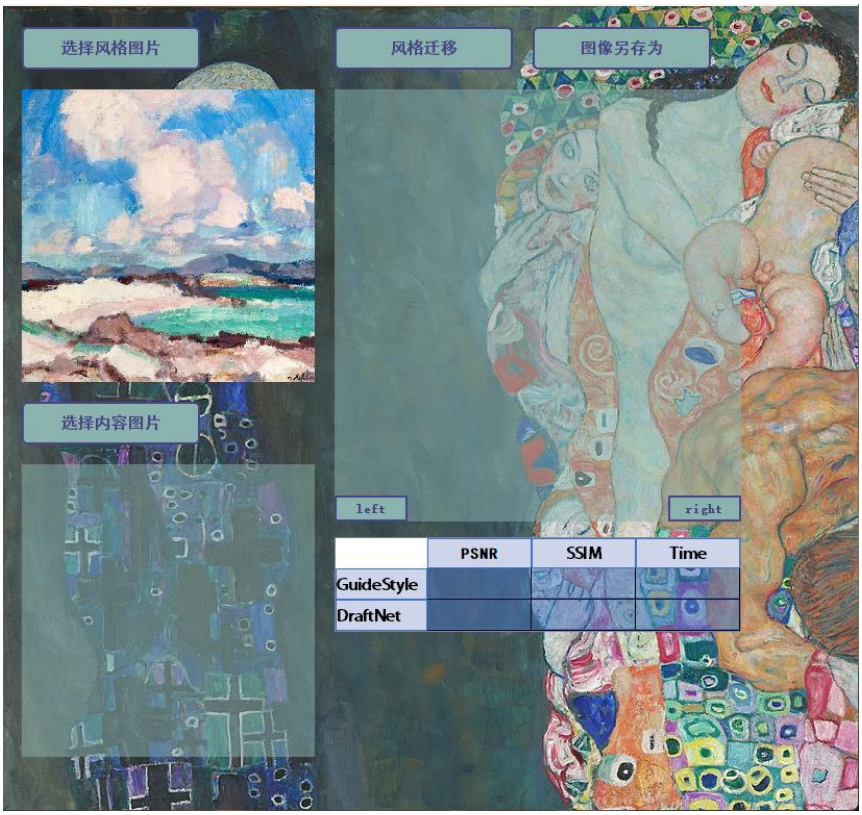


图 4-8 显示风格图像

(2) 点击“选择内容图片”可以从本地文件目录选取特定的内容图像：

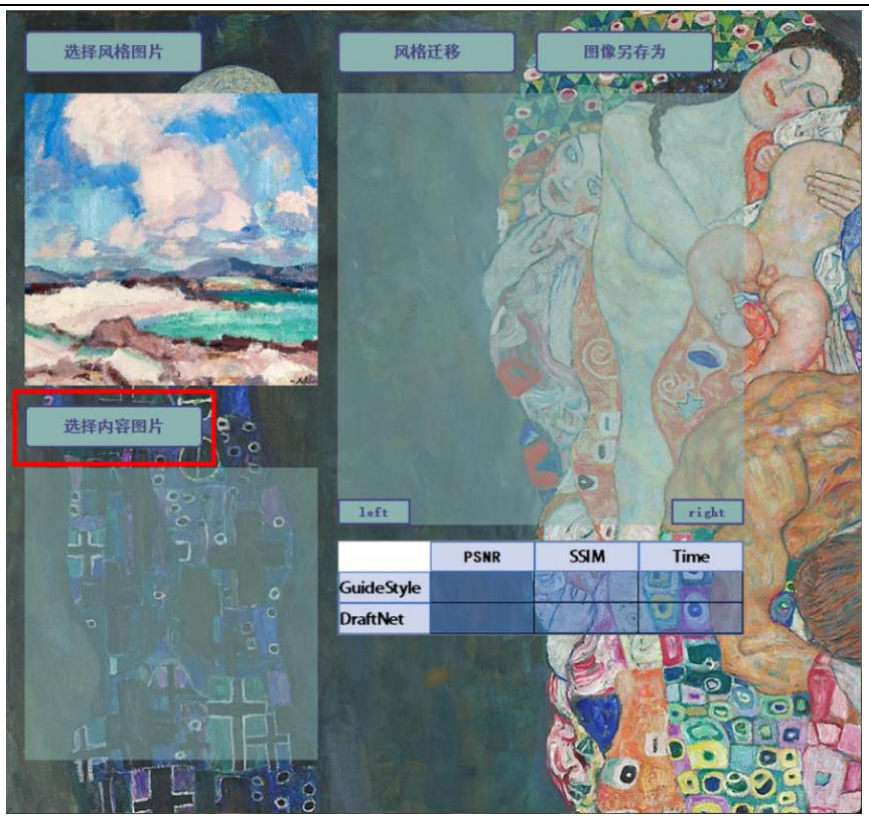


图 4-9 选择内容图像

随后内容图片会显示在界面窗口内。

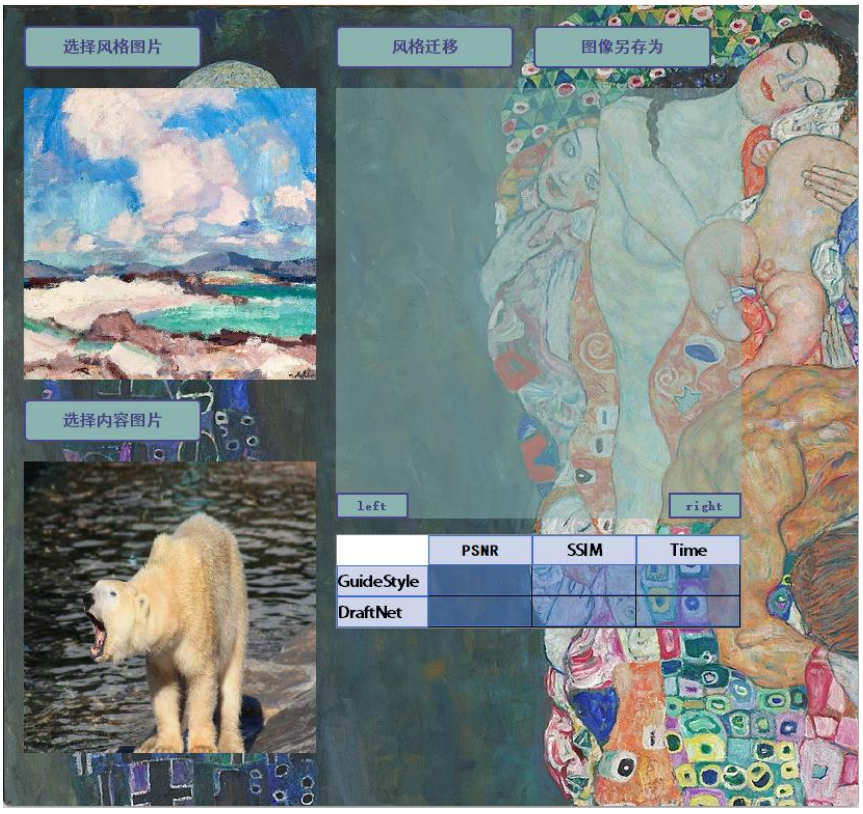


图 4-10 显示内容图像

(3) 点击“风格迁移”按钮进行图像风格迁移操作：

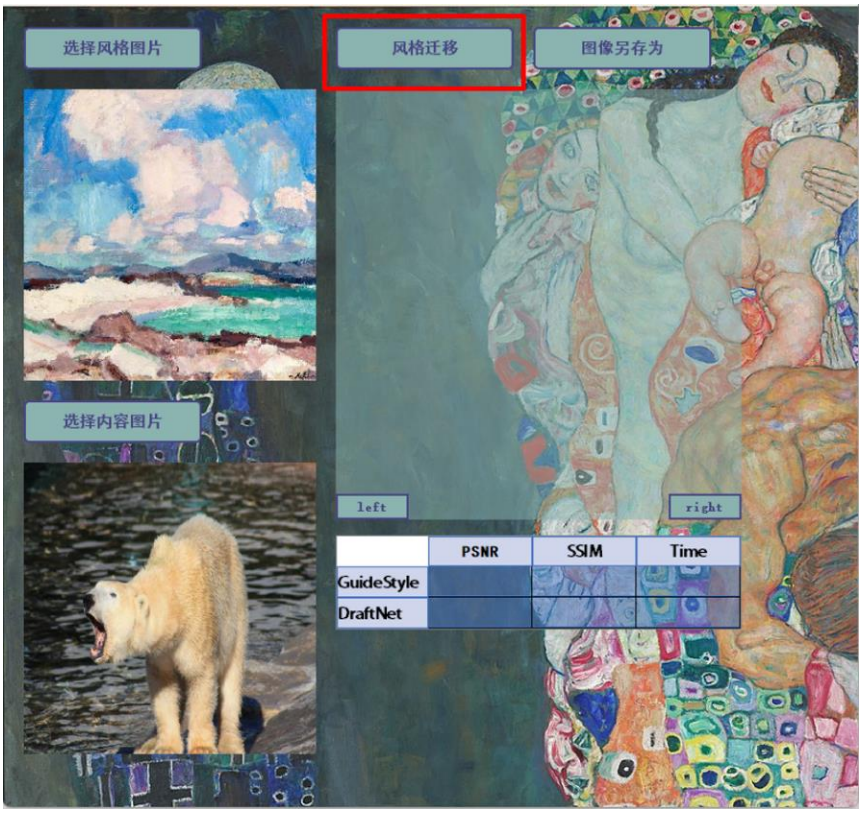


图 4-11 进行风格迁移

迁移完成后在对应窗口显示风格迁移后的图片，并在下方表格中显示该风格化图片与风格图像的SSIM与VIF评估参数，方便客户对迁移效果有直观量化的观察。同时点击“图片另存为”按钮可以将风格化图像保存在用户指定的任意位置。

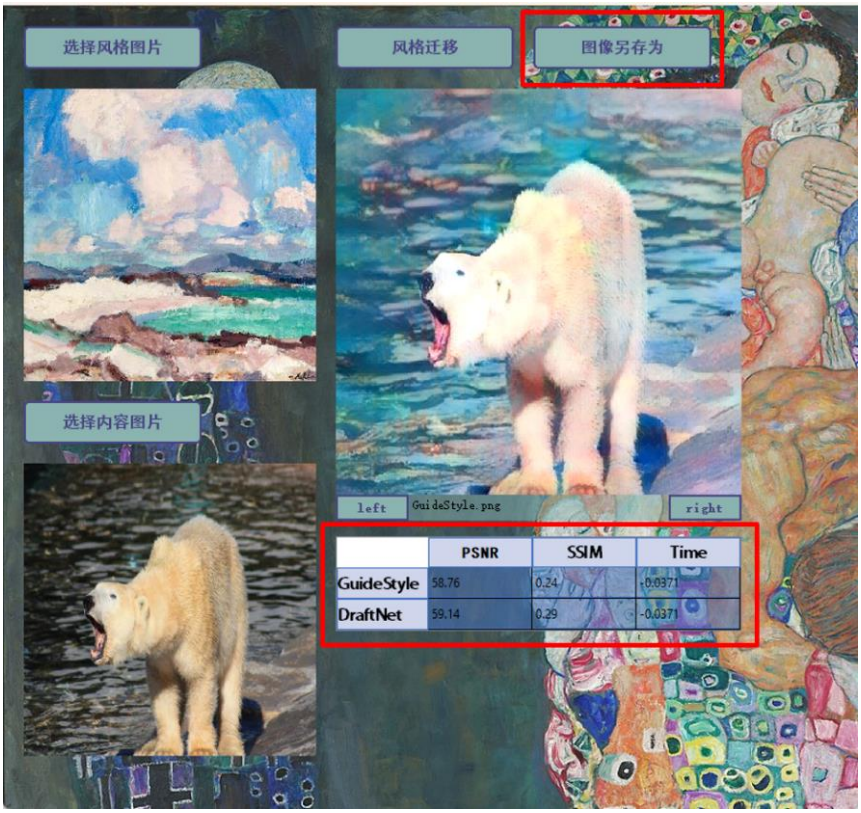


图 4-12 显示迁移效果

此外，如图4-13所示，本文的风格迁移系统点击风格化图像下边框的“right”按钮可

以切换不同的风格化图像，如草稿网络生成的256分辨率的草稿图像。

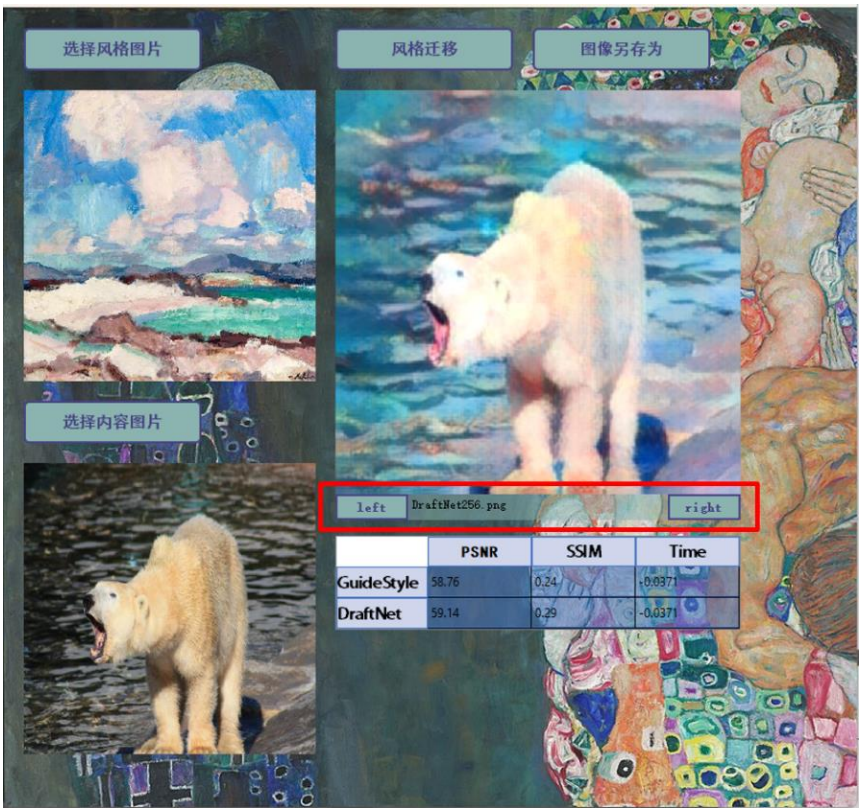


图 4-13 选择窗口显示图像

本文基于自适应实例归一化和引导滤波算法，由草稿网络和修订网络组成风格迁移模型GuideStyle，训练集由COCO和WikiArts构成，实现将内容图像、风格图像融合为风格化图像的风格迁移过程。

如图4-7至图4-13所示，操作的方便性在按键设置的顺序、大小与样式上有所体现，操作流程与按键排布相符，图片的显示与评估参数计算也便于使用者及时接收识别结果，并且该系统可移植性强，进一步增加了可用性。

总体而言，本系统风格迁移的风格化效果有较好的表现，在符合要求的GPU环境下风格迁移时间较短，以本系统设计目的为主要依据，系统操作方便性较优。综合考虑，本系统的目的性和可用性得到了满足。

4.6 本章小结

本章首先介绍了系统性能的两个评价度量指标：峰值信噪比、结构相似性。基于此，在系统模型的风格迁移效果分析和系统整体的使用效果分析两方面开展分析，进行了与AdaIN原模型的对比分析与系统整体的使用效果分析。同时本章提供了该系统的整体使用说明，方便用户了解该系统的操作方法。最后，综合以上不同方面的依据得出结论，该系统能够较好地实现系统目的和满足用户需求。

第 5 章 结论与展望

5.1 结论

结果表明，本文设计的基于自适应实例归一化(AdaIN)和引导滤波的风格迁移模型，在SSIM、PSNR等性能指标上有较好的表现，在保留内容信息的同时，整体风格模式被正确迁移。GuideStyle模型不仅大大减轻了传统AdaIN模型输出风格化图片具有的伪影和扭曲的缺陷，同时在风格化图片中具有更丰富的图像细节和图像深度。通过引入TV损失，本文的生成图像具有更高的平滑度和更少的边缘噪声，同时由于本系统的结构设计，在具有此项优势的同时，风格化图像并不会损失一定的细节信息。

此外，在传统模型中，由于仅具有单一网络，其网络模型适用的图像分辨率是固定的，且在训练高分辨率网络时，由于网络的高复杂度使得在训练时具有大量的显存开销并要付出高昂的时间成本。本文通过引入基于引导金字塔的修正网络来解决这一痛点，通过逐级生成修正残差图像来补全高分辨率图像细节，最后输出的风格化图像是多层网络输出聚合的结果。本文不仅实现了输出图像分辨率的高灵活性和高扩展性，且由于单层网络的结构被设计的十分简单，在训练网络过程中本文大大减少了训练所需的显存容量和时间开销。因此，GuideStyle模型具有迁移速度快，迁移质量高等特点，适用于快速风格迁移系统的设计目的。

最后，本文基于GuideStyle方法开发了对应的图形用户界面，使得图像风格迁移操作更加便捷。该图形用户界面具有操作便捷、结果直观等优点，如在系统风格迁移界面，生成风格化图像的同时会计算对应的风格化图像与内容图像和风格图像的PSNR、SSIM指标，以便用户可以直观的量化图像风格迁移质量。

综上所诉，本文设计的基于自适应实例归一化和引导滤波的快速风格迁移模型GuideStyle能够很好的满足设计目的和要求。

5.2 不足之处及未来展望

由于本次实验的硬件限制，对各超参数的选择仍具有很大的优化空间。且该网络在任意风格迁移效果上与单一风格迁移效果仍有一定差异，因此具有较大的改进空间。在更为复杂的、缺少明确纹理信息的风格图像中，如棋盘格风格图像，模型的风格迁移能力仍然有待改进。

此外，由于本文的草稿网络依旧使用DCNN结构的编解码器，以及使用预训练的VGG-19网络，模型的参数数量较为庞大，使其对于高分辨率图像的快速风格迁移扩展性较差，仍然拥有很大的优化空间。同时，由于通过GuideStyle方法生成超分辨率图像是通过以上采样金子塔的方式堆叠修正网络来完成的，因此对于超分辨率模型来说，其势必面临着模型繁琐、训练复杂、超参数过多等缺点。并且，尽管修正网络和草稿网络均为全卷积网络所以理论上支持任意大小的图像输入，但根据模型训练方式若要获得理想的风格化图像，每层修正网络势必对应着固定的分辨率尺寸，因此GuideStyle模型仅使用单一结构的情况下无法支持任意分辨率图像的高质量图像风格迁移。

随着时代的发展，风格迁移的应用也越发广泛。同时在人工智能领域，技术的发展

日新月异，因此对于风格迁移的更深入探索是十分必要的。此外，目前的工作还仅限于单一风格图像的风格迁移，其语义水平还可以进一步提高。 虽然每个艺术家都有不同的作品风格，但在画作中存在着比较抽象的作者风格特征。如何从一个作者的多幅画作中提取更高层次的特征，并对所得图像进行风格迁移是一个值得探索的方向。

参考文献

[1] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 1501-1510.

[2] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2414-2423.

[3] Li C, Wand M. Combining markov random fields and convolutional neural networks for image synthesis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2479-2486.

[4] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016: 694-711.

[5] Li Y, Wang N, Liu J, et al. Demystifying neural style transfer[J]. arXiv preprint arXiv:1701.01036, 2017.

[6] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

[7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[8] Loffe S, Normalization C S B. Accelerating deep network training by reducing internal covariate shift[J]. arXiv, 2014.

[9] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[10] Ioffe S. Batch renormalization: Towards reducing minibatch dependence in batch-normalized models[J]. Advances in neural information processing systems, 2017, 30.

[11] Li Y, Wang N, Shi J, et al. Revisiting batch normalization for practical domain adaptation[J]. arXiv preprint arXiv:1603.04779, 2016.

[12] Liao Q, Kawaguchi K, Poggio T. Streaming normalization: Towards simpler and more biologically-plausible normalizations for online and recurrent learning[J]. arXiv preprint arXiv:1610.06160, 2016.

[13] Ba J L, Kiros J R, Hinton G E. Layer normalization[J]. arXiv preprint arXiv:1607.06450, 2016.

[14] Salimans T, Kingma D P. Weight normalization: A simple reparameterization to accelerate training of deep neural networks[J]. Advances in neural information processing systems, 2016, 29.

[15] Cooijmans T, Ballas N, Laurent C, et al. Recurrent batch normalization[J]. arXiv preprint arXiv:1603.09025, 2016.

[16] Laurent C, Pereyra G, Brakel P, et al. Batch normalized recurrent neural networks[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016: 2657-2661.

[17]Ren M, Liao R, Urtasun R, et al. Normalizing the normalizers: Comparing and extending network normalization schemes[J]. arXiv preprint arXiv:1611.04520, 2016.

[18]Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: Feed-forward synthesis of textures and stylized images[J]. arXiv preprint arXiv:1603.03417, 2016.

[19]Ulyanov D, Vedaldi A, Lempitsky V. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 6924-6932.

[20]Dumoulin V, Shlens J, Kudlur M. A learned representation for artistic style[J]. arXiv preprint arXiv:1610.07629, 2016.

[21]Dosovitskiy A, Brox T. Inverting visual representations with convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4829-4837.

[22]Chen T Q, Schmidt M. Fast patch-based style transfer of arbitrary style[J]. arXiv preprint arXiv:1612.04337, 2016.

[23]He K, Sun J, Tang X. Guided image filtering[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(6): 1397-1409.

[24]Shaham T R, Dekel T, Michaeli T. Singan: Learning a generative model from a single natural image[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 4570-4580.

[25]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[26]Wang J, Chen K, Xu R, et al. Carafe: Content-aware reassembly of features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3007-3016.

[27]Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[28]Wang H, Li Y, Wang Y, et al. Collaborative distillation for ultra-resolution universal style transfer[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1860-1869.

[29]Sanakoyeu A, Kotovenko D, Lang S, et al. A style-aware content loss for real-time hd style transfer[C]//proceedings of the European conference on computer vision (ECCV). 2018: 698-714.

[30]Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 4401-4410.

[31]Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.

[32]Lin T, Ma Z, Li F, et al. Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 5141-5150.

[33]Wang Z, Zhao L, Zuo Z, et al. MicroAST: Towards Super-Fast Ultra-Resolution Arbitrary Style Transfer[J]. arXiv preprint arXiv:2211.15313, 2022.

致 谢

感谢李辉导师和张泽阳学长对本文的悉心指导！他们在论文撰写以及促成GuideStyle模型提出等多方面提供了诸多启发与帮助。同时，感谢李辉导师与程春阳博士为本项目所需的实验资源提供的诸多硬件资助。

本科时光转瞬即逝，在生活方面，感谢我的父母和陈浩涵同学长久以来的陪伴与鼓励，让我在遇到困难之时得以坚持向前。同时，感谢我的六位舍友，罗骞、马伯源、宋易阳、邓旺、伊克山、吴晗晖四年寝室生活中对我的包容与支持。