基于改进 CycleGAN 网络的图像风格迁移技术研究

吴建磊[1] 杨慧炯[2]
WU Jianlei YANG Huijiong

摘 要 针对 CycleGAN 网络进行图像风格迁移时生成图像存在随机噪音和风格纹理色彩效果较差的问题，文章提出了一种基于 CycleGAN 网络的改进型图像风格迁移方法，分别对生成器的网络结构和损失函数进行改进。网络结构方面，将多头注意力机制加入到生成器中编码器的原始卷积模块中。损失函数方面，在原有损失函数的基础上加入内容损失项和颜色重建损失项。实验结果表明，所提方法生成的图像色彩效果更佳、细节刻画更为丰富，有效避免了生成图像具有随机噪音的问题。并且生成图像的 PSNR 和 SSIM 分别提升了 2.37% 和 12.05%。

关键词 风格迁移；生成对抗网络；CycleGAN；注意力机制；损失函数

doi：10.3969/j.issn.1672-9528.2025.02.019

0 引言

图像风格迁移技术是将一幅图像的风格特征应用到另一幅图像上，从而生成具有新风格的艺术作品。随着深度学习的发展，实现风格迁移的方法大致分为基于神经网络和基于对抗生成网络两种方法。

基于神经网络的图像风格迁移最早由 Gatys 等人[1]提出，其利用 Gram 矩阵将图像表示为内容和风格两部分，通过图像重建使内容图的 Gram 矩阵逼近风格图的 Gram 矩阵，来实现风格迁移。但这种方法需要大量计算资源，且结果可能存在不自然的伪影。Ulyanov 等人[2]引入实例归一化技术，虽然有效改善了纹理重复和颜色失真问题，但有时会导致内容图像结构扭曲。Goodfellow 等人[3]提出生成对抗网络（GAN）后，为图像风格转换提供了新思路，掀起了新研究热潮。Isola 等人[4]的 Pix2Pix 模型则利用输入图像而非噪声进行风格迁移，提升了可控性。但 Pix2Pix 需要配对数据集，这在实际应用中可能难以获得。为了解决这个问题，Zhu 等人[5]提出了循环生成对抗网络（CycleGAN），无须配对数据集，并具备一定的泛化能力，但仍存在生成图像细节模糊、风格纹理的色彩效果较差的问题。针对上述 CycleGAN 网络存在的问题，本文提出以下改进方法：（1）将多头注意力机制加入到生成器中编码器的原始卷积模块中，减少生成图像的随机噪音。（2）引入内容损失函数，保留目标图像的内容特征，进一步提升图像质量。（3）引入颜色重建损失函数，改善生成图像的风格纹理色彩效果，使其更符合人类视觉感知。

1. 太原师范学院计算机科学与技术学院 山西晋中 030600
2. 太原工业学院计算机工程系 山西太原 030008

1 相关工作

1.1 CycleGAN 网络

循环生成对抗网络（CycleGAN）是一种基于 GAN 的算法，主要用于在没有成对训练样本的情况，实现两个不同域之间的图像风格转换。CycleGAN 的核心思想是通过学习两个域之间的映射关系，使得一个域的图像能够转换成另一个域的图像，同时保持原始图像的内容信息。CycleGAN 架构由两对生成器和判别器组成，通过引入循环一致性损失函数，将网络分为两个对称的 GAN 网络。如图 1 所示。一个专注于将图像从源域转换至目标域，而另一个则专注于将图像从目标域逆向转换回源域。

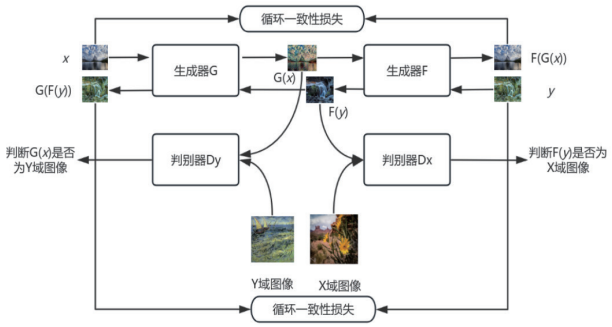

图 1 CycleGAN 网络结构

CycleGAN 的对抗损失函数由两部分组成，分别对应两个 GAN，其定义为：

$$L_{\mathrm{GAN}}(G,F,D_x,D_y)=L_{\mathrm{GAN}}(G,D_y)+L_{\mathrm{GAN}}(F,D_x) \quad (1)$$

式中：$L_{\mathrm{GAN}}(G,D_y)$为 X 域到 Y 域的映射；$L_{\mathrm{GAN}}(F,D_x)$为 Y 域到 X 域的映射，其表达式分别为：

$$L_{\mathrm{GAN}}(G,D_y)=E_{y\sim \mathrm{Pdata}(y)}[\log D_y(y)]+E_{x\sim \mathrm{Pdata}(x)}[\log(1-D_y(G(x))] \quad (2)$$

$$L_{\mathrm{GAN}}(F, D_x) = E_{x\sim \mathrm{Pdata}(x)}[\log D_x(x)] + E_{y\sim \mathrm{Pdata}(y)}[\log(1 - D_x(F(y)))] \quad (3)$$

CycleGAN 循环一致性损失函数：

$$L_{\mathrm{cyc}}(G, F) = E_{x\sim \mathrm{Pdata}(x)}[||F(G(x)) - x||_1] + E_{y\sim \mathrm{Pdata}(y)}[||G(F(y)) - y||_1] \quad (4)$$

式中：G 和 F 分别是从源域到目标域和从目标域到源域的生成器；D_y 判断输入的图像 $G(x)$ 是否为 Y 域图像；D_x 判断输入的图像 $F(y)$ 是否为 X 域图像；x 和 y 分别为源域和目标域的图像；pdata(x) 和 pdata(y) 分别是源域和目标域的数据分布；$\|\cdots\|_1$ 表示 L_1 范数，用于衡量两个图像之间的差异。

1.2 注意力机制

在深度学习领域，注意力机制（attention mechanism，AM）是一种模拟人类视觉注意力特点的方法，能够使模型更加关注图像中的重要信息，从而提高特征表示的准确性。

1.2.1 自注意力机制

自注意力机制 [6]（self-attention mechanism，SAM）允许模型在处理一个序列时，能够同时考虑序列中的每个元素与其他所有元素之间的关系。其结构如图 2 所示。

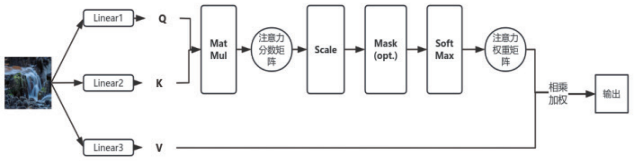

图 2 自注意机制结构

首先图像通过 3 个不同的线性变换分别得到其对应的查询（Query）、键（Key）和值（Value）3 个向量。然后，对查询向量 $\boldsymbol{Q}$ 和键向量 $\boldsymbol{K}$ 进行点积计算（MatMul），从而形成一个注意力分数矩阵，这些分数表示序列中每个元素与其他所有元素的相关性。由于点积的结果较大，导致在后续 SoftMax 函数中计算得到的梯度较小。为避免这一问题，通常在点积操作后引入一个缩放因子，对注意力权重进行缩放操作（Scale），其中，d^k 是键向量 $\boldsymbol{K}$ 的维度。接下来，使用 SoftMax 函数对注意力得分进行归一化，使得所有得分和等于 1，这样，每个得分都可以被视为一个概率值。从而得到了注意力权重（attention weights，AW）。最终，注意力权重与对应的值向量 $\boldsymbol{V}$ 相乘，并对结果进行累加，从而得到自注意力机制的最终输出。整个自注意力机制可以用公式概括为：

$$\mathrm{Self\ Attention}(x) = \mathrm{SoftMax}(\frac{QK^T}{\sqrt{d}})V \quad (5)$$

1.2.2 多头注意力机制

多头注意力机制（multi-head attention，MHA）是自注意力机制的一种扩展，它将自注意力拆分为多个“头”，每个“头”都有自己的参数集，可以学习到输入序列的不同表示。其结构如图 3 所示。首先，多头注意力机制将输入图像通过多个独立的线性变换拆分成多个“头”，每个头都有自己的权重矩阵，且每个头独立进行自注意力机制的计算。接着将所有头的输出连接起来并通过一个额外的线性层进行变换，以组合不同头捕获的信息。最终得到的多头注意力输出将包含来自不同子空间的信息，这有助于模型更好理解序列数据。其结构如图 3 所示。

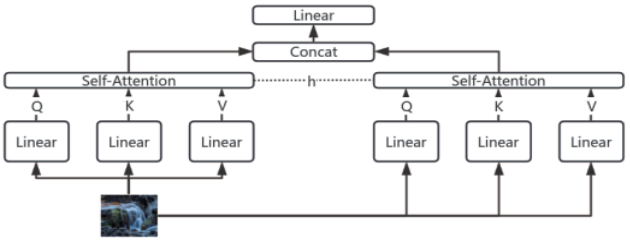

图 3 多头注意机制结构

2 改进方法

2.1 生成器改进

在 Zhu 等人的研究中，CycleGAN 生成器网络由编码器、残差结构、解码器 3 部分组成。生成器的网络结构如图 4 所示，其中，实线所连为生成器网络的具体工作流程，虚线所连为相应模块的具体组成结构。

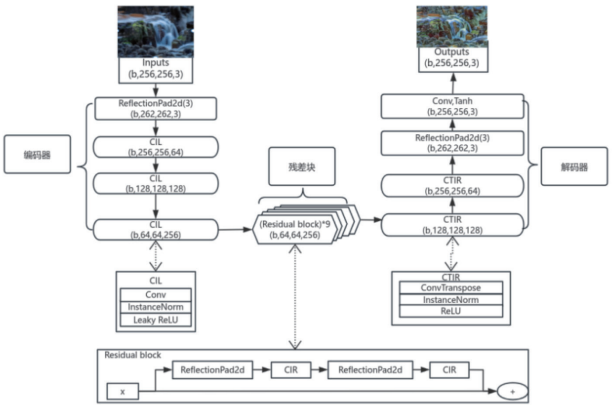

图 4 生成器网络结构

编码器负责对输入图像进行预处理和特征提取，整体流程包括反射填充和下采样两个主要步骤。反射填充通过用 PyTorch 框架中的 ReflectionPad2d 操作实现；下采样通过三个 CIL 模块实现，这些 CIL 模块由卷积层、实例归一化层和 Leaky ReLU 激活函数组成，其作用是逐层提取图像的特征，并减小特征图的尺寸，以适应深层网络的结构和处理需求。

残差结构通过堆叠 9 个残差块，实现深层特征的学习和特征复用。每个残差块内部包含两次反射填充和两个 CIR 模块，后者由卷积、实例归一化和 ReLU 激活函数构成。

解码器负责对抽象特征图进行尺寸恢复和图像后处理，整体流程包括上采样、反射填充和图像重建。其中上采样通过两个 CTIR 模块实现，这些模块由转置卷积层、实例归一

化层和 ReLU 激活函数组成；反射填充通过 ReflectionPad2d 操作实现；最后，通过一个卷积操作将图像恢复到原始输入图像的大小，并使用 Tanh 激活函数来限制输出像素值的范围，使其在 [-1,1] 之间，完成图像的生成过程。

受 Zhang 等人 [7] 研究启发，本文在 CIL 模块中的实例归一化层后加入两个头的多头注意力机制，帮助生成器模型更好地捕捉到特征之间的区分性，使得 CIL 模块能够生成更加精细和有区分度的特征表示，并保留更多关键信息。从而提高整个网络的性能。如图 5 所示。

图 5 卷积模块改进

相较于 Zhang 等人所使用的自注意力机制，多头注意力机制有如下优点：

（1）更强的模型表达能力，多头注意力机制使得模型能够在不同的子空间中学习到不同的特征表示，从而捕捉到多样化的信息，能够更全面地理解序列数据。

（2）更高的模型泛化能力，降低模型对特定特征的依赖，从而在一定程度上减少过拟合的风险。

2.2 损失函数改进

CycleGAN 网络中对抗损失主要关注于生成图像的整体风格和外观，而不是具体的像素级内容，因此生成器可能会忽略输入图像的某些细节。循环一致性损失虽然试图保留原始内容，但不足以完全恢复所有细节，可能导致生成图像在风格转换的同时失去部分原始内容细节并产生随机噪音。受 Johnson 等人的启发，为了更好地保留目标图像的内容细节，减少生成图像噪音，本文在原有损失函数的基础上引入内容损失。同时为更准确地模拟人类视觉系统对颜色的感知，并且能够更有效地处理和调整图像的颜色，本文还引入了颜色重建损失 [8]。

本文使用预训练的 VGG-19 网络提取出的特征图，通过计算生成图像特征图与原图像特征图的 L_1 范数来构建内容损失，其定义为：

$$L_{\text{content}}(G,F)=L_{\text{content}}(G)+L_{\text{content}}(F) \tag{6}$$

式中：$L_{\text{content}}(G)$为 X 域到 Y 域的映射；$L_{\text{content}}(F)$为 Y 域到 X 域的映射。其函数表达式分别为：

$$L_{\text{content}}(G)=E_{x\sim \text{Pdata}(x)}[||\text{VGG}(G(x))-\text{VGG}(x)||_1] \tag{7}$$

$$L_{\text{content}}(F)=E_{y\sim \text{Pdata}(y)}[||\text{VGG}(F(y))-\text{VGG}(y)||_1 \tag{8}$$

式中：VGG(…) 表示图像经过预训练的 VGG-19 网络所提取出的特征图，本文采用第 23 层的特征图作为内容表示。

本文将图像由 RGB 色彩空间转换为 LAB 色彩空间，通过分别计算 L、A、B 三个分量的 L_1 范数构建了颜色重建损失。其中 L 分量代表亮度；A 分量代表从绿色到红色的范围；B 分量代表从蓝色到黄色的范围。其定义为：

$$L_{\text{color}}(G,F)=L_{\text{color}}(G)+L_{\text{color}}(F) \tag{9}$$

式中：$L_{\text{color}}(G)$为 X 域到 Y 域的映射；$L_{\text{color}}(F)$为 Y 域到 X 域的映射。其函数表达式分别为：

$$\begin{aligned}L_{\text{color}}(G)=E_{x\sim \text{Pdata}(x)}[&(||L(G(x))-L(x)||_1)\\&+(||A(G(x))-A(x)||_1)+(||B(G(x))-B(x)||_1)]\end{aligned} \tag{10}$$

$$\begin{aligned}L_{\text{color}}(F)=E_{y\sim \text{Pdata}(y)}[&(||L(G(y))-L(y)||_1)\\&+(||A(G(y))-A(y)||_1)+(||B(G(y))-B(y)||_1)]\end{aligned} \tag{11}$$

本文实验的总损失函数为式（1）（4）（8）（11）相加多所得，其定义为：

$$\begin{aligned}L(G,F,D_x,D_y)=&\omega_{\text{GAN}}L_{\text{GAN}}(G,F,D_x,D_y)+\omega_{\text{cyc}}L_{\text{cyc}}(G,F)\\&+\omega_{\text{content}}L_{\text{content}}(G,F)+\omega_{\text{color}}L_{\text{color}}(G,F)\end{aligned} \tag{12}$$

式中：ω_{GAN}、ω_{cyc}、ω_{content}、ω_{color}分别为对抗损失、循环一致性损失、内容损失、颜色重建损失的权重。本文中ω_{GAN}、ω_{cyc}、ω_{content}、ω_{color}分别为 1.0、10.0、3.0、4.0。

3 实验和结果分析

3.1 实验环境与数据集

本文所使用的硬件设备为 NVIDIA GRID T4-8Q GPU 服务器，在此硬件环境下使用基于 Python3.9 的 PyTorch 深度学习框架进行实验训练。使用 vangogh2photo 数据集进行实验，其中包括 775 张梵高油画和 7038 张照片。不同数据集间没有指定配对关系。数据集图像均为 256 px×256 px 的彩色图像。

3.2 结果分析

本实验批量训练样本数量（batchsize）为 1，训练总轮次（epoch）为 300，学习率（learning rate）设置为 0.000 3，保持学习率不变的轮次为 150，在 150 个轮次后使用指数衰减策略对学习率进行动态调整。

3.2.1 消融实验结果

本文使用 vangogh2photo 数据集进行了一系列的消融实验。并通过峰值信噪比（PSNR）和结构相似性（SSIM）对实验中生成的图像进行了客观的质量评价，以此来验证本文每个所提算法改进点的有效性。

从表 1 中可以看出相较于原始的 CycleGAN 网络模型，改进后的生成器和损失函数同时加入到网络中 PSNR 和 SSIM 分别提升了 2.37% 和 12.05%。由此可以得出本文方法能够在一定程度上提升风格迁移图像的质量。

表 1 消融实验结果

评价指标	AdaIN	DiscoGAN	CycleGAN	本文方法
PSNR	27.856	28.038 5	28.080 3	28.746 7
SSIM	0.219 9	0.578 9	0.584 9	0.655 4

3.2.2 对比实验结果

基于上述网络结构及训练策略，将本文所改进的网络与

AdaIN[9]、DiscoGAN[10]、CycleGAN进行对比，结果如图6所示，细节展示如图7所示。

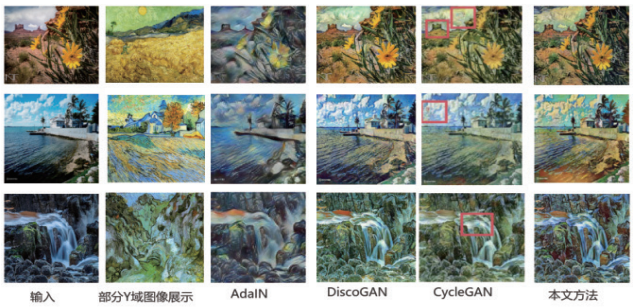

图6 实验结果对比

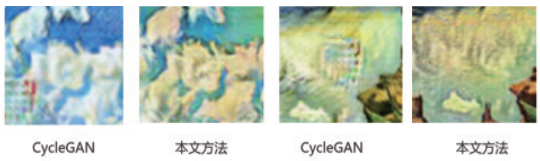

图7 部分细节展示

通过图6可以看出，AdaIN网络内容图像的结构被过度扭曲。DiscoGAN网络的生成结果虽然较好的保留了内容信息但风格化较差。CycleGAN网络的生成结果相较于前者，虽然内容信息的保留和风格化的质量有所提高，但生成图像会出现随机噪音以及仍然存在颜色纹理效果较差的问题。通过对比结果可以看出本文所提出的方法在保留图像细节以及生成更符合人类视觉系统的油画图像方面有很大提升。从图7中可以明显看出本文所提出的改进方法基本解决了CycleGAN网络生成图像存在随机噪音的问题。

本文通过与AdaIN、DiscoGAN、CycleGAN网络进行对比实验，证明本文所提方法的有效性，并以PSNR和SSIM作为评估指标，结果如表2所示。

表2 图像质量评估对比

模型	PSNR	SSIM
CycleGAN	28.080 3	0.584 9
生成器改进	28.115 4	0.617 6
生成器+损失函数改进	28.746 7	0.655 4

从表2可知，本文改进后的CycleGAN网络所生成的图像，PSNR值和SSIM两项指标均高于其他方法，由此说明本文方法相较于上述方法，迁移后的图像质量更佳。

4 结论

本文针对CycleGAN网络进行图像风格迁移时生成图像存在随机噪音和风格纹理色彩效果不理想的问题，提出了一种改进型CycleGAN网络的图像风格迁移方法。通过使用CIML注意力机制卷积模块代替CIL卷积模块并引入内容损失函数，基本上解决了生成图像存在随机噪音的问题。通过引入了颜色重建损失函数改善了生成图像风格纹理色彩效果。

参考文献：

[1] GATYS L A, ECKER A S, BETHGE M.A neural algorithm of artistic style[J].Journal of vision, 2016, 16(9): 326.

[2] ULYANOV D, LEBEDEV V, VEDALDI A, et al. Texture networks: feed-forward synthesis of textures and stylized images[DB/OL].(2016-03-10)[2024-05-19].https://doi.org/10.48550/arXiv.1603.03417.

[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[4] ISOIA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks[C/OL]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway:IEEE,2017[2024-06-19].https://ieeexplore.ieee.org/document/8100115.

[5] ZHU J Y, PARK T, ISOIA P, et al. Unpaired image-to-image translation using cycle-vision consistent adversarial networks[C/OL]// 2017 IEEE International Conference on Computer Vision(ICCV).Piscataway:IEEE,2017[2024-04-13].https://ieeexplore.ieee.org/document/8237506.

[6] ZHAO Y H, WU R H, DONG H.Unpaired image-to-image translation using adversarial consistency loss[DB/OL].(2021-01-18)[2024-04-17].https://doi.org/10.48550/arXiv.2003.04858.

[7] ZHANG H, GOODFELLOW L, METAXAS D, et al.Self-attention generative adversarial networks[DB/OL].(2019-06-14)[2024-05-26].https://doi.org/10.48550/arXiv.1805.08318.

[8] CHEN J, LIU G, CHEN X. AnimeGAN: a novel lightweight GAN for photo animation[J].Artificial intelligence algorithms and applications,2020,1205(5):242-256.

[9] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[DB/OL]. (2017-07-30)[2024-02-22].https://doi.org/10.48550/arXiv.1703.06868.

[10] KIM T, CHA M, KIM H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//ICML'17: Proceedings of the 34th International Conference on Machine Learning. New York: JMLR.org, 2017:1857-1865.

【作者简介】

吴建磊（1997—），男，河北廊坊人，硕士研究生，研究方向：计算机技术、图像处理。

杨慧炯（1972—），男，山西太原人，硕士，教授，研究方向：机器视觉、图形图像处理。

（收稿日期：2024-10-26）