

基金项目论文

# 基于多层次判别器的图像翻译模型

向 晴, 袁健华

(北京邮电大学 理学院, 北京 100876)

**摘 要:** 图像翻译是一类重要的视觉问题, 其目标是学习两个视觉域之间的映射, 同时保持原始结构或语义。图像翻译研究领域主要存在两个难点, 一是缺少配对的图片训练集; 二是很难由单个输入产生多种输出。基于解表示的图像翻译模型 (DRIT) 虽然同时解决这两个难点, 但输出的结果显示仍然存在生成图像质量不佳的问题。造成输出质量不佳的一个原因是由于从判别器到生成器的梯度反向传播过程较长, 使得训练初期生成器的梯度所包含信息量较少。在 DRIT 模型研究基础上, 本文提出了一种基于多层次判别的多目标图像翻译模型, 一定程度上解决了生成图像质量不佳的问题。我们的数值实验结果也表明了本文提出的多层次判别结构对输出质量的提高具有一定的帮助。

**关键词:** 对抗生成网络; 图像翻译; 梯度信息分析; 多层次判别器

**中图分类号:** TP391.41 **文献标识码:** A **DOI:** 10.3969/j.issn.1003-6970.2020.03.003

**本文著录格式:** 向晴, 袁健华. 基于多层次判别器的图像翻译模型[J]. 软件, 2020, 41 (03): 11-17

## Image-to-image Translation Based on Multi-layer Discriminator

XIANG Qing, YUAN Jian-hua

(School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**【Abstract】:** Image-to-image translation is a class of vision problems. Its goal is to learn the mapping between two visual domains, while keeping the original structure or semantics. There are two challenges in this field: one is the lack of aligned training image pairs, and the other one is that it is difficult to generate multiple outputs from a single input. DRIT solved these two difficulties at the same time, but the results show that the quality of the generated images are still unsatisfactory. One of the reasons is that the gradient back-propagation process from the discriminator to the generator is long, so that the gradient of the generator in the initial training period contains less information. In this paper, based on a multi-level discriminator, we propose an image-to-image translation model to solve the problem of poor quality of generated images to a certain extent. Simulation results show that our proposed multi-level discriminator is helpful to improve the quality of the outputs.

**【Key words】:** GAN; Image-to-image translation; Flow of gradients; Multi-level discriminator

## 0 引言

近年来, 随着深度学习技术的发展, 图像翻译 (Image-to-Image Translation) 已成为计算机视觉领域中一个备受关注的研究方向<sup>[1]</sup>。风格迁移<sup>[2]</sup>、图像超分辨<sup>[3]</sup>、图像去噪<sup>[4,5]</sup>、去雾<sup>[6]</sup>、修复<sup>[7]</sup>等等都可被视为图像翻译中的研究问题。

据我们所知, 图像翻译问题的研究工作主要存在着两个挑战。第一个挑战是在图像翻译模型的训练过程缺少配对的图片训练集。实际上, 大部分图

像翻译任务很难获得训练需要的配对数据集, 例如, 我们很难收集到内容完全一致的照片和艺术画作来训练模型。第二个挑战是完成多目标输出的模型的训练比较困难, 研究中单一目标的训练相对容易很多, 但是在现实中, 我们遇到的很多图像翻译问题是多目标的, 如将一张黑白图翻译转换为彩图, 可以有多种上色方式, 求解与评价都存在一定的困难。正由于这两个挑战, 越来越多的研究人员开始研究图像翻译问题。经典的图像翻译模型有很多, 基于不配对数据集训练的模型中最有名的是以生成对抗

基金项目: 国家自然科学基金项目(批准号: 11971073; 11671052)

作者简介: 向晴(1994-), 女, 研究生, 主要研究方向: 图像生成; 袁健华(1979-), 女, 教授, 主要研究方向: 最优化算法及其应用。

网络 (Generative Adversarial Networks, GAN) [8] 为基础的循环一致生成对抗网络 (Cycle-consistent adversarial networks, Cycle-GAN) [9] 模型, 与之类似的还有 DiscoGAN [10] 和 DualGAN [11] 等。虽然这些模型用循环一致损失的方式实现了基于不配对数据集的图像翻译训练, 但它们的缺陷在于无法进行多目标输出。此外, J.-Y. Zhu 等人提出了 Bicycle-GAN [12] 模型, 该模型采用潜码与目标域图像相互映射的方式基本实现了多目标图像输出任务, 问题是这个模型是基于配对的数据集完成训练的, 一定程度上限制它的广泛应用。最近, H.-Y. Lee 等人提出了基于 GAN 的图像翻译模型 (DRIT) [13] 模型, 试图同时解决我们所说的图像翻译问题中的两个挑战, 从他们的研究结果来看, DRIT 模型通过将循环一致损失与潜码映射方式相结合, 实现了基于不配对数据集的多目标图像翻译。然而, 实验模拟的结果表明 DRIT 模型依然存在训练不稳定、生成图像结果模糊等问题, 其输出的图片质量还存在提高的空间。我们的研究工作的出发点就是从 DRIT 模型出发, 期望能够通过分析研究该模型的优缺点, 构建新的图像翻译模型, 提高图片输出质量。

我们在简要分析 DRIT 模型图像翻译结果可能模糊原因的基础上, 提出了一种带有多层次判别器的图像翻译模型, 并且对原始 DRIT 模型和本文提出的模型的输出结果进行了定性分析与定量评估。数值实验结果表明, 我们提出的多层次判别器能够一定程度上提高输出图像的质量。

本文的具体内容如下: 第 1 节介绍 GAN 模型和多目标图像翻译的相关工作。第 2 节详细描述本文对输出图像模糊问题的分析, 并提出带有多层次判别器的图像翻译模型。第 3 节介绍给出数值实验的结果并进行分析。在第 4 节, 我们简单总结了本文的研究结果。

## 1 图像翻译的研究基础

在本节中, 我们介绍生成对抗网络和基于解表示的图像翻译模型 (DRIT) 的基础知识, 并简要分析 DRIT 模型图像翻译结果可能模糊的原因。

### 1.1 生成对抗网络与图像转换

生成对抗网络 (GAN) 是目前深度学习领域的热点之一, 此模型最早在 2014 年由 Ian Goodfellow 等人 [8] 提出。GAN 模型由生成器 G 和判别器 D 构成, 训练过程由这两部分交替对抗地学习完成。G 的输入通常为的一组满足正态分布的随机噪声, 输出为一

张图片; D 的输入为一张图片, 输出为 1 或 0, 表示它判定输入图像为真实图像还是由 G 生成的图像。对抗式学习发生在 G 与 D 之间: G 需要生成尽可能能够欺骗 D 的图像, 而 D 则需要尽可能分辨出图像的真假。这个零和游戏的对抗损失函数定义为。

$$L_{adv}(G, D) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim P_G} [\log(1 - D(G(z)))] \quad (1)$$

具体训练时, 生成器 G 尽量将对抗损失函数最小化, 而判别器 D 尝试将其最大化, 从而尽可能实现提高训练质量的目标。最近的研究结果 [9, 14, 15] 已经证明了 GAN 在生成合成图像问题方面具有较突出的优点, 这也是 GAN 研究工作越来越受到众多研究人员关注的原因。

近几年, GAN 逐渐被人们用于解决各种问题 [8, 9, 13, 16]。在图像翻译方面, Isola 等人 [16] 首先提出了 Pix2pix 模型, 将原始图像作为 GAN 的输入, 训练模型输出目标域图像。Pix2pix 模型在 GAN 对抗损失的基础上加上了均方误差损失, 大大提高了图像翻译的清晰度, 使得翻译后图像的质量有了很大提高。但由于计算损失需要用到均方误差, 所以 Pix2pix 需要配对数据集才可训练。随后, 为了解决基于不配对数据集训练的问题, Cycle-GAN [9] 采用循环一致性来规范训练, 即让  $a \approx G_{B \rightarrow A}(G_{A \rightarrow B}(a))$ 。加上对抗损失, Cycle-GAN 可以基于不配对图片集进行图像翻译模型训练。但是, 这些方法都没有实现多目标输出, 训练完成后单次输入只能产生一个输出。在多目标转换方面, 由 Choi Y 等人 [17] 提出的 StarGAN, 它用标记人物表情类别的方式来限制 GAN 的生成过程。虽然有效, 但这种方法只能生成固定数量的表情类型, 例如数据集中共标记了 5 种表情, 则只能生成 5 种表情。随后, J.-Y. Zhu 等人 [12] 提出的 Bicycle-GAN 算法, 通过约束输出图像风格与潜在噪声向量之间的一对一关系, 来解决图像翻译中多目标输出的问题。图像翻译过程将原图  $x$  和潜在噪声向量  $z$  同时注入生成器 G 生成目标风格图像。此模型中, 由于潜在噪声向量来自高斯分布采样, 所以潜在噪声向量的数量是无穷无尽的, 对应地就可以让生成图像的风格样式也是没有数量限制的。但 Bicycle-GAN 的训练过程需要配对图片集。所以, 在很难收集到配对数据集的图像翻译问题上, Bicycle-GAN 的推广应用受到了限制。

### 1.2 DRIT 模型

基于解表示的 DRIT (Diverse image-to-image translation via disentangled representations) 模型于

2018 年首次被 H.-Y. Lee 等人提出, 首次解决了基于不配对数据集的多目标图像翻译问题。DRIT 模型将 Cycle-GAN 中的循环一致性方法与 Bicycle-GAN 模型中的约束潜在噪声向量方法集于一身, 结合解表示方法, 最终达到了训练目的。具体来说, 就是将图像翻译网络分解为结构编码器、风格编码器与生成器, 用编码器对图像结构和风格进行编码。然后通过重组结构码与风格码, 生成对应风格的图像, 再送入判别器进行判别, 得到  $A \rightarrow B$  与  $B \rightarrow A$  的转换。为使模型可基于不配对数据集进行训练, DRIT 模型将第一次图像翻译得到的图再做一次编码与重组, 得到  $A \rightarrow B \rightarrow A$  与  $B \rightarrow A \rightarrow B$  的转换, 从而可以计算循环一致性损失。在多目标图像翻译任务方面, 约束风格码与潜在噪声向量  $z$  之间的一对一关系。最终, DRIT 模型基于不配对数据集, 完成了多目标图像翻译任务。然而, 这样的排列与重组, 使得网络在训练的过程中梯度流过长, 导致模型编码与生成初期的梯度比较随机, 带有的信息量较少<sup>[18]</sup>。所以本文提出了一种带有多层次判别器的图像翻译模型, 让判别器的梯度流信息直接作用到生成器初期, 使得模型的输出图像质量有了一定提高。

## 2 基于多层次判别器的多目标图像翻译模型

在本节中, 我们介绍本文主要的研究结果。基于对 DRIT 模型梯度流分析后, 我们提出了一种基于多层次判别的多目标图像翻译模型, 有望在一定程度上解决生成图像质量不佳的问题。

### 2.1 模型框架

本文模型与 DRIT 模型类似, 包括编码、生成和判别三部分, 其中编码器包含一系列的卷积操作, 作用是提取出图像的结构码与风格码; 生成器由一系列反卷积组成, 可将结构码与风格码相结合, 还原出图像; 判别器将图像映射到  $[0, 1]$  区间, 作用是判别图像来自真实图像还是由生成器生成。我们以如图 1 所示的自然照片集 (A) 到艺术画作集 (B) 为例来具体说明我们的模型。图 1 中的  $a \in A$  是一张自然风景图片,  $b \in B$  为一张莫奈画作。我们将自然风景图片  $a$  与莫奈画作  $b$  分别进行结构和风格的解码, 得到各自的结构码与风格码。如图 1 所示, 我们再将  $a$  的结构码与  $b$  的风格码相结合, 生成由  $a$  到  $b$  的风格迁移图像。判别器的目标是努力区分真实莫奈艺术品与图像翻译得到的莫奈艺术品。同样由  $B$  到

$A$  也是如此。另外, 为了在测试时生成多样风格的输出, 我们对风格码进行正则化, 以便可以从高斯分布  $N(0,1)$  中随机采样, 并以此作为随机风格码。

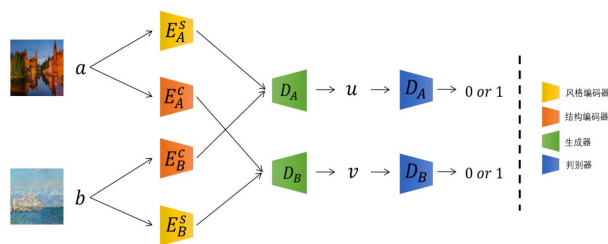


图 1 基于不配对数据集的多目标图像翻译模型框架  
Fig.1 Framework of Image-to-image translation based on multi-layer discriminator

由上述过程, 本文将编码过程表示为。

$$\begin{aligned} c_a &= E_c(a), s_a = E_s(a) \\ c_b &= E_c(b), s_b = E_s(b) \end{aligned} \quad (2)$$

其中,  $c_a$  和  $c_b$  分别代表原始图  $a$  和目标风格图  $b$  的结构码,  $s_b$  代表目标风格图  $b$  的风格码。由此, 基于结构码与风格码的图像生成过程可表示为。

$$\begin{aligned} u &= G(c_b, s_a) \\ v &= G(c_a, s_b) \end{aligned} \quad (3)$$

其中,  $u$  表示由  $b$  图像风格转换到  $A$  类的图,  $v$  表示由  $a$  图像风格转换到  $B$  类的图。最终, 我们将生成的图送入风格判别器, 得到对抗损失。

### 2.2 多层次判别器

在训练过程中, GAN 训练不稳定的原因之一是训练初期当生成器和判别器网络较深时, 从判别器  $D$  到生成器  $G$  的反向传播过程较远, 使得反向求导得到的梯度信息 (红色箭头) 达到生成器初期时带有的信息较少<sup>[18]</sup> (如图 2 (a) 所示 DRIT 模型)。受 MSG-GAN<sup>[18]</sup>与 U-Net<sup>[19]</sup>跳跃连接的启发, 本文采用了一种如图 2 (b) 所示的多层次判别结构, 来重新构造 DRIT 模型中的判别器。主要的做法是, 在判

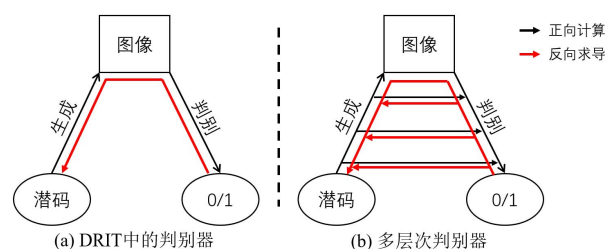


图 2 DRIT 判别器与本文多层次判别器的对比示意图  
Fig.2 Framework of DRIT's discriminator (left) and our Multi-layer discriminator (right)



别器计算的过程中加入生成器每层的输出结果,这样就为生成与判别建立了多层连接。当网络更新时,反向求导操作会使得梯度信息通过新添加的连接层作用到生成器上,使得模型的整个训练过程更加稳定,也使生成的图像在细节上更加精细。

图3展示了多层次判别器的整体结构图。其中图3(a)展示了生成器G的图像生成过程,将生成器G根据特征图的大小分解为一系列由反卷积组成的 $g^*$ 的操作,每个 $g^*$ 包含一个反卷积块。所以生成器G可以写成。

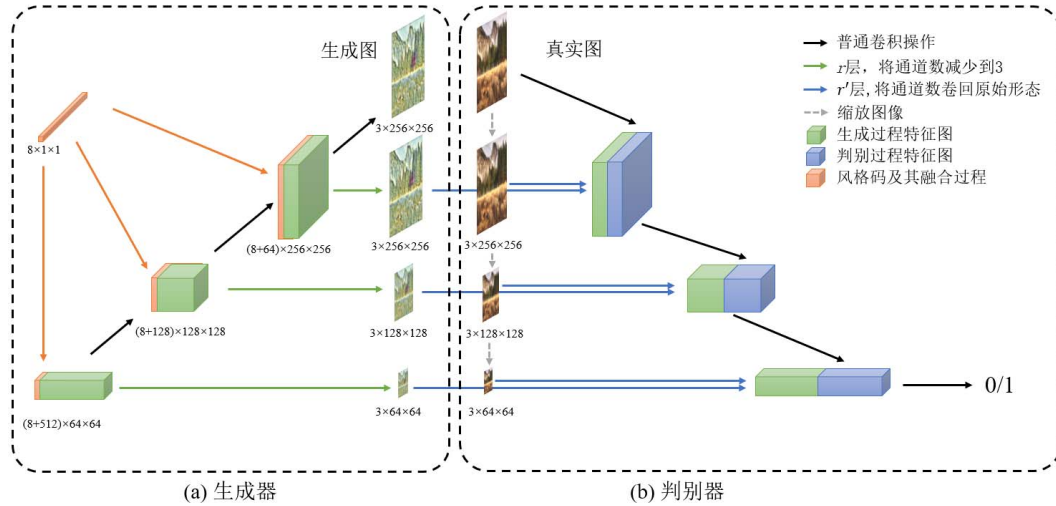


图3 多层次判别器结构图

Fig.3 Detailed framework of our Multi-layer discriminator.

$$G(c, s) = g^k \otimes g^{k-1} \otimes \dots \otimes g^1 \otimes g^0(c, s) \quad (4)$$

其中 $g^i \otimes g^{i-1}$ 操作表示 $g^i(g^{i-1}(\dots))$ ,  $i=1, 2, \dots, k$ ,  $k$ 表示生成器或判别器的层数。定义一系列 $r^*$ 函数,将每一个 $g^*$ 生成的多通道的特征图用 $1 \times 1$ 的卷积处理成3通道的图像。这个过程可表示为。

$$Img_f^i = r^i(g^i(c, s)) \quad (5)$$

同时对真实的目标域图像进行下采样,得到与 $Img_f^i$ 大小完全相同的 $Img_r^i$ 。图3(b)中展示了多层次判别器的判别过程,我们将判别器分解为一系列判别操作 $d^*$ 。与常规判别操作(6)式不同,我们将(5)式得到的 $Img_f^i$ 融合进对应的判别层,得到如下(7)式。

$$D(*) = d^k \otimes d^{k-1} \otimes \dots \otimes d^1 \otimes d^0(*) \quad (6)$$

$$D(*) = d^k(Img_*^k) \otimes d^{k-1}(Img_*^{k-1}) \otimes \dots \otimes d^1(Img_*^1) \otimes d^0(*) \quad (7)$$

其中, $*$ 表示待判别的图像, $d^i(Img^i) \otimes d^{i-1}$ 操作表示 $d^i(Img^i; d^{i-1}(\dots))$ ,  $i=1, 2, \dots, k$ ,  $[\cdot]$ 代表串联(concatenation)操作。由此,可使网络在更新参数的过程中,直接对生成过程的底层特征产生作用,从而加强稳定性。多层次判别的对抗损失 $L_{adv}^s$ 可表

示为。

$$L_{adv}^s(G, D) = \mathbb{E}_a[\log D(a)] + \mathbb{E}_a[\log(1 - D(u))] + \mathbb{E}_b[\log D(b)] + \mathbb{E}_b[\log(1 - D(v))] \quad (8)$$

### 2.3 其他损失函数

#### (1) 结构判别对抗损失函数

为了实现内容与风格的分离,使得结构码只带有结构信息而不带风格语义,本文借鉴了DRIT<sup>[13]</sup>模型的结构判别器 $D_c$ 来强制将来自A与B类图像的结构特征编码到同一空间。训练过程中,结构判别器 $D_c$ 会尽量区分结构码来自A还是B,相对的,结构编码器为了能“骗过”结构判别器,会努力编码出不带有各自风格特征的结构码。这部分的对抗损失可表示为。

$$L_{adv}^c(E_c, D_c) = \mathbb{E}_a[\log D_c(c_a)] + \mathbb{E}_b[\log(1 - D_c(c_b))] + \mathbb{E}_a[\log D_c(c_a)] + \mathbb{E}_b[\log(1 - D_c(c_b))] \quad (9)$$

#### (2) 自一致性损失函数

理论上,当编码与生成过程达到完美时, $G(c_a, s_a)$ ,  $G(c_b, s_b)$ 应分别与原图 $a, b$ 完全一致,所以首先构建自一致性损失。

$$L_{rec} = \mathbb{E}_a \|a - G(c_a, s_a)\|_1 + \mathbb{E}_b \|b - G(c_b, s_b)\|_1 \quad (10)$$

#### (3) KL 散度损失

为了在测试时可用随机抽样来决定生成图像的风格，使得风格码尽可能接近高斯分布，我们采用了 KL 散度损失。

$$L_{KL} = \mathbb{E}[D_{KL}(s_b \| N(0,1))] \quad (11)$$

其中， $D_{KL}(p \| q) = -\int p(z) \log \frac{p(z)}{q(z)} dz$ 。

综合 (8) ~ (11) 式，本文最终的目标函数可表示为。

$$L = \lambda_1 L_{adv}^s(G, D) + \lambda_2 L_{adv}^c + \lambda_3 L_{rec} + \lambda_4 L_{KL} \quad (12)$$

其中，每项损失的比重由超参数  $\lambda_*$  决定。

### 3 实验结果及分析

为展示本文提出的模型在图像翻译效果上的优势，我们用相同的数据集训练 DRIT 模型和我们提出的基于多层次判别的多目标图像翻译模型。本节

中，我们将给出对比实验的结果并进行定性和定量分析。

#### 3.1 实验设置

我们的数值实验采用的数据集为公开不配对数据集 monet2photo。其中，训练集中普通照片 6287 张，莫奈艺术作品 1072 张，测试照片 751 张，训练前均处理为 256×256 像素大小。

#### 3.2 定性实验结果对比

图 4 展示了本文基于多层次判别的多目标图像翻译模型的生成结果。我们将风景照片（第一列）的结构码提取出来，与莫奈艺术品（第一行）的风格码相结合，得到了各种不同的莫奈艺术品。从图上可以看到，自然照片的纹理与色彩都产生了相应的转变，并与目标风格区域一致，说明我们的模型能够将照片很好地转换成莫奈艺术风格，实现多目标图像翻译的目的。



图 4 本文图像翻译模型的结果展示。每张图由其左端的图像结构码与上方的图像风格码生成  
Fig.4 The results of the image-to-image translation in this paper. Each image is generated by the structure code of the left image and the style code of the top image

图 5 展示了 DRIT 模型结果图和本文基于多层次判别的多目标图像翻译模型的结果对比。在结果图上，我们用框体标注了特别对比的部分。这些特殊部分的对比结果表明 DRIT 模型在对图像翻译后，有

可能会出现网格状粗糙颗粒，而本文模型则一定程度上缓解了出现网格状粗糙颗粒的情况，使转换后的图像在一定程度上细腻柔和很多。模拟实验结果也说明多层次判别器确实可以起到稳定训练的作用。



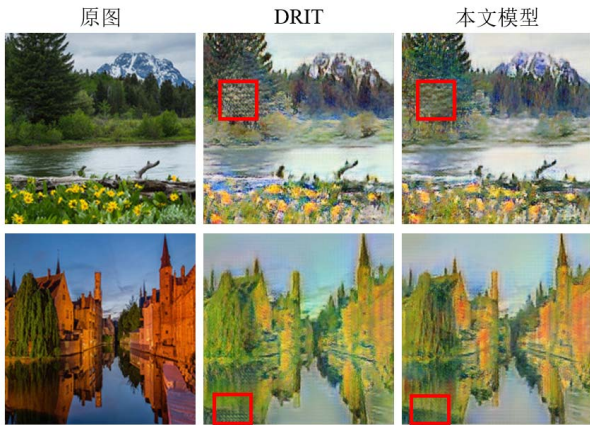


图5 DRIT模型与本文模型的定性结果对比  
Fig.5 Comparison of qualitative results between the DRIT and our model

### 3.3 定量实验结果对比

为测试翻译后图像的质量与多样性,本文采用了与DRIT模型原文相同的两个测试指标FID<sup>[20]</sup>和LPIPS<sup>[21]</sup>来对图像进行评估。

FID是一种常用的生成图像质量评估标准<sup>[20]</sup>。它可以测量生成图像分布与真实图像分布之间的某种距离。其原理是,以训练好的GoogLeNet-v3分类模型为模型基础,输入待测图像,取网络倒数第二层的输出(一个2048维的向量)作为该图像的特征。计算出所有图片的特征后,用如下公式计算生成图像与真实图像两分布间的距离。

$$FID = \|m - m_r\|_2^2 \text{Tr} \left( C + C_r - 2(CC_r)^{\frac{1}{2}} \right) \quad (13)$$

其中, $m$ 和 $C$ 分别是生成图像特征的均值和协方差, $m_r$ 和 $C_r$ 分别是真实图像特征的均值和协方差。当生成图像集和真实图像集的风格越接近,FID值越小,说明生成图像质量越高。本文分别计算了DRIT $\leftrightarrow$ 真实莫奈艺术品,和本文模型 $\leftrightarrow$ 真实莫奈艺术品的FID值(如表1所示)。可以看出,本文模型FID指标显然低于DRIT,说明本文模型生成的图片更接近于真实的莫奈艺术品。

表1 FID与LPIPS测试指标。

Tab.1 The scores of FID and LPIPS metrics

	FID	LPIPS
DRIT	81.2573	0.523
本文模型	74.4013	0.527

LPIPS指标是一种多样性评估指标,已被证明与人类感知相似度密切相关<sup>[21]</sup>。与<sup>[12]</sup>一致地,本文

在预先训练的AlexNet<sup>[22]</sup>的深层特征空间中,计算了每张原始风景照与生成的莫奈风格图像之间的LPIPS距离<sup>[21]</sup>,然后取所有测试值的平均值,作为最终LPIPS多样性指标(如表1所示)。越高的LPIPS值表示风格迁移的越彻底,生成的图像更具有多样性。由表1所示,本文模型的LPIPS值略高于DRIT模型的值,说明本文模型生成的图像多样性不输于DRIT模型。

## 4 总结

为缓解基于不配对数据集的多目标图像翻译的模型不稳定问题,本文构建了一个带有多层次判别器的生成网络,来进行图像到图像的转换生成,由此构造了基于多层次判别的多目标图像翻译模型。应用DRIT模型和我们提出的基于不配对数据集的多目标图像翻译的模型,我们进行了图像翻译的数值实验,从定性与定量结果来看,这种简单的方法提高了生成图像的质量,使得多尺度图像生成有了一定的提高。当然,图像生成质量还有提升空间,在以后的工作中我们将继续改进这个模型。

## 参考文献

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [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] 尹宗天, 谢超逸, 刘苏宜, 等. 低分辨率图像的细节还原[J]. 软件, 2018, 39(5): 199-202.
- [4] 曹妍, 陈伟, 徐森. 图像去噪方法研究与仿真[J]. 软件, 2015, 36(4): 33-36.
- [5] 刘鑫森, 康朝红, 薛乐乐. 基于ICA的遥感图像去噪融合研究[J]. 软件, 2015, 36(7): 53-56.
- [6] 段世杰, 黄华, 王鹏飞, 等. 一种单帧图像的快速去雾方法[J]. 软件, 2015, 36(5): 1-6.
- [7] 胡云康, 姜苏, 吴志荣, 等. 基于改进的纹理合成图像修复算法[J]. 软件, 2016, 37(4): 60-63.
- [8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
- [9] 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. 2017: 2223-2232.
- [10] Kim T, Cha M, Kim H, et al. Learning to discover cross-

- domain relations with generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 1857-1865.
- [11] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks[C]//Advances in neural information processing systems. 2017: 700-708.
- [12] Zhu J Y, Zhang R, Pathak D, et al. Toward multimodal image-to-image translation[C]//Advances in Neural Information Processing Systems. 2017: 465-476.
- [13] 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 (ECCV). 2018: 35-51.
- [14] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 172-189.
- [15] Taigman Y, Polyak A, Wolf L. Unsupervised cross-domain image generation[J]. arXiv preprint arXiv: 1611.02200, 2016.
- [16] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [17] Choi Y, Choi M, Kim M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8789-8797.
- [18] Karnewar A, Wang O, Iyengar R S. MSG-GAN: Multi-Scale Gradient GAN for Stable Image Synthesis[J]. CoRR, 2019.
- [19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
- [20] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[C]//Advances in Neural Information Processing Systems. 2017: 6626-6637.
- [21] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 586-595.
- [22] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

(上接第 10 页)

- [8] K. Wisaeng, W. Sa-Ngiamvibool. Exudates Detection Using Morphology Mean Shift Algorithm in Retinal Images[J]. IEEE Access, 2019, 1(3): 11946-11958.
- [9] 陈园, 侯赞, 刘军华, 等. 基于改进K-Means聚类医学图像配准[J]. 软件, 2018, 39(01): 75-82.
- [10] Parham Khojasteh, Leandro Aparecido Passos Júnior, et al. Exudate detection in fundus images using deeply-learnable features[J]. Computers in Biology and Medicine, 2019, 104: 62-69.
- [11] Srivastava, R, Duan, L, Wong, D.W.K., et al. Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels[J]. Computer Methods and Programs in Biomedicine, 2017, 138(): 83-91.
- [12] Prenta I P, Lon Ari S. Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion[J]. Computer Methods and Programs in Biomedicine, 2016: S0169260716308033.
- [13] Z. Feng, J. Yang, L. Yao, Y. Qiao, Q. Yu, and X. Xu. Deep Retinal Image Segmentation: A FCN-Based Architecture with Short and Long Skip Connections for Retinal Image Segmentation[C]. International Conference on Neural Information Processing, 2017, PP. 713-722.
- [14] Tan J H, Fujita H, Sivaprasad S, et al. Automated Segmentation of Exudates, Haemorrhages, Microaneurysms using Single Convolutional Neural Network[J]. Information Sciences, 2017: S0020025517308927.
- [15] 曹妍, 陈伟, 徐森. 图像去噪方法研究与仿真[J]. 软件, 2015, 36(4): 33-36
- [16] S6nchez CI, Hornero R, L6pez MI, et al. A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis[J]. Medical engineering and physics, 2008, 30(3): 350-357.
- [17] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. 2015.
- [18] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4): 640-651.
- [19] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. CVPR. IEEE Computer Society, 2016.
- [20] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]. CVPR. IEEE Computer Society, 2017.
- [21] [Diptoneel K, Sreeparna B. A New Dynamic Thresholding Based Technique for Detection of Hard Exudates in Digital Retinal Fundus Image[C]. International Conference on Signal Processing and Integrated Networks (SPIN), 2014, Noida, India.