联合图像域间和域内信息建模的图像风格转换

甘益波, 谭智一*, 鲍秉坤

(南京邮电大学通信与信息工程学院 南京 210000)
(tzy@njupt.edu.cn)

摘 要: 针对当前图像风格转换算法缺乏建模图像域间语义信息和域内长范围信息的能力, 提出了一种联合图像域间和域内信息建模的图像风格转换算法 SSC-GAN. 通过提出语义残差连接, 提取图像域内的语义特征, 增强模型建模图像域间语义信息差异的能力; 同时, 将注意力机制引入图像风格转换任务中, 解决卷积缺乏图像域内长范围信息建模能力的问题. SSC-GAN 可以在不增加计算量的情况下, 显著提升图像风格转换的表现. 在图像风格转换数据集 vangh2photo 和 selfie2anime 上对 SSC-GAN 进行训练、评估和验证, 结果表明, SSC-GAN 不仅能取得极佳的视觉效果, 而且在 FID 和 KID 指标上分别平均下降了 1.3 和 1.1, 证明了 SSC-GAN 的有效性.

关键词: 图像风格转换; 生成对抗网络; 残差连接; 注意力机制
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2022.19784

Joint Intra-Domain and Inter-Domain Information Modeling for Image-to-Image Translation

Gan Yibo, Tan Zhiyi*, and Bao BingKun

(*College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing* 210000)

Abstract: To solve the disability of modeling inter-domain semantic and intra-domain long-range information in current image style transferring algorithms this article proposes a novel image style transferring algorithm SSC-GAN. This method extracts the semantic features by constructing the semantic shortcut connections, thereby enhancing its capability of modeling semantic difference between image domains. Meanwhile, the self-attention mechanism is introduced for the modeling of long-range dependency in the image domain. The SSC-GAN can significantly improve the performance of image style transferring without extra computation. Through the extensive experiments on vangh2photo and selfie2anime datasets, the proposed method achieves excellent visual effects and reduces FID and KID by 1.3 and 1.1 on average respectively, which verifies the effectiveness of SSC-GAN.

Key words: image-to-image translation; generative adversarial network; shortcut connection; self-attention

图像风格转换是一类将源域输入图像转换为目标域输出图像的计算机视觉任务. 例如, 从灰度图转换为彩色图, 从梵高风格的图像转换为现实风格的图像. 图像风格转换在诸多领域有着广泛的应用, 如图像修复、图像去噪[1]、图像去雾[2]和图像超分辨率重建[3]等. 图像风格转换的思想由 Hertzmann 等[4]提出, 其使用非参数模型在单一输入输出配对图像中实现了图像风格转换. 由于该

收稿日期: 2022-06-15; 修回日期: 2022-07-14. **基金项目**: 国家重点研发计划(2020AAA0106200); 国家自然科学基金(61936005, 61872424); 江苏省自然科学基金(BK20200037, BK20210595). **甘益波**(1998—), 男, 硕士研究生, 主要研究方向为图像风格转换、生成对抗网络; **谭智一**(1986—), 男, 博士, 讲师, 论文通信作者, 主要研究方向为多媒体计算、多媒体数据挖掘与序列建模、人工智能; **鲍秉坤**(1982—), 女, 博士, 教授, 博士生导师, 主要研究方向为多媒体计算、社交多媒体、计算机视觉、人工智能.

方法在处理不同类型图像时存在较大的局限，因此多年来发展缓慢.

随着深度学习的发展和硬件设备的进步，尤其得益于卷积神经网络在计算机视觉领域取得的巨大成功，基于卷积神经网络[5-10]的图像风格转换取得了较大的进展. 目前，图像风格转换分为 2 个流派：一是以循环一致性生成对抗网络(cycle-consistent generative adversarial network, CycleGAN)循环一致性损失为代表[11-14]的像素层面的损失函数流派，循环一致性损失较好地解决了图像风格转化需要匹配数据集的问题，但存在生成器和鉴别器的冗余的缺陷；二是以 PatchNCE(patch *n* negative cross entropy)为代表[15-19]的特征层面的损失函数流派，其虽然能够改善循环一致性损失的缺陷，但是会混淆图像域内的内容和风格特征. 由于特征层面的损失函数会混淆图像域内特征，因此，本文将基于第 1 种流派进行研究.

近年来，虽然图像风格转换模型取得了较大的进展，但是现有研究表明，它们均未能同时考虑图像域间和域内信息的差异. 在语义信息较少的图像域向语义信息丰富的图像域转换时，图像域间信息的差异会导致生成图像质量欠佳，如素描图像向现实图像转换的难度远远大于现实图像向素描图像转换的难度. 图像域内信息可以分为内容特征和风格特征，它们分别偏向于全局信息和局部信息，而图像风格转换模型由于卷积提取全局特征的能力欠佳，导致模型只能生成结构简单的图像.

为了解决上述问题，本文提出了一种联合图像域间和域内信息建模的图像风格转换算法. 所提出的语义残差连接生成对抗网络 SSC-GAN (semantic shortcut connection generative adversarial network, SSC-GAN)模型通过语义残差连接和注意力机制，分别对图像域间和域内信息进行建模. 针对图像域间语义信息差异较大的问题，受 ResNet 残差连接的启发，本文提出语义残差连接，通过语义编码器提取语义特征用于图像重构，以弥补图像域间的语义信息差异，改善语义信息较低的图像域向语义信息丰富的图像域转换时生成图像效果欠佳的问题；针对图像域内内容特征和风格特征范围不同的问题，本文将能够高效建模长范围信息的注意力机制引入图像风格转换模型，以解决模型难以提取全局特征的问题.

本文的主要贡献如下：

(1) 本文提出语义残差连接，利用语义编码器提取到的语义特征通过语义残差连接补偿生成器，使图像风格转换模型能够对图像域间的语义信息差异建模.

(2) 将注意力机制引入图像风格转换任务，使模型能够对图像全局信息进行建模，提高生成图像的质量.

(3) 大量实验证明，本文提出的方法能够在诸多数据集上表现出色，取得了超越现有方法的结果.

1 相关工作

深度学习算法和计算机算力的突飞猛进使得图像风格转换任务取得了长足的进步，尤其归功于生成对抗网络(generative adversarial network, GAN)的提出. 随后基于 GAN 的 Pix2Pix[20], CycleGAN[21], DRIT++[22], CouncilGAN[23], MUNIT[24], U-GAT-IT[25] 等算法陆续被提出. 基于这些算法，本文提出了一种联合图像域间和域内信息建模的图像风格转换算法.

1.1 GAN

GAN 由 Goodfellow 等[26]在 2014 年提出，是一种无监督的生成式算法，其由生成器和鉴别器 2 个网络组成. GAN 成功的关键在于零和博弈的思想，鉴别器的存在迫使生成图像尽可能与真实图像难以区分，从而使得生成图像的数据分布能够尽可能地靠近真实图像分布. 条件 GAN[27](conditional GAN, CGAN)同时在生成器和鉴别器中引入条件信息，改进了 GAN 不能控制生成图像内容的缺点. GAN 由于卷积算子缺乏对长范围信息建模的能力，只能在生成结构简单的图像时表现良好. Zhang 等[28]在 GAN 的生成器中引入了注意力机制，提出了自注意力 GAN(self-attention GAN, SAGAN)，其能够有效地对全局信息进行建模的同时，不会增加过多的计算量，极大地提升了生成图像的质量. 本文提出的 SSC-GAN 将注意力机制迁移至图像风格转换任务上，并取得了显著的效果.

1.2 图像风格转换

随着 GAN 的提出和完善，基于 GAN 思想的图像风格转换算法的研究如同雨后春笋. Isola 等[20]在 CGAN 和 U-Net[29]的基础上提出了 Pix2Pix，其能够完成给图像上色、从边缘图像中恢复原图等任务. Pix2Pix 的提出使图像风格转换算法不再依赖手工设计特征映射函数，其通过匹配的数据集训练使模型学习输入图像到输出图像之间的映射. 然而，匹配数据集的获取需要大量的人力和物力，

这限制了 Pix2Pix 的应用场景. 为此, Zhu 等[21]提出了 CycleGAN, 其采用 2 个生成器和 2 个鉴别器构成环形网络, 生成器 G 学习图像域 X 到图像域 Y 的映射; 相反, 生成器 F 学习图像域 Y 到图像域 X 的映射. 损失函数部分除了循环一致性损失外, 整个网络沿用 GAN 的对抗损失. 网络通过循环一致性损失尽可能保证输入图像和输入图像先后通过生成器 G 和 F 的输出图像保持一致, 即 $F\left(G\left(X\right)\right)\approx X$. CycleGAN 的提出, 将图像风格转换的研究推向了高潮. 随后, Lee 等[22]在循环一致性的基础上提出了 DRIT++, 其将图像域内的信息分为内容特征和属性特征, 通过内容编码器和属性编码器将图像进行解耦, 不同图像域之间的图像只要通过交换属性特征便可实现风格转换. CouncilGAN[23]利用多个 GAN 的协作去除循环一致性损失, 取得了良好的效果. MUNIT[24]将图像风格转换视为求联合概率分布的问题, 其重点是解决一对多的图像风格转换任务. U-GAT-IT[25]提出了一种在图像风格转换任务上表现良好的归一化方法自适应层归一化(adaptive layer-instance normalization, AdaLIN), 其能够自适应控制风格转换时内容特征和属性特征产生的形变, 并借鉴辅助分类器的类激活映射(class activation mapping, CAM)[30]思想, 提出注意力模块: 利用源域与目标域的辅助分类器得到全连接权重, 对特征图通道进行加权, 较好地处理了不同特征图通道对生成图像贡献不同的问题. 虽然上述方法在图像风格转换任务上均取得了较大的成功, 但其均未考虑图像域间语义差异、图像域内内容特征和风格特征范围不同的问题, 限制了现有模型的性能.

2　本文方法

SSC-GAN 的框架如图 1 所示. 其由 2 个生成器 $G_{s\to t}$ 和 $G_{t\to s}$, 以及 2 个鉴别器 D_s 和 D_t 组成. 其中, $G_{s\to t}$ 表示将源域图像向目标域转换的生成器, D_t 表示目标域图像鉴别器.

本文将语义编码器提取到的语义特征通过残差连接来补偿生成器, 建模图像域间语义信息; 同时, 将注意力机制模块整合到生成器的解码器中, 增强模型对长范围信息建模的能力. 其中, 生成器和鉴别器的详细结构如表 1 和表 2 所示.

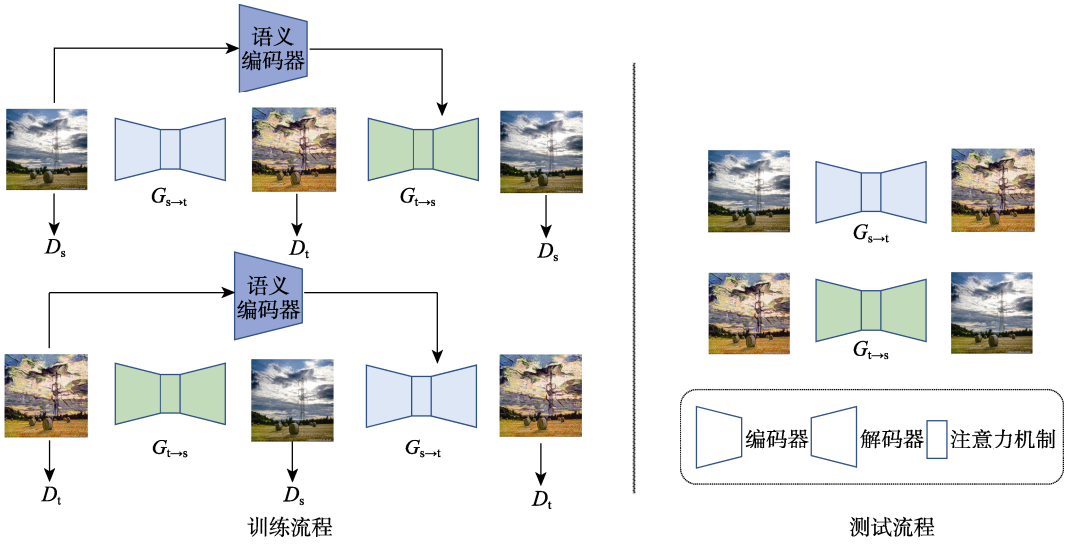


图 1　SSC-GAN 整体框架

2.1　生成器

2.1.1　语义残差连接

循环一致性虽然能够使图像风格转换任务摆脱对匹配数据集的依赖, 但是实验证明图像在循环中的重建效果欠佳. 究其原因是循环一致性无法建模图像域间的语义差异. 因此, 本文受残差连接[31]的启发, 提出语义残差连接解决图像域间语义信息差距大的问题.

如图 2 所示, 输入图像分别通过生成器和语义编码器提取特征, 将语义编码器提取到的语义特征通过残差连接分别补偿 $G_{s\to t}$ 和 $G_{t\to s}$, 以此弥补图像重构过程中的语义差异并且增强循环一致性, 使模型能够对图像域间语义信息差异进行建模. 其中, 残差支路中语义编码器的网络架构和主干中的编码器架构一致.

表 1 生成器网络结构

阶段	输入→输出	参数设置
编码器下采样层	$(h,w,3)\rightarrow(h,w,64)$	CONV-(N_{64}, K_7, S_1, P_3), IN, ReLU
	$(h,w,64)\rightarrow(9h/2,w/2,128)$	CONV-(N_{128}, K_3, $S2$, P_1), IN, ReLU
	$(h/2,w/2,64)\rightarrow(h/4,w/4,128)$	CONV-(N_{256}, K_3, S_2, P_1), IN, ReLU
编码器瓶颈层	$(h/4,w/4,128)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
解码器注意力层	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_1, S_1, P_0), IN, ReLU
解码器瓶颈层	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
	$(h/4,w/4,256)\rightarrow(h/4,w/4,256)$	CONV-(N_{256}, K_3, S_1, P_1), IN, ReLU
解码器上采样层	$(h/4,w/4,256)\rightarrow(h/2,w/2,128)$	UP-CONV-(N_{128}, K_3, S_1, P_1), IN, ReLU
	$(h/2,w/2,128)\rightarrow(h,w,64)$	UP-CONV-(N_{64}, K_3, S_1, P_1), IN, ReLU
	$(h,w,64)\rightarrow(h,w,3)$	UP-CONV-(N_3, K_7, S_1, P_3), IN, ReLU

注. N, K, S, P 分别表示通道数、卷积核大小、步长、填充; IN 表示归一化方式; ReLU 表示激活函数.

表 2 鉴别器网络结构

阶段	输入→输出	参数设置
编码器下采样层	$(h,w,3)\rightarrow(h/2,w/2,64)$	CONV-(N_{64}, K_4, S_2, P_1), SN, Leaky-ReLU
	$(h/2,w/2,64)\rightarrow(h/4,w/4,128)$	CONV-(N_{128}, K_4, S_2, P_1), SN, Leaky-ReLU
	$(h/4,w/4,128)\rightarrow(h/8,w/8,256)$	CONV-(N_{256}, K_4, S_2, P_1), SN, Leaky-ReLU
	$(h/8,w/8,25)\rightarrow(h/8,w/8,256)$	CONV-(N_{512}, K_4, S_2, P_1), SN, Leaky-ReLU
分类器	$(h/8,w/8,512)\rightarrow(h/8,w/8,1)$	CONV-(N_1, K_4, S_{11}, P_1), SN

注. N, K, S, P 分别表示通道数、卷积核大小、步长、填充; IN 表示归一化方式; Leaky-ReLU 表示激活函数.

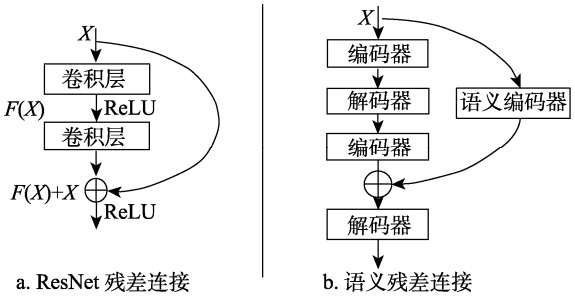


图 2 残差连接和语义残差连接

2.1.2 注意力机制

在图像风格转换任务中，图像域内信息通常可以划分为 2 个维度: 内容和风格. 其中，内容信息偏向于全局信息，而风格信息则侧重于局部信息. 目前，大多数用于图像生成的 GAN 模型使用卷积层进行构建，卷积算子擅长提取图像的局部信息，但在提取全局信息时卷积算子显得捉襟见肘. 因此，本文将 SAGAN[28]中提出的用于 GAN 中的注意力机制模块引入图像风格转换任务，用于增强生成器对图像全局信息的提取的能力，从而提升模型对图像域内信息的建模能力，相关结构如图 3 所示.

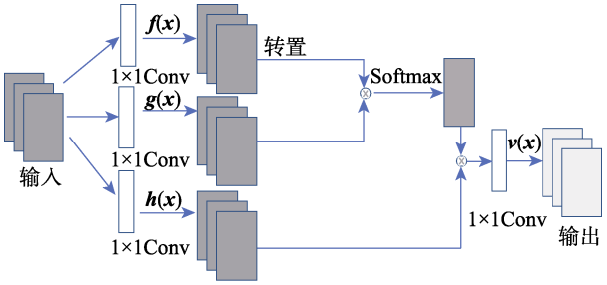


图 3 注意力机制模块

如图 3 所示，本文将图像特征分别进行 3 次 1×1的卷积计算，编码得到 $\boldsymbol{f}(\boldsymbol{x}),\boldsymbol{g}(\boldsymbol{x}),\boldsymbol{h}(\boldsymbol{x})$ 3 个特征图，即注意力机制中的 query, key 和 value. 这 3 个特征图之间的计算公式为

$$\boldsymbol{v}(\boldsymbol{x})=\mathrm{Softmax}\left(\frac{\boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})}{\sqrt{d}}\right)\boldsymbol{h}(\boldsymbol{x}) \tag{1}$$

其中，$\boldsymbol{v}(\boldsymbol{x})$ 表示通过注意力机制输出的特征图；d 表示特征的通道维度. 本文通过引入注意力机制，以增加少量计算量的代价极大地提升了模型对图像域内信息差异建模的能力.

2.2　语义编码器

本文的语义编码器在结构上和生成器中的编码器部分保持一致，但是考虑在训练过程中如果对语义编码器不加以约束，那么可能导致语义编码器的作用减弱，因此，本文引入了一个新的损失函数来对语义编码器进行约束.

2.3　鉴别器

$x\in\left\{X_{\mathrm{t}},G_{\mathrm{s}\to\mathrm{t}}\left(X_{\mathrm{s}}\right)\right\}$ 表示目标域和源域转换到目标域后的图像样本对. 与其他图像风格转换模型一致，在鉴别器的选择上，本文与 CycleGAN 一致，均采用多尺度鉴别器，网络结构如表 2 所示.

2.4　目标损失函数

本文的损失函数分为 4 个部分：对抗损失、循环一致性损失、恒等损失和语义损失.

本文利用对抗损失

$$L_{\mathrm{adv}}^{\mathrm{s}\to\mathrm{t}}=E_{x\sim X_{\mathrm{s}}}\left[\left(D_{\mathrm{t}}(x)\right)^2\right]+E_{x\sim X_{\mathrm{s}}}\left[1-\left(D_{\mathrm{t}}\left(G_{\mathrm{s}\to\mathrm{t}}(x)\right)\right)^2\right] \tag{2}$$

$$L_{\mathrm{adv}}^{\mathrm{t}\to\mathrm{s}}=E_{x\sim X_{\mathrm{t}}}\left[\left(D_{\mathrm{s}}(x)\right)^2\right]+E_{x\sim X_{\mathrm{t}}}\left[1-\left(D_{\mathrm{s}}\left(G_{\mathrm{t}\to\mathrm{s}}(x)\right)\right)^2\right] \tag{3}$$

将生成图像的数据分布匹配到目标图像的数据分布，其中 X 表示输入图像. 为了稳定模型的训练过程，相较于已有算法，本文采用最小平方损失函数.

为了缓解模式崩塌问题，本文使用了循环一致性损失

$$L_{\mathrm{cycle}}^{\mathrm{t}\to\mathrm{s}}=E_{\mathrm{x}\sim X_{\mathrm{s}}}\left[\left|x-G_{\mathrm{t}\to\mathrm{s}}\left(G_{\mathrm{s}\to\mathrm{t}}(x)\right)\right|\right]_1 \tag{4}$$

$$L_{\mathrm{cycle}}^{\mathrm{s}\to\mathrm{t}}=E_{\mathrm{x}\sim X_{\mathrm{t}}}\left[\left|x-G_{\mathrm{s}\to\mathrm{t}}\left(G_{\mathrm{t}\to\mathrm{s}}(x)\right)\right|\right]_1 \tag{5}$$

对生成器进行约束. 给定一个图像 $X\in X_{\mathrm{t}}$，图像 X 在经过循环一致性之后应该被转换回原来的图像域.

为了保证输入图像和输出图像的颜色分布相似，本文对生成器应用了身份一致性约束.

$$L_{\mathrm{identity}}^{\mathrm{s}\to\mathrm{t}}=E_{\mathrm{x}\sim X_{\mathrm{t}}}\left[\left|x-G_{\mathrm{s}\to\mathrm{t}}(x)\right|\right]_1 \tag{6}$$

$$L_{\mathrm{identity}}^{\mathrm{t}\to\mathrm{s}}=E_{\mathrm{x}\sim X_{\mathrm{s}}}\left[\left|x-G_{\mathrm{s}\to\mathrm{t}}(x)\right|\right]_1 \tag{7}$$

给定一个图像 $X\in X_{\mathrm{t}}$，图像 X 在经过生成器 $G_{\mathrm{s}\to\mathrm{t}}$ 转换之后不发生改变.

为了保证语义编码器能够提取到图像的语义特征，本文对语义编码器应用了语义损失

$$L_{\mathrm{semantic}}=-E\left[\left|E_{\mathrm{semantic}}\left(x_{\mathrm{t}}\right)-E_{\mathrm{semantic}}\left(x_{\mathrm{s}}\right)\right|\right]_1 \tag{8}$$

进行约束. 其中，E_{semantic} 表示语义编码器，通过该损失函数将语义编码器提取到的不同图像域的特征距离尽可能拉远，使语义编码器能够提取到 2 种不同图像域的语义特征.

本文算法的总体损失函数为

$$L_{\mathrm{total}}=\lambda_1 L_{\mathrm{adv}}+\lambda_2 L_{\mathrm{cycle}}+\lambda_3 L_{\mathrm{identity}}+\lambda_4 L_{\mathrm{semantic}} \tag{9}$$

其中，$\lambda_1=1,\ \lambda_2=10,\ \lambda_3=10,\ \lambda_4=1$.

3　实验结果与分析

3.1　训练设置

3.1.1　数据集

本文算法在 vangh2photo 和 selfie2anime 数据集上进行实验，vangh2photo 包含 800 幅梵高风格的图像和 6932 幅现实场景的图像. 其中，训练集包含 400 幅梵高风格的图像和 6281 幅现实场景的图像，测试集包含 400 幅梵高风格的图像和 751 幅现实场景的图像; selfie2anime 包含 3500 幅动漫风格的人像图像和 3500 幅自拍人像图像，其中训练集和测试集各占一半.

3.1.2　训练方式和超参数设置

本文采用 Adam 优化器进行网络参数优化，其中 $\beta_1=0.5,\beta_2=0.999$. 本文采用以 0.5 的概率将图像水平反转，将图像大小调整为 286×286 像素，随机裁剪到 256×256 像素. 本文所有实验的批量大小均设置为 1. 本文以固定的学习率 0.0001 训练所有模型，直至迭代 50 万次，随后线性衰减，总计训练迭代 100 万次. 同时，本文采用权重系数为 0.0001 的权重衰减，权重初始化采用的方式为均值为 0、标准差为 0.02 的正态分布.

3.1.3　实验环境

本文实验的硬件环境为 Intel Xeon(R) silver 4214R 2.40 GHz CPU; Tesla V100 32 GB GPU; 软件环境为 Centos 操作系统，深度学习框架采用 Pytorch 1.6.0 版本, CUDA 版本为 10.2 .

3.2　实验分析

3.2.1　实验设置与对比算法

本文采用 2 种评价指标(FID, KID)，从 2 种不同的角度分别与 CycleGAN, MUNIT, DRIT++, CouncilGAN, U-GAT-IT 进行比较. 同时为验证本文算法的有效性，实验与上述 5 种当前主流的图像风格转换算法在生成图像可视化效果和 FID, KID

评测指标上分别进行对比.

3.2.2 定性评估

图 4 展示了在 selfie2anime 数据集上 SSC-GAN 与上述对比算法的风格转换结果比较. SSC-GAN 在保留原图像特征的同时, 在所有对比算法中取得了最好的视觉感知质量. CouncilGAN 能较好地实现完成人脸到动漫头像的转换, 但是在转换过程中发色发生了改变. CycleGAN 在转换过程中人脸产生了较大的形变, 且生成的图像真实性欠佳. MUNIT 生成的图像同时存在 CouncilGAN 和 CycleGAN 的缺点. DRIT++生成的图像虽然视觉效果不错, 但是其前景和背景产生了严重的混淆. U-GAT-IT 在图像转换过程中人脸的五官不够突出且不够规则. SSC-GAN 能够同时关注到人脸的发色、五官、前景和背景, 能够处理好图像生成过程中人脸发生的形变, 证明了注意力机制的引入提升了模型对全局特征提取的能力. 此外, 为说明 SSC-GAN 的泛化能力, 本文对比了在 photo2vangh 数据集上的风格转换表现, 如图 5 所示, 在该数据集上本文算法也取得了极佳的视觉效果.

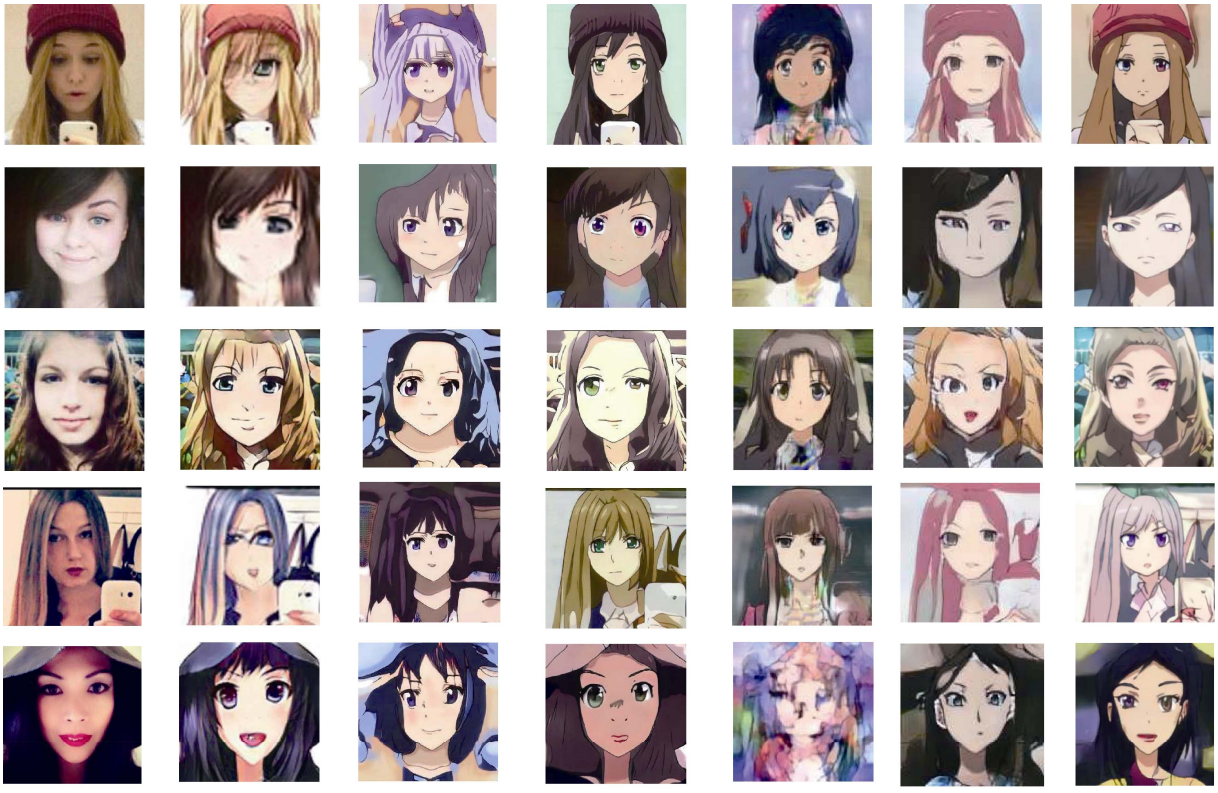
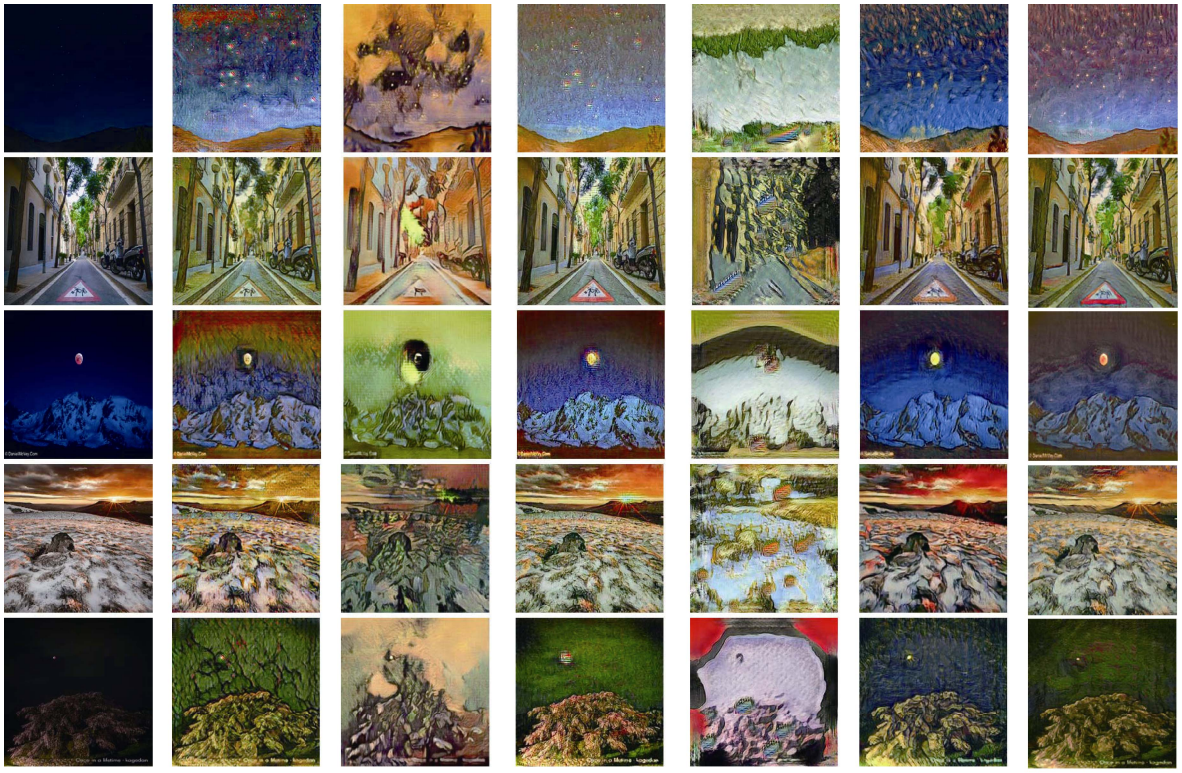


图 4 不同算法在 selfie2anime 数据集上的定性结果比较

3.2.3 定量评估

本文实验采用目前在图像风格转换中广泛使用弗雷切特空间距离(Fréchet inception distance, FID)和核空间距离(kernel inception distance, KID) 2 个指标来对比本文选取的基准模型的定量风格转换表现. FID 衡量生成图像与真实图像间特征向量的距离, FID 值越小说明生成图片多样性越丰富, 图像质量越高. KID 衡量真实图像特征和生成图像特征之间的最大平均差异, 值越小表示生成图像与原图像相似度越高, 其中的特征表示是从 Inception[32]网络中提取的. 与 FID 指标相比, 无偏估计使 KID 指标更加可靠. 除了 photo2vangh 的转换, SSC-GAN 在 FID 和 KID 指标上均取得了最佳表现, 结果如表 3 和表 4 所示. 由表 3 可以看出, SSC-GAN 在 FID 指标表现最佳, 其在 vangh2photo 上明显优于 photo2vangh, 充分说明了 SSC-GAN 能够有效地对图像域间语义信息进行建模. 由表 4 可以看出, 在 KID 指标上, SSC-GAN 在转换中均取得了最佳的 KID 数值表现, 虽然动漫头像和人脸之间存在较大的形变, 但是 SSC-GAN 依旧取得了最佳的转换效果, 进一步证明了 SSC-GAN 的有效性.



a. 输入图像　b. CycleGAN[21]　c. DRIT++[22]　d. CouncilGAN[23]　e. MUNIT[24]　f. U-GAT-IT[25]　g. SSC-GAN

图 5　不同算法在 photo2vangh 数据集上的定性结果比较

表 3　6 种算法在 3 个数据集上的 FID 评分比较

算法	数据集		
	vangh2photo	photo2vangh	selfie2anime
CycleGAN[21]	163.4	151.4	149.4
DRIT++[22]			109.2
CouncilGAN[23]			115.1
MUNIT[24]			131.7
U-GAT-IT[25]	80.4	**89.4**	114.1
SSC-GAN	**75.1**	92.8	**112.2**

注. 加粗表示最优的数值结果.

表 4　6 种算法在 4 个数据集上的 KID 评分比较

算法	数据集			
	vangh2 photo	photo2 vangh	selfie2 anime	anime2 selfie
CycleGAN[21]	4.68	5.46	13.08	11.84
DRIT++[22]	7.72	12.65	15.08	14.85
CouncilGAN[23]			15.21	14.94
MUNIT[24]	9.53	13.08	13.85	13.94
U-GAT-IT[25]	5.61	4.28	11.61	11.52
SSC-GAN	**3.67**	**3.26**	**10.30**	**11.21**

注. 加粗表示最优的数值结果.

4　结　论

针对现有图像风格转换模型未考虑数据集图像域间语义信息差异大和卷积算子缺乏全局特征提取能力的问题，本文提出了 SSC-GAN. 大量实验结果表明，在图像域间语义信息建模方面，经过 SSC-GAN 风格转换后的图像更加贴近于目标域，同时保留了大部分原图像的细节；在图像域内语义信息建模方面，经过 SSC-GAN 风格转换后的图像能够较好地保留图像的全局特征，不会丢失源图像的细节、混淆前景和背景. 然而，目前本文算法只适用于一对一的图像风格转换任务，如何实现一对多的风格转换是今后需要完善的工作.

参考文献(References):

[1] Guan Xinping, Zhao Lixing, Tang Yinggan. Mixed filter for image denoising[J]. Journal of Image and Graphics: Series A, 2005, 10(3): 332-337(in Chinese)
(关新平, 赵立兴, 唐英干. 图像去噪混合滤波方法[J]. 中国图象图形学报: A 辑, 2005, 10(3): 332-337)

[2] Yang Jie, Zhao Shubin, Wang Qiang. Research on image defogging based on generative adversarial network[J]. Command Control and Simulation, 2022, 44(1): 44-50(in Chinese)
(杨杰, 赵书斌, 王强. 基于生成对抗网络的图像去雾研究[J]. 指挥控制与仿真, 2022, 44(1): 44-50)

[3] Gao Yuan, Liu Zhi, Qin Pinle, *et al*. Medical image super-resolution algorithm based on deep residual generative adversarial network[J]. Journal of Computer Applications, 2018,

38(9): 2689-2695(in Chinese)
(高媛, 刘志, 秦品乐, 等. 基于深度残差生成对抗网络的医学影像超分辨率算法[J]. 计算机应用, 2018, 38(9): 2689-2695)

[4] Hertzmann A, Jacobs C E, Oliver N, *et al*. Image analogies[C] //Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001: 327-340

[5] LeCun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324

[6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012: 1097-1105

[7] Szegedy C, Liu W, Jia Y Q, *et al*. Going deeper with convolutions[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 1-9

[8] Huang G, Liu Z, van der Maaten L, *et al*. Densely connected convolutional networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2261-2269

[9] Lin M, Chen Q, Yan S. Network in network[OL]. [2022-06-15]. https://arxiv.org/abs/1312.4400v3

[10] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778

[11] Yi Z L, Zhang H, Tan P, *et al*. DualGAN: unsupervised dual learning for image-to-Image translation[OL]. [2022-06-15]. https://arxiv.org/abs/1704.02510v4

[12] 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. Los Alamitos: IEEE Computer Society Press, 2019: 4396-4405

[13] Ohkawa T, Inoue N, Kataoka H, *et al*. Augmented cyclic consistency regularization for unpaired image-to-image translation[C] //Proceedings of the 25th International Conference on Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 362-369

[14] Han J, Shoeiby M, Petersson L, *et al*. Dual contrastive learning for unsupervised image-to-image translation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 746-755

[15] Snell J, Ridgeway K, Liao R, *et al*. Learning to generate images with perceptual similarity metrics[C] //Proceedings of the IEEE International Conference on Image Processing. Los Alamitos: IEEE Computer Society Press, 2017: 4277-4281

[16] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2016: 694-711

[17] Mechrez R, Talmi I, Zelnik-Manor L. The contextual loss for image transformation with non-aligned data[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 768-783

[18] Park T, Efros A A, Zhang R, *et al*. Contrastive learning for unpaired image-to-image translation[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2020: 319-345

[19] Zheng C, Cham T J, Cai J. The spatially-correlative loss for various image translation tasks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 16407-16417

[20] Isola P, Zhu J Y, Zhou T H, *et al*. Image-to-image translation with conditional adversarial networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5967-5976

[21] Zhu J Y, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2242-2251

[22] Lee H Y, Tseng H Y, Huang J B, *et al*. Diverse image-to-image translation via disentangled representations[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 35-51

[23] Nizan O, Tal A. Breaking the cycle-colleagues are all you need[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 7857-7866

[24] Huang X, Liu M Y, Belongie S, *et al*. Multimodal unsupervised image-to-image translation[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 172-189

[25] Kim J, Kim M, Kang H, *et al*. U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation[OL]. [2022-06-15]. https://arxiv.org/abs/1907.10830

[26] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets[C] //Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. 2014: 2672-2680

[27] Mirza M, Osindero S. Conditional generative adversarial nets[OL]. [2022-06-15]. https://arxiv.org/abs/1411.1784v1

[28] Zhang H, Goodfellow I, Metaxas D, *et al*. Self-attention generative adversarial networks[C] //Proceedings of the 36th International Conference on Machine Learning. Lille: PMLR Press, 2019: 7354-7363

[29] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241

[30] Zhou B L, Khosla A, Lapedriza A, *et al*. Learning deep features for discriminative localization[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 2921-2929

[31] He K M, Zhang X Y, Ren S Q, *et al*. Identity mappings in deep residual networks[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2016: 630-645

[32] Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 2818-2826