第 58 卷 第 12 期/2021 年 6 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

目标边缘清晰化的图像风格迁移

沈瑜，杨倩*，苑玉彬，王霖

兰州交通大学电子与信息工程学院，甘肃 兰州 730070

摘要 针对图像风格迁移时出现前后景边界模糊造成风格化图像主要目标模糊的问题，提出了目标边缘清晰化的图像风格迁移算法。通过将用于提取内容图像轮廓的深度抠图神经网络与风格迁移网络合并，形成透明遮罩约束风格迁移过程，凸显风格化图像中主要目标的轮廓；通过对迁移网络中最大池化层进行替换，保留图像的背景信息，细化风格化图像的整体结构；通过替换迁移网络中较大卷积核，减少网络模型参数，减少风格迁移计算量；通过对常规卷积层的归一化，实现相似风格迁移之间的参数共享，提升风格迁移速度。用 VGG-19 神经网络作为特征提取器对输入的内容图像和风格图像提取特征图，把输入图像到输出图像的变换约束在色彩空间局部仿射中，在输入图像 RGB 通道上合并目标遮罩，使得风格化图像的主要目标在遮罩约束中实现纹理合成。实验表明，与传统方法比较，该方法产生的迁移结果前后景边缘明显，内容结构保留较好，解决了风格化图像主要目标边缘模糊的问题。

关键词 图像处理；风格迁移；神经网络；抠图算法；深度学习；结构约束

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.1210021

Image Style Transfer with Clear Target Edges

Shen Yu, Yang Qian*, Yuan Yubin , Wang Lin

College of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, *China*

Abstract In the process of image style transfer, the main target of the stylized image is blurred due to the blurring of the foreground and background boundaries in image reconstruction. Image style transfer algorithm with clear target edges is proposed. The deep matting neural network used to extract outline of the content image is merged with the style transfer network to form a transparent mask to constrain the style transfer process, highlighting the outline of the main target of the stylized image. By replacing the max-pooling layer in the transfer network, more image background information is retained, and the overall structure of the stylized image is refined. The network model parameters are reduced by replacing the larger convolution kernel in the transfer network, and the calculation process of style transfer is simplified. Then, the normalization of the conventional-convolutional layer realizes parameter sharing between similar style transfers, improving the speed of style transfer. The VGG-19 neural network as a feature extractor extracts feature maps from the input content and style images, constrains the transformation from the input image to the output image in local affine of color space, and merges the target mask on the RGB channel of the input image so that the main goal of stylized images is to achieve texture synthesis in mask constraints. Experimental results show that compared with the traditional method, this method has obvious boundaries in the foreground and background of the stylized image, and the content structure is preserved well.

Key words image processing; style transfer; neural network; matting algorithm; deep learning; structural constraints

OCIS codes 100.4996; 100.4994; 100.4997; 100.3010

收稿日期：2020-08-27；**修回日期**：2020-09-22；**录用日期**：2020-10-29

基金项目：国家自然科学基金(61861025，61663021，61761027，61669010)、长江学者和创新团队发展计划(IRT_16R36)、兰州市人才创新创业项目(2018-RC-117)、光电技术与智能控制教育部重点实验室(兰州交通大学)开放课题(KFKT2018-9)

* **E-mail**：13662175532@163.com

1 引 言

图像风格渲染是计算机视觉领域中重要的研究方向，在电影产业、动画制作、游戏渲染等领域有着大量的应用[1]。图像风格艺术化是指一幅图像的语义内容用另一幅图像的风格去表示。传统的风格迁移大多采用手动建模，需要专业经验和复杂的数学公式[2]。深度神经网络凭借其强大的图像表示能力，推动了神经样式转移方法的发展。因此，利用高效准确的图像风格化算法提取风格及内容图像，可使风格化图像的主要目标清晰可辨且风格化效果显著[3-4]。

Gatys 等[5-9]开创性地提出了基于卷积神经网络的图像风格迁移，将图像的内容特征和风格特征进行分离，并独立处理高层特征表示来实现图像风格的迁移，获得了艺术效果。Ulyanov 等[10-11]用实例标准化模块来替代批量标准化，在纹理集合中采用无偏差采样来提高风格化速度。Huang 等[12]提出了自适应实例标准化(AdaIN)层与迭代优化相结合的前馈方法，实现了任意样式的实时转换。Li 等[13]采用马尔可夫生成对抗网络直接将内容图像转换成艺术画作，提升了图像合成的速度。Mechrez 等[14]通过比较具有相似语义的区域进行特征相似匹配，实现了风格迁移。Johnson 等[15]从预训练网络中提取高级特征来定义和优化感知损失函数，提高了风格转换的速度。Liu 等[16]通过训练具有额外深度损失函数的前馈变换网络，并与优化的图像生成方法相结合，保留了图像的语义内容和布局。Chen 等[17]提出的卡通化生成对抗网络，能够从真实照片中生成高质量的卡通图像。

在图像迁移过程中，图像布局被破坏且前景、背景以及其他物体之间的边界变得模糊，视觉效果不够理想。本文针对风格化图像目标轮廓模糊的问题，提出了目标边缘清晰化的图像风格迁移算法。首先，搭建编解码深度神经网络，编码器将输入通过后续的卷积层和最大池化层转换成下采样的特征映射，解码器使用反池化层和反卷积层对特征映射进行上采样。其次，在迁移网络中，在相同的感受野条件下，用小卷积核代替大卷积核，增加网络深度和非线性，减少参数；用平均池化层替换最大池化层，减小因邻域大小受限造成的估计值方差增大量，保留更多的背景信息。在迁移网络的常规卷积层后添加归一化层，归一化层通过学习仿射参数匹配内容图像和风格图像的统计信息，实现风格类似的图像之间的参数共享，减少迁移模型的计算量。最后，在色彩空间局部仿射变换中合并深度抠图网络和风格迁移网络，用 VGG-19 模型[18]提取特征图进行风格重建和内容重建，用遮罩约束风格迁移，实现图像风格化。

2 风格迁移

图像神经风格迁移的主要目标是将自然图像用参考图像的色彩和纹理进行表示，可主要通过以下两个方面实现：1)通过对不同的色彩通道进行不同的处理，实现用户对颜色和纹理的控制；2)为了使迁移后的图像符合自然图像的语义内容，通常先对内容图像进行语义分割，再对其进行风格转换，或者先对不同区域进行标记，再进行不同纹理的迁移。

由于风格迁移是对风格图像和内容图像进行特征提取，然后对风格化图像进行重建，故重建过程会出现图像前后景及周围物体边缘模糊。因此，本文针对风格化图像目标轮廓模糊的问题，提出了目标边缘清晰化的图像风格迁移算法。1)本文算法利用 VGG-19 网络进行特征提取，采用网络 Conv4_4 和 Conv5_4 的高维内容特征图进行内容表示，融合卷积 Conv1_2，Conv2_2，Conv3_4，Conv4_4，Conv5_4 的风格特征图进行风格表示；2)将风格特征图像、内容特征图像和抠图网络产生的遮罩叠加在高斯白噪声图像上，并将其约束在色彩空间局部仿射上，进行风格化图像重建；3)计算风格化图像与提取内容特征图之间的欧氏距离和均方误差，并按权重叠加得到整体风格化误差；4)用反向传播进行迭代更新，使误差函数最小得到理想的风格化图像，解决了前后景边界模糊的问题。网络模型如图 1 所示。

2.1 VGG-19 网络

VGG-19 感知网络[19]是使用 ImageNet 进行对象分类的预训练模型，其结构如图 2 所示。VGG-19 网络由多层非线性函数组成，用于计算风格图像和内容图像的特征表示，然后计算目标图像对应特征与合成图像之间的欧氏距离，并使用训练好的深度网络作为一个损失函数的计算器计算损失。

风格迁移保留内容图像的语义内容和空间布局。内容表示选取图像的高维特征，提取出的高维特征之间的欧氏距离越小，则表明生成图像与原始内容图像的内容越相似。风格表述在不同层的特征表达有不同的视觉效果，计算不同卷积层间的格莱姆矩阵建立风格损失函数，可使风格表达更加丰富，能够达到风格全局表述的目的。

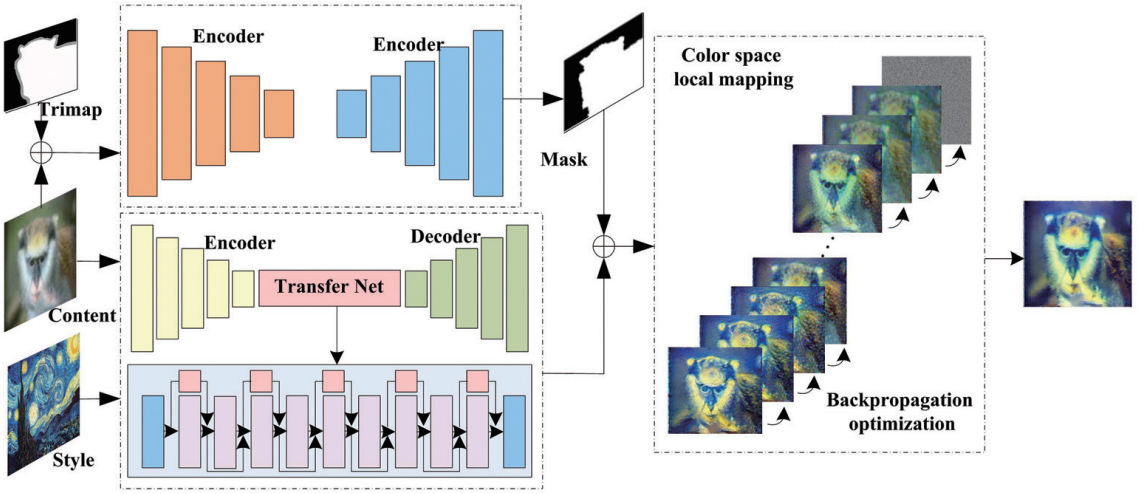

图 1　网络框架图

Fig. 1　Network framework diagram

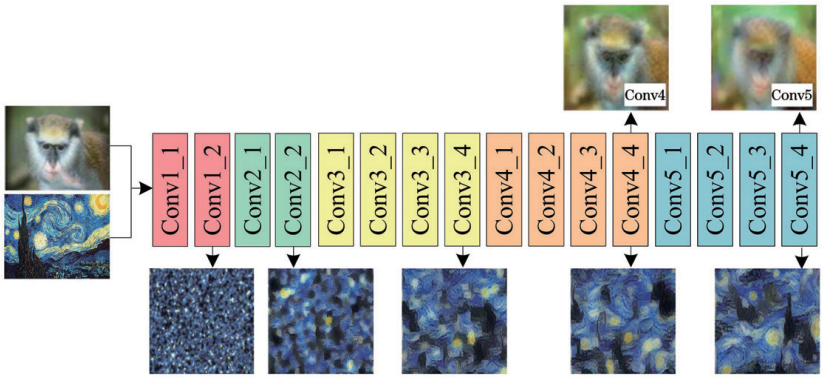

图 2　VGG-19 网络特征提取

Fig. 2　VGG-19 network feature extraction

2.2　深度抠图神经网络

2.2.1　抠图理论

抠图是将图像分成前景和背景两块，目的是取出前景。抠图过程是将图像区域属于前景或背景的概率值返回，但在前景与背景交互区域会产生渐变的效果[20-21]，抠图表示如图 3 所示。

图 3　图像的前后景表示

Fig. 3　Representation of foreground and background of the image

抠图原理表达式为

$$T_i=\alpha r+(1-\alpha)s, \tag{1}$$

其中，T_i 是图像可观察的像素；α 是遮罩，r 是前景像素，s 是背景像素。原始图像可以被看作是由前景和背景按照一定权重叠加而成。对于完全确定是前景的像素，$\alpha=1$；对于完全确定是背景的像素，$\alpha=0$；对于不确定是前景还是背景的像素，α 为介于 0 到 1 之间的浮点数。

遮罩可以在局部表示为图像颜色通道的线性组合[22]，如图 4 给出的两个代表性示例所示。如果图像的局部窗口是一个前景和背景颜色相当统一的窗口[图 4(1)]，则 α 遮罩与图像具有很强的归一化相关性，可通过将其中一个颜色通道乘以比例因子并添加一个常数来生成。如果图像局部窗口的 α 遮罩在整个窗口中都是常量[图 4(2)]，都可以通过将图像通道乘以 0 并添加一个常数来获得。由于大多数图像窗口上的典型遮罩是常数(0 或 1)，因此这些窗口中的遮罩可以表示为图像的线性函数。

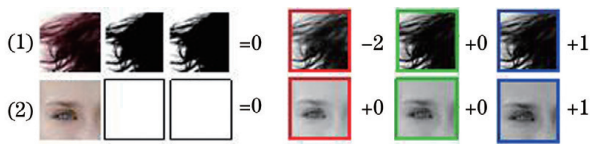

图 4　遮罩颜色通道的线性组合

Fig. 4　Linear combination of mask color channels

2.2.2　深度抠图模型

抠图网络由编码网络和解码网络组成。本文抠图网络的解码子网分为 5 个阶段，网络结构如图 5 所示。对每个阶段的卷积层设置不同的步长，如表 1 所示。

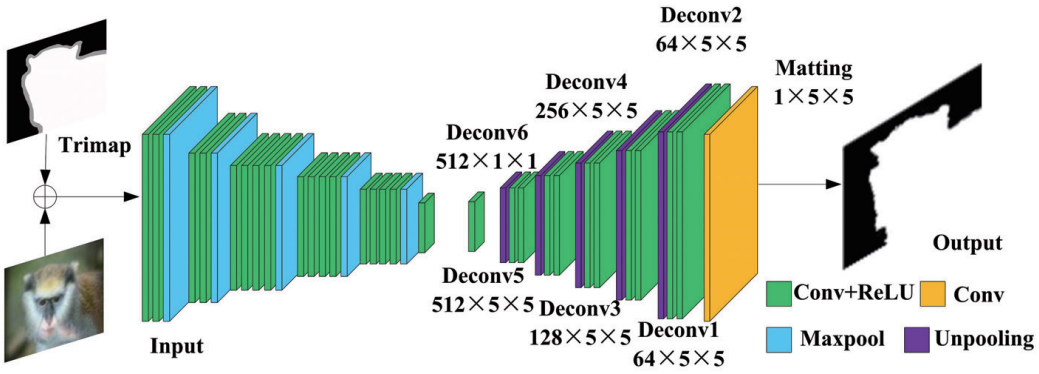

图 5　深度抠图网络结构

Fig. 5　Deep matting network structure

表 1　步长和感受野参数设置

Table 1　Stride length and receptive field parameter settings

Layer	Conv1	Conv2	Conv3	Conv4	Conv5
Stride	1	2	4	8	16
Receptive field	5	14	40	92	196

本文抠图网络具有如下特点：1）编码器使用 VGG-19 模型的前 17 层，将第 17 层的全卷积层转换为卷积层；2）解码网络采用小卷积核堆叠代替大卷积核，有 11 个卷积层、5 个反池化层，网络结构更小；3）编码网络的输入内容图像和相应的 Trimap，沿着通道尺寸连接在一起，形成 4 通道输入；4）网络的最后一层是具有 Sigmoid 函数的 1×1 的卷积层，将特征图的通道数减少为 1，其中的元素阈值设为[0，1]，以生成透明的遮罩。

多层次多尺度的反卷积解码能产生更加精确的抠图遮罩（图 6），与传统的抠图算法相比，本文的抠图算法能更好地区分前景和后景[图 6(d1)]，对于前景和后景颜色相近的图像，也能更好地抠取出前景[图 6(d2)]。

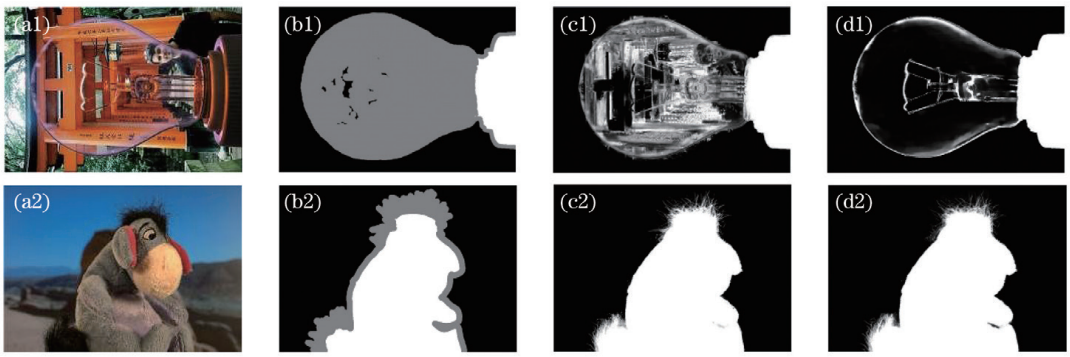

图 6　抠图效果对比图。(a1) (a2)原始图像；(b1) (b2) Trimap；(c1) (c2)传统抠图；(d1) (d2)本文深度抠图

Fig. 6　Comparison of matting effects. (a1) (a2) Original images; (b1) (b2) Trimap; (c1) (c2) traditional matting; (d1) (d2) depth matting of this paper

2.3　风格迁移网络

本文搭建的风格迁移网络分为编码器、转换网络和解码器三个部分。编解码器是用 VGG-19 网络来实现对输入图像的特征提取。转换网络主要由 Fire Module 构成，为了在特征通道中产生较高的平均激活度，在 Conv1 后增加归一化层和 ReLU 层，实现了纹理合成和色彩迁移。网络主体如图 7 所示，网络参数如表 2 所示。在转换网络中，本文在 Conv1 用 3 个 3×3 的卷积核代替一个 7×7 的卷积核，并将 3 个 3×3 卷积核堆叠，增加了多层非线性组合，提高了网络学习复杂内容的能力，并且使判决函数更具判决性，起到了隐式正则化的作用。本文将最大池化用平均池化代替，平均池化能减小邻域大小受限造成的估计值方差增大量，更多地保留图像的背景信息。

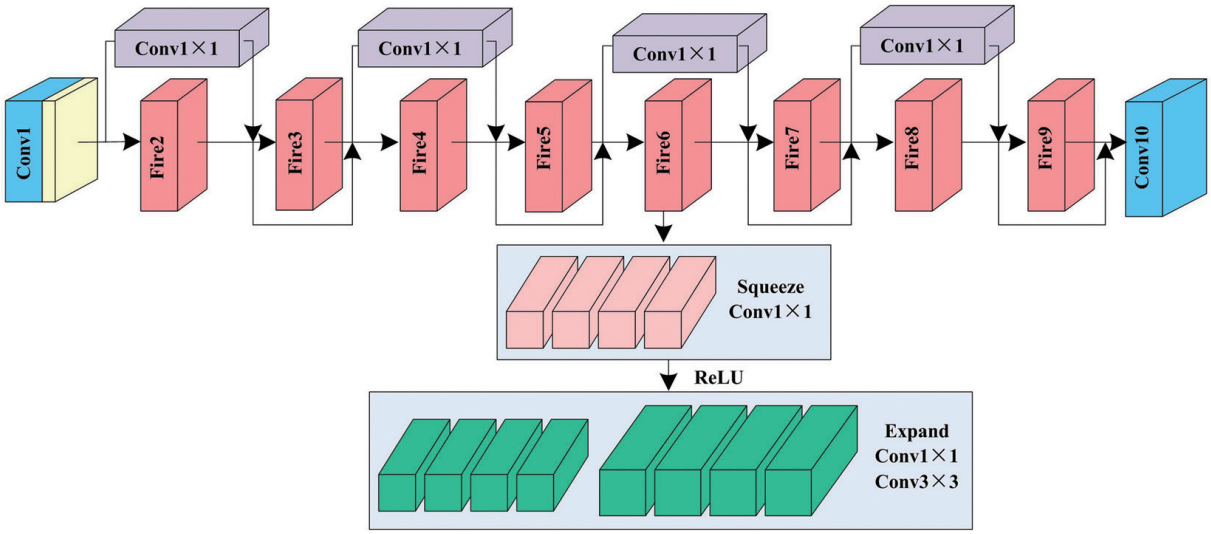

图 7 转换网络结构

Fig. 7 Transform network structure

表 2 转换网络参数

Table 2 Transfer network parameters

Layer	Output size	Size, Stride	Depth	Parameter
Input	224×224×3			
Conv1	111×111×9	3×3×3, 2(×96)	1	10728
Maxpool1	55×55×96	3×3, 2	0	
Fire2, Fire3	55×55×128		2	12004
Fire4	55×55×256		2	20646
Avgpool4	27×27×256	3×3, 2	0	
Fire5	27×27×256		2	24742
Fire6, Fire7	27×27×384		2	90936
Fire8	27×27×512		2	77581
Maxpool8	13×12×512	3×3, 2	0	
Fire9	13×13×512		2	77581
Conv10	13×13×1000	1×1, 1(×1000)	1	103400
Avgpool10	1×1×1000	13×13, 1	0	

网络训练是学习各种参量的过程，其参数量由网络深度、卷积核尺寸以及通道数来决定。在单个卷积层中，使用 3 个 3×3 卷积核的参数量为 $3\times3\times3\times C\times C=27\times C^2$，使用 7×7 卷积核的参数量为 $7\times7\times C\times C=49\times C^2$（$C$ 为通道数），前者只有后者的一半左右，前者经 3 个非线性操作，后者经 1 个非线性操作，故后者具有强的特征学习能力，且网络学习参数较少，网络风格迁移的速度大幅提升。

本文将深度抠图网络与风格迁移网络相结合，凸显了主要目标的高层次特征，忽略了弱细节，使得生成图像边缘轮廓模糊，语义扭曲严重[图 8(c)]。大卷积核通常会获得更大的感受野，生成尺寸较大的特征图，经过池化压缩特征图去除冗余信息的同时会损失背景信息。本文用 3×3 卷积核代替 7×7 卷积核，能够保证相同感受野的条件下，产生更加准确的特征响应，并且增加网络深度，使得网络中的函数能够更好地逼近输入图像的特征，能更好地捕捉自然图像的统计属性，让目标轮廓越来越清晰，细节纹理更精细，视觉效果更好[图 8(d)]。

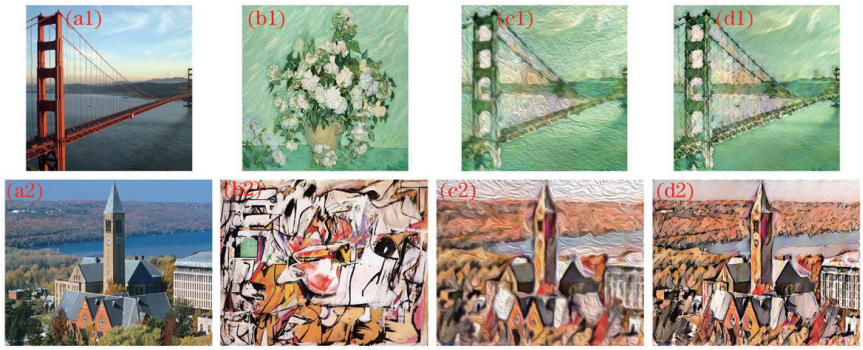

图 8　不同卷积核风格迁移纹理对比。(a1)(a2)内容图像;(b1)(b2)风格图像;(c1)(c2) 7×7 卷积核迁移纹理;(d1)(d2) 3×3 卷积核迁移纹理

Fig. 8　Comparison of style transfer texture of different convolution kernels. (a1)(a2) Content images; (b1)(b2) style images; (c1)(c2) 7×7 convolution kernel transfer texture; (d1)(d2) 3×3 convolution kernel transfer texture

3　神经风格迁移损失

假设网络的第 l 层有 $N^{(l)}$ 个不同的滤波器，每个特征响应的大小为 $H^{(l)}\times W^{(l)}$，其中 $H^{(l)}$ 和 $W^{(l)}$ 分别为第 l 层特征映射的高度和宽度，该层的响应可表示为

$$\boldsymbol{M}^{(l)}\in\mathbb{R}^{(H^{(l)}\times W^{(l)})\times N^{(l)}},\tag{2}$$

其中，每一个值 $M^{(l)}_{(i,j),k}$ 都是在 l 层 (i,j) 位置第 k 个滤波器的激活函数值。

内容图像 $\boldsymbol{x}$ 与风格图像 $\boldsymbol{y}$ 和风格化图像 $\hat{\boldsymbol{y}}$ 在第 l 层的损失函数分别为

$$L^{(l)}_{\text{content}}(\boldsymbol{x},\hat{\boldsymbol{y}})=\frac{1}{H^{(l)}W^{(l)}N^{(l)}}\left\|\boldsymbol{M}^{(l)}(\boldsymbol{x})-\boldsymbol{M}^{(l)}(\hat{\boldsymbol{y}})\right\|_2^2,\tag{3}$$

$$L^{(l)}_{\text{style}}(\boldsymbol{y},\hat{\boldsymbol{y}})=\left\|\boldsymbol{G}^{(l)}(\boldsymbol{y})-\boldsymbol{G}^{(l)}(\hat{\boldsymbol{y}})\right\|_{\text{F}}^2,\tag{4}$$

式中：$\boldsymbol{G}^{(l)}$ 是格莱姆矩阵，是一个 $N^{(l)}\times N^{(l)}$ 的对称矩阵；$\boldsymbol{F}^{(l)}_{i,j}$ 是第 l 层的第 i 个和第 j 个矢量化特征图的归一化内积，即

$$\boldsymbol{G}^{(l)}_{i,j}(\boldsymbol{y})=\frac{1}{H^{(l)}W^{(l)}N^{(l)}}\boldsymbol{F}^{(l)}_{(h,w),i}(\boldsymbol{x})\boldsymbol{F}^{(l)}_{(h,w),j}(\boldsymbol{x})。\tag{5}$$

特征提取是将输入图像先编码再解码的过程，条件归一化层[23]通过编码器将图像映射在特征空间中，然后将其统计信息进行匹配，匹配公式为

$$\boldsymbol{A}^{(c)}=\frac{\sigma[\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{x})]}{\sigma[\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{y})]}\{\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{x})-\mu[\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{x})]\}+\mu[\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{y})],\tag{6}$$

式中：$\boldsymbol{A}^{(c)}$ 为第 c 个通道的统计匹配特征图；$\mu(\boldsymbol{M}^{(c)}_{\text{conv}})$ 和 $\sigma(\boldsymbol{M}^{(c)}_{\text{conv}})$ 是在特征图 $\boldsymbol{M}^{(c)}_{\text{conv}}$ 的所有位置上计算的平均值和标准偏差，表达式为

$$\mu(\boldsymbol{M}^{(c)}_{\text{conv}})=\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\boldsymbol{M}^{(h\times w\times c)}_{\text{conv}},\tag{7}$$

$$\sigma(\boldsymbol{M}^{(c)}_{\text{conv}})=\left\{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}[\boldsymbol{M}^{(h\times w\times c)}_{\text{conv}}-\mu(\boldsymbol{M}^{(c)}_{\text{conv}})]^2\right\}^{\frac{1}{2}}。\tag{8}$$

解码器将生成的特征映射转换到图像空间，经归一化操作后产生的内容损失和风格损失可分别表示为

$$L_{\text{N,C}}=\frac{1}{HWC}\left\|\hat{\boldsymbol{y}}-\boldsymbol{M}^{(l)}_{\text{conv}}(\boldsymbol{x})\right\|_2^2,\tag{9}$$

$$L_{\text{N,S}}=\left\|\mu(\hat{\boldsymbol{y}})-\mu[\boldsymbol{M}^{(l)}_{\text{conv}}(\boldsymbol{y})]\right\|_2^2+\left\|\sigma(\hat{\boldsymbol{y}})-\sigma[\boldsymbol{M}^{(l)}_{\text{conv}}(\boldsymbol{y})]\right\|_2^2。\tag{10}$$

图像表示的内容差异和风格差异通过搭建的转换网络和归一化操作产生，网络学习通过随机梯度下降法使加权和损失最小化来优化损失函数，优化的损失函数的表达式为

$$L_{\text{loss}}=\alpha\sum_{l=1}^{L}(L_{\text{content}}+L_{\text{N,C}})+\beta\sum_{l=1}^{L}(L_{\text{style}}+L_{\text{N,S}})+\gamma L_{\text{R}},\tag{11}$$

式中：L_{R} 为正则化项[24]，以增加生成图像的平滑性；L 为产生损失函数的层数；α，β，γ 分别为内容损失函数、风格损失函数和正则化项的权重系数。

深度神经网络每一层的参数更新会导致上一层的输入数据分布发生变化，通过层层叠加，使输入数据不再满足独立同分布的假设，使每一层的输入数据的整体分布向非线性激活函数取值区间的上下两端偏移，从而导致梯度消失。为了降低分布变化的影响，可使用归一化策略，把数据分布映射到一个确定的区间，以防止梯度消失。此方式可以让模型更好地拟合真实数据，使网络训练更加稳定。如图 9 所示，本文转换网络中添加了归一化层，通过学习参数 $\frac{|\sigma[\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{x})]|}{|\sigma[\boldsymbol{M}^{(c)}_{\text{conv}}(\boldsymbol{y})]|}$ 和 μ 来调整数据分布，可以解决

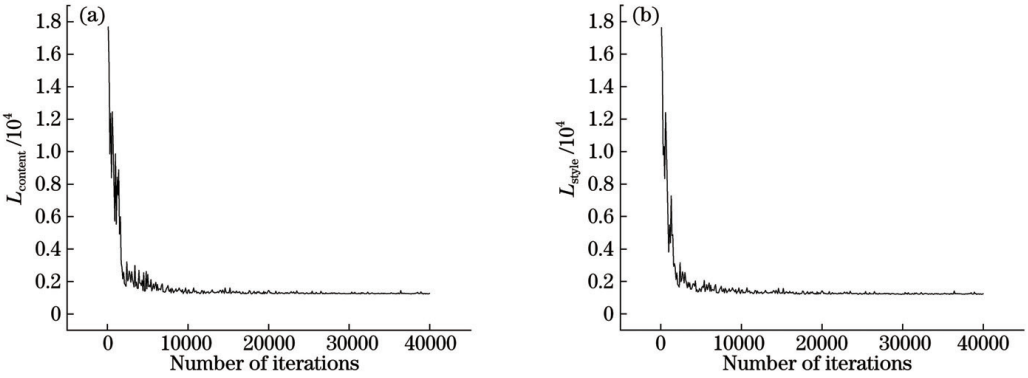

图 9　损失函数收敛图。(a)内容图像;(b)风格图像

Fig. 9　Graphs of loss functions. (a) Content image; (b) style image

内部协变量的偏移,使损失函数收敛速度大幅提升,误差小,网络性能好。

转换网络的所有卷积权重可以在多种风格之间共享,对每种风格进行归一化可使风格图像和内容图像的特征统计量互相适应,从而输出不同风格的图像。网络训练的实质是学习各种参量的过程。对两幅图像进行归一化特征匹配实现参数共享,可减少网络需要学习的参数量,进而减少计算量。在风格迁移模型中,特征图各通道的均值和方差会影响最终生成图像的风格。通过(6)式实现风格图像和内容图像均值和方差的匹配,并在编解码时分别进行归一化与去归一化,获得目标图像的风格,最终实现风格迁移。由图 10 可以看出:添加归一化之前图像细节太平滑,前后景边缘轮廓模糊;归一化后生成图像的笔触较小,纹理精细,边缘轮廓得以增强,使得风格化后的图像整体视觉效果更有层次感。

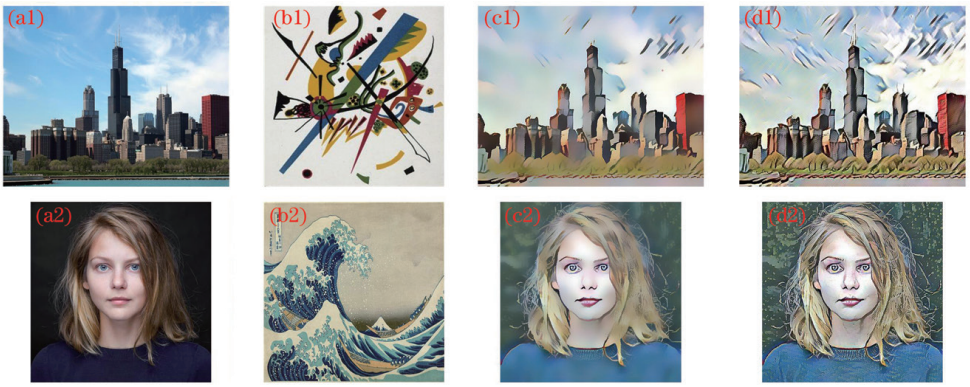

图 10　纹理比较。(a1)(a2)内容图像;(b1)(b2)风格图像;(c1)(c2)归一化前;(d1)(d2)归一化后

Fig. 10　Texture comparison. (a1)(a2) Content images; (b1)(b2) style images; (c1)(c2) before normalization; (d1)(d2) after normalization

在训练过程中,调整(11)式中的内容损失和风格损失的权重比 η 来控制样式转移的程度。如图 11 所示:$\eta=0$ 时,重建内容图像,未进行风格化;$\eta=1.00$ 时,风格化程度最大。通过调整内容损失和风格损失的权重比 η,使其在 0 到 1 之间变化,能够观察到内容相似度和风格相似度的过渡变化。

图 11　风格和内容权衡效果图。(a) $\eta=0$;(b) $\eta=0.25$;(c) $\eta=0.50$;(d) $\eta=0.75$;(e) $\eta=1.00$;(f)风格图像

Fig. 11　Effect of style and content trade-off. (a) $\eta=0$; (b) $\eta=0.25$; (c) $\eta=0.50$; (d) $\eta=0.75$; (e) $\eta=1.00$; (f) style image

4　实验设置与结果分析

4.1　实验设置

在迁移网络中，用高斯白噪声进行网络的初始化，利用 VGG-19 网络中 Relu4_4 和 Relu5_4 层的激活函数值计算内容损失，利用 Relu1_2、Relu2_2、Relu3_4、Relu4_4 和 Relu5_4 层计算风格损失，(11)式中默认参数 α，β，γ 分别为 1×10^{-4}，1×10^{-4}，1×10^{-3}。在抠图网络训练过程中，VGG-19 用作编码器进行特征提取，网络具有 4 通道输入，对网络的 Trimap 通道的第一层卷积进行零初始化。解码器网络参数均使用 Xavier 随机变量作初始化处理。本次实验是基于 TensorFlow 的框架搭建的，设备为 Intel I9 9900K 5.0 GHz，RAM 为 64G，NVIDIA RTX 2080Ti。使用 Adam 优化器[25]进行优化，内容样式图像的批处理大小为 16，学习率为 1×10^{-4}，对其进行 40000 次迭代优化。

4.2　主观分析

图像风格迁移过程可理解为两幅图像中色彩和纹理的融合，也就是尽可能地将内容图像的语义内容用风格图像的色彩纹理去表示。为了证明本文算法的有效性，图 12 展示了本文算法的效果图，并将这些方法的风格化结果与 Gatys 等[5]的优化方法(灵活但缓慢)、Ulyanov 等[10]的优化方法(快速但不灵活)、Huang 等[12]的自适应实例规范化方法、Chen 等[26]的碎片交换方法、Li 等[27]的白化和着色变换方法进行了对比，实验分 4 组，结果如图 13 所示。

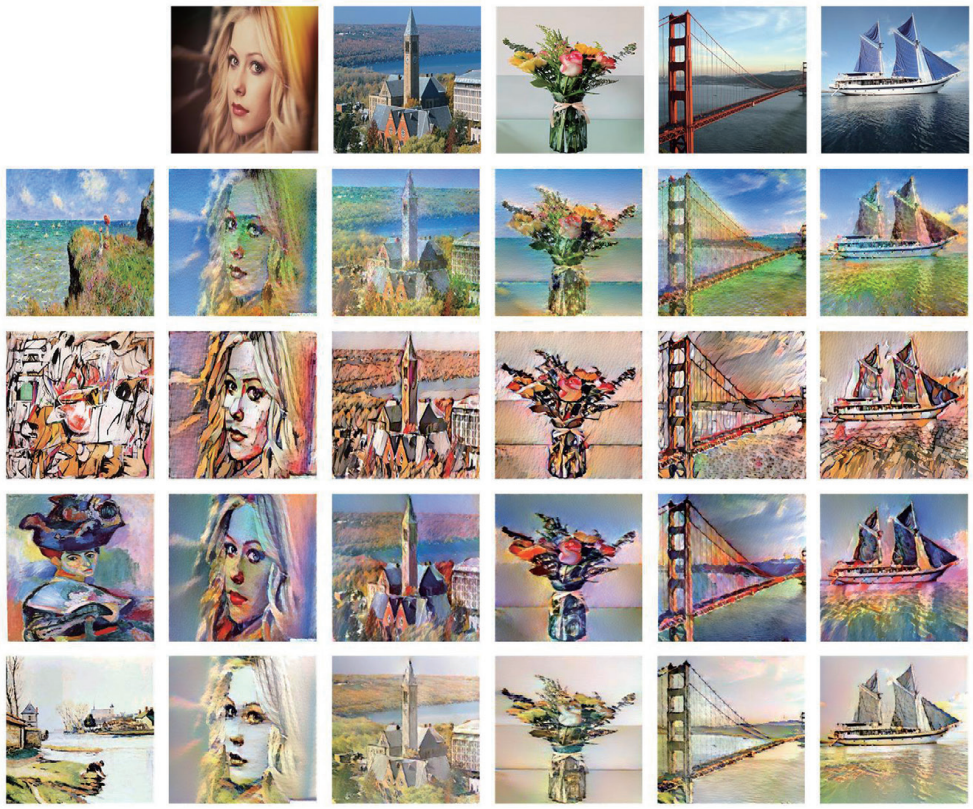

图 12　本文迁移结果

Fig. 12　Transfer results of this article

在神经风格迁移中，Gatys 等[5]最早提出运用神经网络实现神经风格迁移的思路，实现了任意风格的迁移，但是风格化速度较慢且风格化结果不稳定，出现了伪影和图像内容扭曲[图 13(c3)和图 13(c4)]。Ulyanov 等[10]的方法是一种快速的风格迁移方法，迁移结果中颜色纹理表示很好，但是这种训练模型只能迁移一种风格，不能扩展到新的风格[图 13(d)]。Huang 等[12]的方法将同一层的内容特征映射的均值和方差与风格特征映射的均值和方差对齐，能够快速实现任意风格的转换，灵活性较高，但是风格化图像有细小的颗粒覆盖，会导致细节模糊，且存在内容扭曲[图 13(e3)]，对于风格纹理复杂的迁移效果较差。与图 13(c)、图 13(d)的图像相比，图 13(e)的图像质量较差。Chen 等[26]将每个内容碎片与其最匹配的风格碎片相匹配，图像的扭曲较小，但是内容图像没有与风格图像很好地结合，

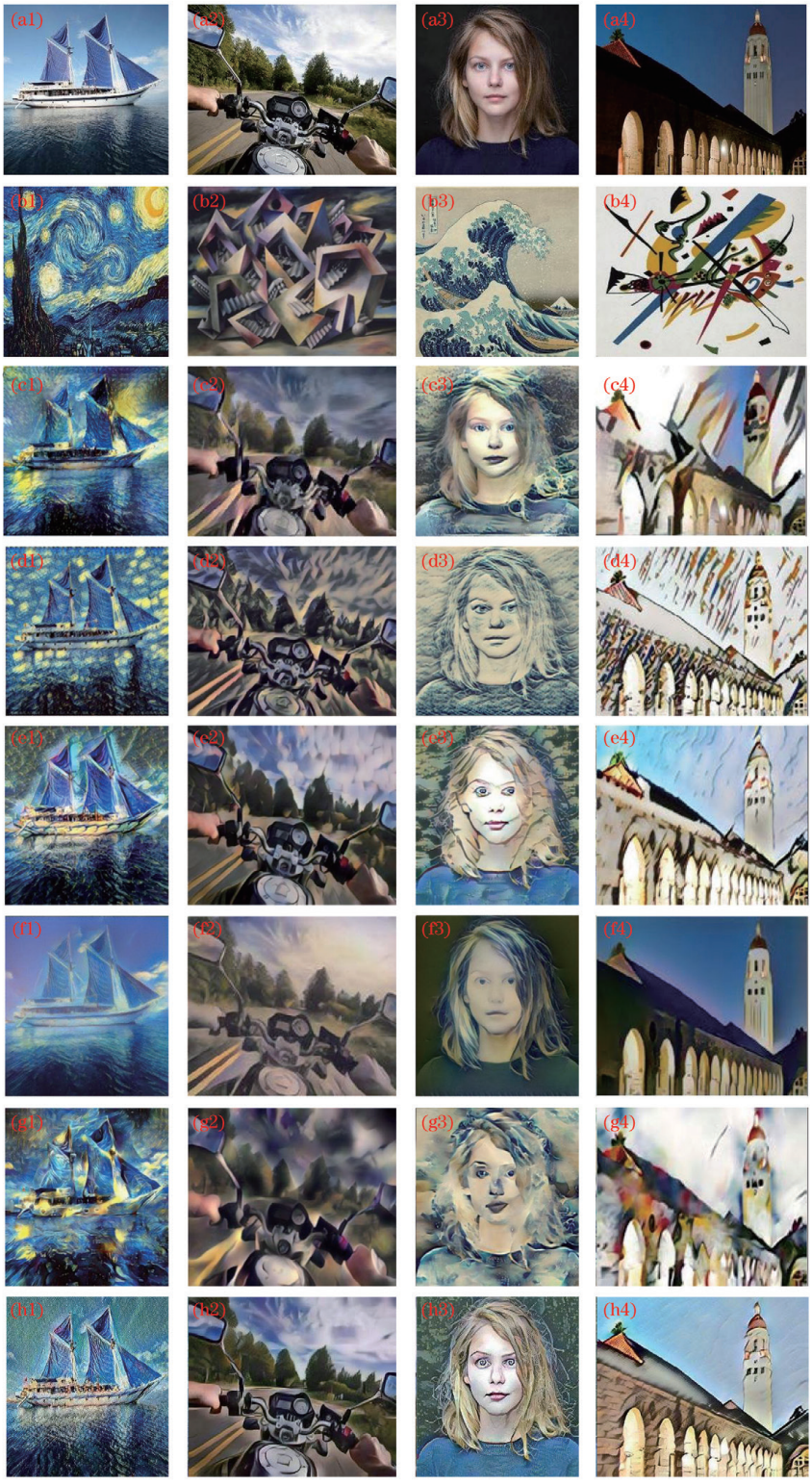

图 13 实验结果对比。(a1)～(a4)内容图像;(b1)～(b4)风格图像;(c1)～(c4) Gatys 等[5]算法的图像;(d1)～(d4) Ulyanov 等[10]算法的图像;(e1)～(e4) Huang 等[12]算法的图像;(f1)～(f4) Chen 等[25]算法的图像;(g1)～(g4) Li 等[26]算法的图像;(h1)～(h4) 本文算法的图像

Fig. 13 Comparison of experimental results. (a1)-(a4) Content images; (b1)-(b4) style images; (c1)-(c4) images based on method proposed by Gatys *et al*[5]; (d1)-(d4) images based on method proposed by Ulyanov *et al*[10]; (e1)-(e4) images based on method proposed by Huang *et al*[12]; (f1)-(f4) images based on method proposed by Chen *et al*[26]; (g1)-(g4) images based on method proposed by Li *et al*[27]; (h1)-(h4) images based on our method

纹理表示和颜色表示效果都较差，风格化效果不理想。Li 等[27]的方法用白化和着色转换的思想将内容图像的特征协方差与给定样式图像的特征协方差直接匹配，它可以将学习到的一种风格扩展为多种风格，但是不能产生有效的清晰细节和细微笔触[图 13(g1)]，对于人脸这类易变形的输入，迁移结果产生了语义内容扭曲[图 13(g3)和图 13(g4)]。

图 13(h)为本文风格迁移的结果图，由图 13(h)可以看出，通过将抠图遮罩和风格迁移合并，本文迁移结果前后景及周围主要景物的边界清晰，主要物体清晰可辨，既结合了风格图像的纹理和色彩，也没有出现多余的纹理分布，最大程度地保留了内容图像的语义内容，语义内容扭曲程度很小。

4.3　客观分析

通过对风格化图像的主观评价，可以直观地感受算法性能的好坏。视觉评价常常带有一定的主观性，而单一的客观评价指标无法全面衡量图像的质量，因此本研究采用以下 4 个指标作为评价标准来定量评价不同迁移方法的性能。

MSSIM 表示内容图像与迁移图像的结构相似性(SSIM)值[28]和风格图像与迁移图像的 SSIM 值的平均值，SSIM 用于表征两幅图像的结构相似性，数值范围为[0，1]，值越接近 1，表示迁移图像与两幅图像越相似。

图像 $\boldsymbol{X}$ 与图像 $\boldsymbol{F}$ 的结构相似性定义为

$$\zeta_{\mathrm{SSIM\_}\boldsymbol{X},\boldsymbol{F}}=\sum_{x,f}\frac{2\mu_x\mu_f+C_1}{\mu_x^2+\mu_f^2+C_1}\cdot\frac{2\sigma_x\sigma_f+C_2}{\sigma_x^2+\sigma_f^2+C_2}\cdot\frac{\sigma_{xf}+C_3}{\sigma_x\sigma_f+C_3},\tag{12}$$

式中：μ_x、μ_f 分别表示图像 $\boldsymbol{X}$ 和 $\boldsymbol{F}$ 的均值；σ_x、σ_f 分别表示图像 $\boldsymbol{X}$ 和 $\boldsymbol{F}$ 的标准差；σ_x^2、σ_f^2 分别表示图像 $\boldsymbol{X}$ 和 $\boldsymbol{F}$ 的方差；σ_{xf} 代表图像 $\boldsymbol{X}$ 和 $\boldsymbol{F}$ 的协方差；C_1，C_2 和 C_3 为常数，是为了避免分母为 0 而维持稳定。通常取 $C_1=(K_1\times L)^2$，$C_2=(K_2\times L)^2$，$C_3=C_2/2$，一般地 $K_1=0.01$，$K_2=0.03$，$L=255$。

平均结构相似度的计算公式为

$$\zeta_{\mathrm{MSSIM}}=0.5\times\zeta_{\mathrm{SSIM\_}\boldsymbol{C},\boldsymbol{F}}+0.5\times\zeta_{\mathrm{SSIM\_}\boldsymbol{S},\boldsymbol{F}},\tag{13}$$

其中，$\zeta_{\mathrm{SSIM\_}\boldsymbol{C},\boldsymbol{F}}$ 和 $\zeta_{\mathrm{SSIM\_}\boldsymbol{S},\boldsymbol{F}}$ 分别表示内容图像 $\boldsymbol{C}$ 与风格化图像 $\boldsymbol{F}$、风格图像 $\boldsymbol{S}$ 与风格化图像 $\boldsymbol{F}$ 的结构相似性。

峰值信噪比(PSNR)由图像信号峰值与均方误差来决定，能够反映图像的失真程度。图像 $\boldsymbol{X}$ 与参考图像 $\boldsymbol{R}$ 的峰值信噪比定义为

$$R_{\mathrm{PSN\_}\boldsymbol{X},\boldsymbol{R}}=10\log\frac{k^2}{\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[X(i,j)-R(i,j)]^2},\tag{14}$$

其中，M，N 为图像尺寸，(i,j)为像素位置，k 为图像的最大灰度等级。本实验的峰值信噪比的计算公式为

$$R_{\mathrm{PSN}}=(R_{\mathrm{PSN\_}\boldsymbol{C},\boldsymbol{F}}+R_{\mathrm{PSN\_}\boldsymbol{S},\boldsymbol{F}})/2,\tag{15}$$

其中，$R_{\mathrm{PSN\_}\boldsymbol{C},\boldsymbol{F}}$ 和 $R_{\mathrm{PSN\_}\boldsymbol{S},\boldsymbol{F}}$ 分别为以内容图像 $\boldsymbol{C}$ 与风格图像 $\boldsymbol{S}$ 为参考图像时，风格化图像 $\boldsymbol{F}$ 的峰值信噪比。PSNR 的值越大，风格化图像的效果越好。

平均梯度(AvG)量化了迁移图像的梯度信息，并可表征其细节和纹理特征。AvG 的值越大，迁移图像包含的细节纹理越多。设 $\boldsymbol{F}$ 是大小为 $M\times N$ 的图像，$\boldsymbol{F}$ 在位置(i,j)处的灰度值为 $F(i,j)$，则平均梯度的表达式为

$$\Delta_{\mathrm{AvG}}=\frac{1}{(M-1)(N-1)}\sum_{i-1}^{M-1}\sum_{j-1}^{N-1}\sqrt{\frac{\left[\frac{\partial F(i,j)}{\partial i}\right]^2+\left[\frac{\partial F(i,j)}{\partial j}\right]^2}{2}}。\tag{16}$$

$Q^{\mathrm{AB/F}}$[29]利用局部度量来估计来自输入显著信息在风格化图像中的表现程度，$Q^{\mathrm{AB/F}}$ 的值越高，表示风格化图像的质量越好，定义式为

$$Q^{\mathrm{AB/F}}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[Q^{\mathrm{AF}}(i,j)\omega^{\mathrm{A}}(i,j)+Q^{\mathrm{BF}}(i,j)\omega^{\mathrm{B}}(i,j)]}{\sum_{i=1}^{M}\sum_{j=1}^{N}[\omega^{\mathrm{A}}(i,j)+\omega^{\mathrm{B}}(i,j)]},\tag{17}$$

其中，Q^{AF}和Q^{BF}为边缘强度和方向的保留值，ω^{A}和ω^{B}表示源图像对于风格化图像重要性的权重。$Q^{AB/F}$的数值越接近 1，代表边缘信息的保留效果越好。

图 14 给出了 6 种方法的客观评价数据，Chen 等[26]的方法的平均结构相似性和峰值信噪比明显高于其他方法，同时它的平均梯度和$Q_{AB/F}$也最小，这说明此方法最大程度地保留了内容图像的语义内容，图像失真程度最小，同时也反映出内容图像没有和风格图像的纹理色彩相结合，没有真正实现风格迁移。本文方法与其他 4 种方法相比较，无论在平均结构相似性、峰值信噪比、平均梯度还是$Q^{AB/F}$上都有较好的表现，这说明本文方法在保留语义内容和结合色彩纹理方面都呈现出了较好的效果，最大程度地实现了图像风格化。

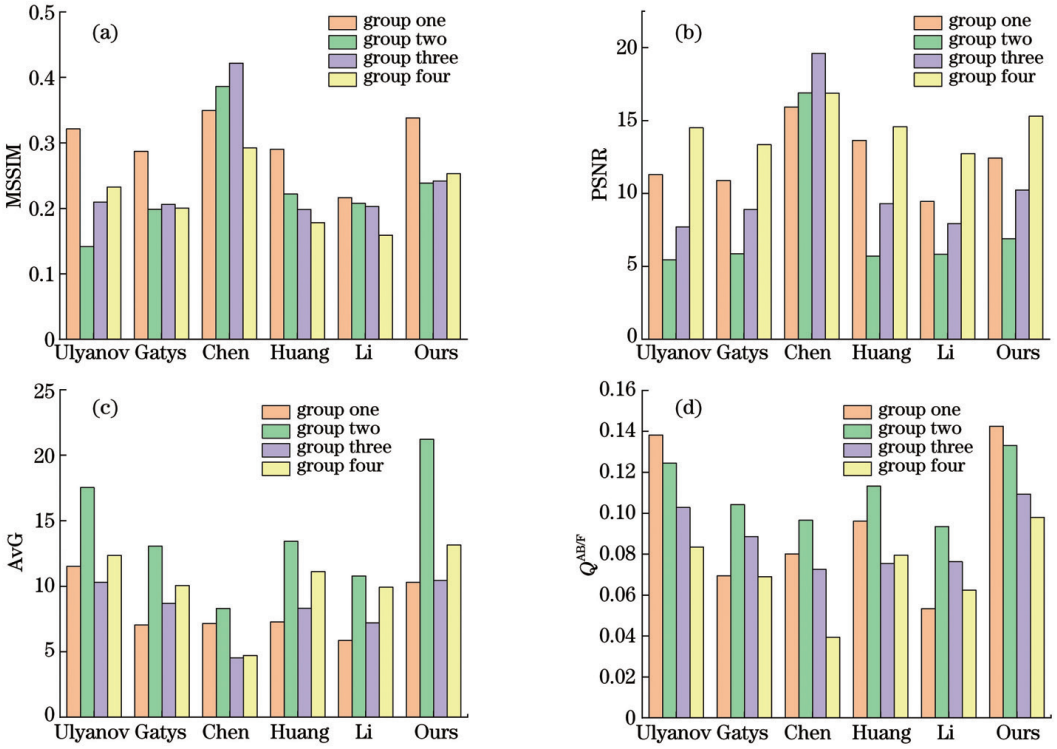

图 14　客观评价。(a)平均结构相似性；(b)峰值信噪比；(c)平均梯度；(d) $Q^{AB/F}$

Fig. 14　Objective evaluation. (a) MSSIM; (b) PSNR; (c) AvG; (d) $Q^{AB/F}$

5　结　　论

将基于神经网络的抠图网络与神经风格迁移模型合并起来，通过抠图获得精确遮罩，使风格转换网络过程凸显主要目标轮廓；在迁移网络中，本研究将其常规卷积层的大卷积核进行了替换，细化了迁移结果的纹理，使得风格化图像纹理更加清晰；将迁移网络中的最大池化层换成平均池化层，减少了邻域大小受限造成的估计方差增大量，保留了更多的背景信息；在常规卷层后添加了归一化层，细化了风格化结果，一定程度上保留了图像的结构布局。本文算法解决了风格迁移过程中图像重建造成的细节信息丢失和前后景边界模糊的问题。通过客观数据表明，本文算法能更好地保留内容结构，细化语义信息，结合风格纹理。

参　考　文　献

[1] Pei F, Liu J F, Li X H, et al. An image style transformation model compression algorithm for mobile terminal [J]. Laser & Optoelectronics Progress, 2020, 57(6): 061021.
裴斐, 刘进锋, 李峭河, 等. 一种面向移动端的图像风格迁移模型压缩算法[J]. 激光与光电子学进展, 2020, 57(6): 061021.

[2] Sun J G, Liu X S. Local style migration method based on residual neural network [J]. Laser & Optoelectronics Progress, 2020, 57(8): 081012.
孙劲光, 刘鑫松. 基于残差式神经网络的局部风格迁移方法[J]. 激光与光电子学进展, 2020, 57(8): 081012.

[3] Kyprianidis J E, Collomosse J, Wang T H, et al. State of the "art": a taxonomy of artistic stylization techniques for images and video[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(5): 866-885.

[4] Jing Y, Yang Y, Feng Z, et al. Neural style transfer: a review[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(11): 3365-3385.

[5] Gatys L A, Ecker A S, Bethge M, et al. Image style transfer using convolutional neural networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2414-2423.

[6] Gatys L, Ecker A, Bethge M, et al. A neural algorithm of artistic style[J]. Journal of Vision, 2016, 16(12): 326.

[7] Gatys L A, Ecker A S, Bethge M, et al. Controlling perceptual factors in neural style transfer[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3730-3738.

[8] Caracalla H, Roebel A. Sound texture synthesis using convolutional neural networks[EB/OL]. (2019-05-09)[2020-08-07]. https://arxiv.org/abs/1905.03637.

[9] Gatys L A, Bethge M, Hertzmann A, et al. Preserving color in neural artistic style transfer[EB/OL]. (2016-06-19)[2020-08-07]. https://arxiv.org/abs/1606.05897v1.

[10] Ulyanov D, Vedaldi A, Lempitsky V, et al. Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4105-4113.

[11] Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: feed-forward synthesis of textures and stylized images[EB/OL]. (2016-03-10)[2020-08-07]. https://arxiv.org/abs/1603.03417.

[12] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C]// 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 1510-1519.

[13] Li C, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2479-2486.

[14] Mechrez R, Talmi I, Manor L Z, et al. The contextual loss for image transformation with non-aligned data[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 800-815.

[15] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[M]// Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2018, 9906: 694-711.

[16] Liu X C, Cheng M M, Lai Y K, et al. Depth-aware neural style transfer[C]//NPAR′17: Proceedings of the Symposium on Non-Photorealistic Animation and Rendering, July 29-30, 2017, New York, NY, United States. New York: ACM Inc, 2017: 1-10.

[17] Chen Y, Lai Y K, Liu Y J, et al. CartoonGAN: generative adversarial networks for photo cartoonization[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9465-9474.

[18] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition[C]// 2015 International Conference on Learning Representations, May 7, 2015, San Diego, CA, USA. San Diego: The Hilton San Diego Resort & Spa, 2015: 1150-1210.

[19] Yao L S, Xu G M, Zhao F, et al. Facial expression recognition based on local feature fusion of convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041513.
姚丽莎, 徐国明, 赵凤, 等. 基于卷积神经网络局部特征融合的人脸表情识别[J]. 激光与光电子学进展, 2020, 57(4): 041513.

[20] Xu N, Price B, Cohen S, et al. Deep image matting[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 311-320.

[21] Levin A, Lischinski D, Weiss Y, et al. A closed-form solution to natural image matting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 228-242.

[22] Tang H, Huang Y J, Jing M E, et al. Very deep residual network for image matting[C]//2019 IEEE International Conference on Image Processing (ICIP), September 22-25, 2019, Taipei, Taiwan, China. New York: IEEE Press, 2019: 4255-4259.

[23] Radford A, Metz L, Chintala S, et al. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2015-11-19)[2020-08-07]. https://arxiv.org/abs/1511.

06434.

[24] Luan F J, Paris S, Shechtman E, et al. Deep photo style transfer[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6997-7005.

[25] Kingma D, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2017-01-30)[2020-08-07]. https://arxiv.org/abs/1412.6980v4.

[26] Chen T Q, Schmidt M. Fast patch-based style transfer of arbitrary style[EB/OL]. (2016-12-13)[2020-08-07]. https://arxiv.org/abs/1612.04337.

[27] Li Y, Fang C, Yang J, et al. Grammar transfer in a second order recurrent neural network[M]// Dietterich T G, Becker S, Ghahramani Z, et al. Advances in neural information processing systems 14. Cambridge: The MIT Press, 2002.

[28] 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.

[29] Piella G, Heijmans H. A new quality metric for image fusion[C]// Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429), September 14-17, 2003, Barcelona, Spain. New York: IEEE Press, 2003: III-173.