天津科技大学学报
Journal of Tianjin University of Science & Technology
ISSN 1672-6510,CN 12-1355/N

《天津科技大学学报》网络首发论文

题目：基于 StarGAN 的多属性风格图像生成的轻量化网络
作者：孙志伟，曾令贤，马永军
DOI：10.13364/j.issn.1672-6510.20230048
收稿日期：2023-03-07
网络首发日期：2023-09-28
引用格式：孙志伟，曾令贤，马永军．基于 StarGAN 的多属性风格图像生成的轻量化网络[J/OL]．天津科技大学学报．
https://doi.org/10.13364/j.issn.1672-6510.20230048

DOI:10.13364/j.issn.1672-6510.20230048

基于 StarGAN 的多属性风格图像生成的轻量化网络

孙志伟，曾令贤，马永军
(天津科技大学人工智能学院，天津 300457)

摘　要：生成对抗网络已广泛用于图像到图像的翻译任务，其中多属性变换得到了越来越多的研究和应用，目前的网络架构参数多而且模型复杂，需要较高的计算能力和存储成本；网络压缩技术如蒸馏和剪枝，主要侧重于视觉识别任务，很少实现对生成任务的压缩。本文提出了一种利用 StarGAN 的低级和高级特征训练参数较少的学生网络 stuStarGAN 的方法，首先采用知识蒸馏对生成器进行蒸馏，并设计学生判别器让教师判别器蒸馏学生判别器；然后在学生网络设计中采用 skip-connection 进行跨模块的特征融合；接着增加内容损失函数保持生成图像和原图像的内容信息的一致性；最后采用深度可分离卷积进一步降低参数量并提高图片生成质量。在 CelebA 和 Fer2013 数据集上的实验结果表明：模型能够在保证生成质量不降低的情况下，用较少参数生成多属性风格的图像，可以方便移植在多种应用场景。

关键词：生成对抗网络；知识蒸馏；skip-connection；深度可分离卷积；内容损失
中图分类号：TP391.9　　**文献标志码**：A　　**文章编号**：1672-6510（0000）00-0000-00

Lightweight Network for Multi-Attribute Style Image Generation Based on StarGAN

SUN Zhiwei, ZENG Lingxian, MA Yongjun
(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: The generation countermeasure network has been widely used in image-to-image translation tasks, in which multi-attribute transformation has been studied and applied increasingly. The existing network architecture has many parameters and complex models, requiring high computing and storage costs; the traditional network compression technology mainly focuses on visual recognition tasks, and rarely implements the compression of generated tasks. This paper proposes a method stuStarGAN to train the student network with fewer parameters by learning the low-level and high-level features of StarGAN. First, distill the generator with knowledge distillation, and design the student discriminator so that the teacher discriminator distills the student discriminator; then in the student network design, skip-connection is used to provide cross module feature fusion; secondly, the content loss function is added to keep the consistency of the content information between the generated image and the original image; finally, depth separable convolution is used to further reduce the number of parameters and improve the quality of image generation. The experimental results on benchmark datasets show that the model can generate multi-attribute style images with fewer parameters without reducing the generation quality, making it easy to transplant to various application scenarios.

Key words: GAN; knowledge distillation; skip-connection; depth separable convolution; content loss

深度生成视觉是计算机视觉领域的一个重要研究方向，将人工生成的过程转化为智能生成的过程，以大幅减少重复性的人工劳动，甚至可以进行创造性的智能创作[1]。生成对抗网络

收稿日期：2023-03-07；修回日期：2023-05-25
基金项目：国家自然科学基金资助项目（61976156）；天津自然科学基金资助项目（18JCQNJC69500）
作者简介：孙志伟（1973—），男，河北保定人，副教授，zhwsun@tust.edu.cn

（generative adversarial network，GAN）自 2014 年由 Goodfellow 等[2]提出，是实现计算机深度生成视觉的主要技术之一，随后发展并衍生了许多变体，如生成器和判别器上都增加约束条件的条件生成对抗网络 CGAN[3]；生成器和判别器均采用深度卷积的 DCGAN 模型[4]；为解决生成对抗网络长期以来的训练不稳定和模式坍塌的问题，Arjovsky 等[5]提出了 WGAN 模型，使用 W（Wasserstein）距离代替 JS（Jensen-Shannon）散度计算生成样本分布与真实样本分布间的距离。但是 WGAN 模型依然存在训练困难，收敛速度慢的问题。Gulrajani 等[6]提出的 WGAN-GP 直接将判别器的梯度作为正则项加入到判别器的损失函数中。对 GAN 结构的改变还包括 Zhang 等[7]将自注意力模块与 GAN 的思想相结合提出的 SAGAN 模型，为图像生成任务提供了注意力驱动的长距离依赖的模型。

以上 GAN 的衍生模型都是通过网络结构、损失函数等的改变，提高 GAN 的性能和稳定性。这些 GAN 的衍生模型被用在图像任务中，如用于解决图像修复的超分辨率重建 SRGAN[8]、IIZUKA[9]等方法。

随着图像到图像翻译任务的发展，对 GAN 所生成图像的要求也越来越多，如两个图像领域的转换问题（cross-domain）、属性编辑问题等，出现了能进行多领域图像转换的 CycleGAN[10]、DualGAN[11]、DualStyleGAN[12]等方法。CycleGAN 主要是解决 pix2pix[13]中进行风格转换时需要成对数据的问题，使用两个生成器和两个判别器分别处理源域到目标域的转换，并且提出了循环一致性损失进行控制，但仍然存在生成器数量和域数量一对一的问题。DualGAN 主要受自然语言翻译任务中对偶学习的启发，使图像翻译器能够在两个无标签的图像域中学习，但是也存在生成器过多的问题。在单属性编辑任务中已有的模型不能很好地完成多属性的转换，往往在 k 个属性之间相互转换时，需要 $k*(k-1)$个生成器，并且由于是一对一的属性变换并不能有效学习到全局特征以及充分利用全部训练数据[14]，多属性风格变换有助于拓展属性变换任务需求。Choi 等[14]提出的 StarGAN 解决了 1 个生成器只能处理单一属性的问题，生成器的形状像星星一样，可以根据不同的输入属性要求产生不同的输出，在人脸数据集上取得了很好的效果。

StarGAN v2[16]是基于 StarGAN[14]的跨域的多样性图像生成网络，其多样性在于通过最大化两个风格编码所生成的图像的距离控制生成图像的多样性，但是不同于多属性生成，模型能够生成某一个域多样性的图片，而不是具体的多属性转换。

多属性图像生成网络在很多场景下具有重要的应用价值，然而该模型结构复杂、计算量大。轻量化旨在保持模型精度基础上减少模型参数量和复杂度，轻量化网络即包含了对网络结构的探索，又有知识蒸馏、剪枝等模型压缩技术的应用，推动了深度学习在移动端和嵌入式端的应用落地，在智能家居、安防、自动驾驶等领域都有重要贡献。传统的模型压缩方法很难对生成模型进行压缩，主要原因包括：生成器需要大量的参数建立潜在向量到生成图像的映射关系，这种极度复杂的映射结构相较于图像识别任务更难确定冗余的权重；目标检测和图像分割等其他视觉任务都是有标签的训练数据，而 GAN 中的很多任务并没有任何标签用来评判生成的图像，如超分辨率重建和风格迁移。

为了解决上述问题，Aguinaldo 等[17]提出一种压缩和加速 GAN 训练的网络框架，利用知识蒸馏技术以 MSE 损失最小化学生网络和教师网络的距离，但是该方法仅能应用于噪声到图像的网络架构，而如今 GAN 的应用主要是图像到图像[17]。为了解决这些问题，Chen 等[18]以 CycleGAN[10]为基准提出了一个新的基于知识蒸馏的小型 GAN 的框架，在像素层面上最小化学生网络和教师网络生成图像的距离，教师网络生成的图像对学生判别器而言是真实样本，因此设计了学生判别器。但是，该方法只能进行单一图像域的转换，而不能进行多属性的图像生成任务。

由于实际场景中实际采集样本的各属性分布不均，多属性生成是目前的研究重点之一，然而现有的模型较为复杂，计算量大，而且图片生成的效果需进一步提高，因此本文提出了一种基于 StarGAN 的可进行多属性风格图像生成的轻量化网络。

1 相关工作

1.1 StarGAN

本文以 StarGAN 为基准模型设计的多属性风格图像生成的轻量化网络。StarGAN 作为跨多领

域的图像到图像翻译任务的生成对抗网络，其结构为 1 个可以生成多属性的条件生成器和 1 个判别器。生成器包括下采样模块、特征提取模块和上采样模块，生成器接受原图像以及目标属性条件作为输入，生成同样尺寸的目标属性图像。判别器接受生成器生成的图像或者真实图像作为输入，但是判别器有两个输出，一个是二分类的输出，判断图像是来自真实样本还是生成器生成样本；另一个是类别输出，判别图像是哪一个属性类别。

1.2 知识蒸馏

Hinton 等[19]提出知识蒸馏用于模型的轻量化过程，主要是设计学生网络，让小型的学生网络学习大型教师网络的低层特征和高层语义信息。知识蒸馏及其变种主要研究教师网络向学生网络传递知识的链接方式，最初的蒸馏对象是 logit 层，让学生网络和教师网络的 logit KL 散度尽可能小。FitNets[20]出现开始蒸馏中间层，一般使用均方误差（MSE）损失函数使学生网络和教师网络特征图尽可能接近，如图 1 所示。Zagoruyko 等[21]提出的 Attention Transfer 进一步发展了 FitNets，提出使用注意力图引导知识的传递。Tian 等[22]在 FitNet 基础上进一步引入对比学习进行知识迁移。

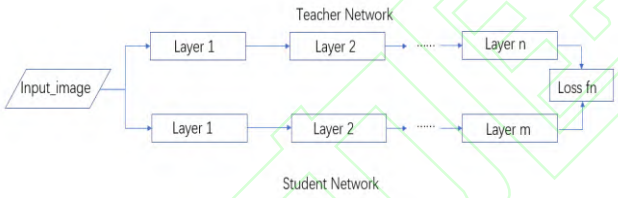

图 1　蒸馏网络示意图

Fig. 1 Diagram of distillation network

这些多数用于 CNN 等神经网络，很少对生成网络 GAN 的蒸馏，这主要在于 GAN 学习的是很复杂的从噪声向量到生成图像的映射关系，而且 GAN 多数是没有标签的数据，导致网络学习到的知识很难衡量，难以确定冗余权重。pix2pix[13]在论文中提供了成对的数据集，Chen 等[18]基于这个成对有标签数据集对 pix2pix 蒸馏，用判别器衡量标签图像、学生生成器以及教师生成器生成的图像三者之间的距离训练学生生成器，并且在 CycleGAN[10]上有较好的效果，主要是其对判别器也同时进行了蒸馏，让学生判别器对教师生成器的输出判定为真，使教师网络和学生网络的判别器接近教师生成器的结果。

1.3 深度可分离卷积

深度可分离卷积是一种广泛应用于卷积神经网络模型结构中的模块，可以取代传统的卷积操作，用于提取图像特征。传统的卷积神经网络，一个卷积核对输入特征图的所有通道进行卷积，卷积核的通道数为输入通道数，个数为输出通道数，而深度可分离卷积将卷积过程进行分解，卷积核个数分别由输入通道数和输出通道数决定。

深度可分离卷积示意图如图 2 所示，核心思想是将卷积分成了逐通道卷积（depthwise convolution）和逐点卷积（pointwise convolution），前者是对输入特征图的每一个通道进行卷积，卷积核个数等于输入通道数；后者主要指 1×1 的卷积[23]，具有不改变特征图尺寸的情况下加深特征图的通道数，能够进行跨通道的特征融合，卷积核个数等于输出通道数。

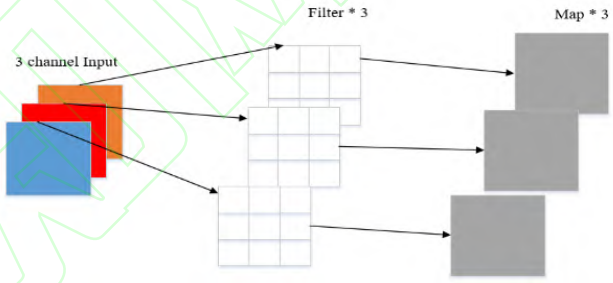

图 2　深度可分离卷积示意图

Fig. 2 Diagram of depth separable convolution

1.4 内容损失函数

在图像到图像转换的早期任务中，如将一张图形的风格转换到另一张内容图像上的风格迁移。内容损失和风格损失比较示意图如图 3 所示。为了确保迁移图像的风格和风格损失函数控制，而使生成的图像和原内容图像的结构等信息不变，则用内容损失函数控制。在 Yang 等[24]提出的 L2M-GAN 模型中，内容损失函数在潜在空间中对人脸的语义信息获取有提升作用。

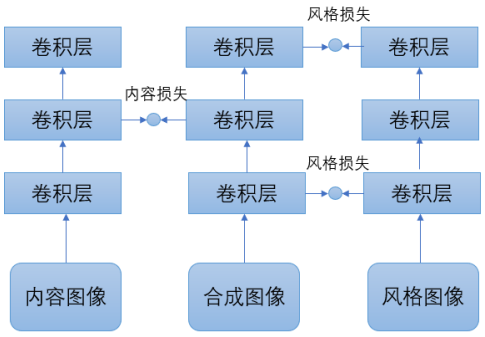

图 3　内容损失和风格损失比较示意图

Fig.3 Diagram of content loss and style loss

2 本文方法

本文是对 StarGAN 进行轻量化，提出了能够生成多风格属性图像的学生网络（student network based on StarGAN，stuStarGAN），在保证生成图片质量的前提下，减少网络的参数量，降低了模型的复杂度。

由于跨域的生成模型，StarGAN v2[16]等只能生成属于该图像域的整体变换，主要建立两个数据集之间的映射关系，而不能生成确定属性和多属性风格图像，并且其中的多样性损失与蒸馏损失发生冲突，故本文以 StarGAN 为教师网络和基准模型。

模型包括以下过程：首先使用知识蒸馏技术降低参数量，提出进一步采用学生判别器蒸馏损失提升性能；然后为了保证生成图像质量，采用 skip-connection 提供跨模块的连接；使用内容损失，确保不改变原始图像的内容信息；最后用深度可分离卷积取代普通卷积，进一步降低参数量并提高图片生成质量。

2.1 蒸馏生成器

为了让学生生成器学习教师生成器的知识，直接最小化两个生成器生成的图像的欧氏距离，为

$$L_{\mathrm{L1}}(G_{\mathrm{S}})=\frac{1}{n}\sum_{i=1}^{n}\| G_{\mathrm{T}}(x_i)-G_{\mathrm{S}}(x_i)\|_1^2 \qquad (1)$$

式中：G_{T} 表示教师生成器，G_{S} 表示学生生成器，其中 $\|\cdot\|$ 表示 L1 正则化。通过最小化式（5），学生生成器的结果可以从像素层面上与教师生成器相似，L1 距离损失的目标只是最小化平均合理的结果。但是生成器的训练是伴随着判别器的，因此只蒸馏生成器对学生生成器的学习是不够的。由于教师判别器与生成器任务高度相关，要求教师判别器能够评估学生生成器是否像教师生成器那样生成了高质量的图像，即生成器的感知损失，为

$$L_{\mathrm{perc}}(G_{\mathrm{S}})=\sum_{i=1}^{n}\| D_{\mathrm{T}}(G_{\mathrm{T}}(x_i))-D_{\mathrm{T}}(G_{\mathrm{S}}(x_i))\|_1^2 \qquad (2)$$

式中：D_{T} 表示教师判别器，同时生成器的输入中省略了属性标签信息。

因此，对生成器的蒸馏损失为

$$L_{\mathrm{KD}}^{\mathrm{G}}(G_{\mathrm{S}})=L_{\mathrm{L1}}(G_{\mathrm{S}})+\gamma L_{\mathrm{perc}}(G_{\mathrm{S}}) \qquad (3)$$

其中 γ 是平衡两个损失函数的超参数。

2.2 蒸馏判别器

Aguinaldo 等[17]对生成式网络压缩时，没有利用教师判别器对学生判别器蒸馏，而判别器对 GAN 的训练也很重要，本文首先设计学生判别器，然后对学生判别器进行蒸馏。

学生判别器用来协同学生生成器训练，同时用教师判别器进行蒸馏，在蒸馏过程中，使用了和蒸馏生成器相同的方法，直接对教师和学生的判别器的输出作 L1 损失，即对判别器蒸馏损失为

$$L_{\mathrm{L1}}(D_{\mathrm{S}})=\frac{1}{n}\sum_{i=1}^{n}\| D_{\mathrm{T}}(x_i)-D_{\mathrm{S}}(x_i)\|_1^2 \qquad (4)$$

其中 D_{S} 表示学生判别器，对设计的学生判别器通过对其输出值进行蒸馏。判别器在多数情况类似于二分类，而分类模型中的标签平滑能够让判别器更好地学习教师判别器，而判别器对于生成器的训练至关重要，整体的蒸馏网络模块如图 4 所示。

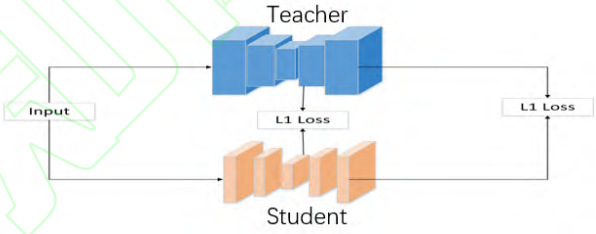

图 4 整体蒸馏网络模块示意图

Fig. 4 Diagram of whole distillation network module

此外，本文采用了教师网络和学生网络的对抗学习，教师网络经过很好的训练，学生判别器在教师网络的监督之下训练，通过教师生成器生成的图像应该被学生判别器判别为真，损失函数定义为

$$L_{G_{\mathrm{T}}}(D_{\mathrm{S}})=\frac{1}{n}\sum_{i=1}^{n}D_{\mathrm{S}}(G_{\mathrm{T}}(x_i),\mathrm{True}) \qquad (5)$$

2.3 skip-connection

StarGAN 网络包括下采样模块、骨干网络以及上采样模块。在生成模型中，下采样主要用于编码功能，完成对潜在向量的编码，骨干网络主要提取图像特征，而上采样模块主要用于解码，还原为图像。在学生网络设计中，骨干网络选择与教师网络相同，都是 ResNet 模块，在下采样和上采样之间使用 skip-connection 提供跨模块的连接。跨模块连接示意图如图 5 所示。在学生网络中将具有同样尺寸的上采样和下采样中的模块进行 skip-connection，连接的过程不是简单的求和，而是通道的叠加，尽可能地保留低层的信息。

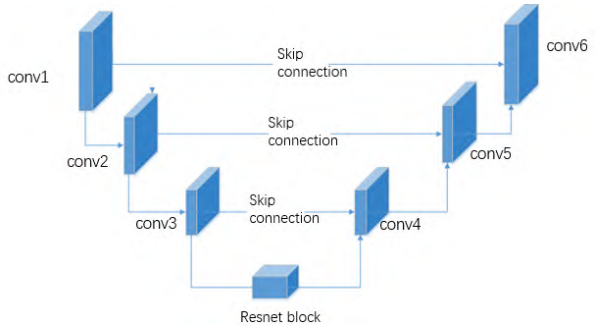

图 5　跨模块连接示意图

Fig. 5 Diagram of skip-connection

2.4 内容损失函数

为了在图像的风格变换后，与风格无关的信息能够保留下来，而对生成图像和原图像作内容损失。本文中图像的属性改变，但是不希望与属性无关的其他信息，如结构、背景等发生变化，本文采用内容损失函数，在消融实验中验证其有效性。

内容损失函数通过 1 个预训练好的网络作为特征提取器，由于模型高层的输出是高维的语义特征，包含更多具体的内容信息，因此，本文采用 ResNet-18 模型的最后 1 个卷积层的输出作为特征提取器。分别提取原图像和学生网络生成器生成的图像的特征，在像素层面上作 L1 损失，为

$$L_{\text{content}} = E_x[F(x) - F(G(x,c))] \tag{6}$$

其中：c 表示 x 要转换的目标属性类别，生成器以原图像和目标属性类别作为输入；F 则是 1 个预训练好的 ResNet-18[25]模型。

2.5 深度可分离卷积

在学生网络设计中，替换掉骨干网络 Resnet 模块中的普通卷积，改为深度可分离的卷积，进一步减少模型计算量，降低网络复杂度。在 Google 的 MobileNet[26]证明了深度可分离卷积的性能，以及相较于普通卷积，能极大降低计算量。

假定输入通道数为 M，输出通道数为 N，标准卷积的卷积核大小为 $D_K \cdot D_K$，特征图大小为 $D_F \cdot D_F$，则采用标准卷积计算量为

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \tag{7}$$

若采用深度可分离卷积，计算量为

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \tag{8}$$

联立二者计算量大小，可得出其比值为

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \tag{9}$$

通常标准卷积核大小为 3×3，即 D_K 为 3，N 为输出通道数通常较大，一般为 256 或 512，因此深度可分离卷积的计算量只有大约标准卷积的 1/9。

3 实验与分析

3.1 数据集

为了对所提学生网络的轻量化和性能进行验证，本文首先在 CelebA 上做消融实验验证模型不同部分的有效性，然后进一步在 CelebA 和 Fer2013 数据集进行算法对比实验。

CelebA 数据集：是人脸识别和人脸表情研究领域具有权威性和完整性的名人人脸属性数据集，包含 202599 张人脸图像，图像大小为 128×128，在原始数据集中每张图像都有 40 个属性标注。以 20%作为测试集，即 40000 张图像，其余为训练集。

Fer2013 数据集：是一个灰度图像数据集，主要用于人脸表情变化的研究，该数据集共有 7 种表情，分别对应数字标签 0~6，这 7 种表情图片共有 35886 张，其中训练集包含 28708 张，其余为测试集 7178 张。每张图片的大小为 48×48，以 csv 格式的文件存储像素值表示。

3.2 实验细节及评价指标

实验以 PyTorch 框架在 Nvidia GeForce RTX 3060 上实现，显存为 12 GB。在训练过程中，将学生网络输入和教师网络输入尺寸固定一样，参数设置：批大小 batch size 为 8，迭代次数 200000 次，生成器和判别器的学习率均为 0.0001，每 1000 次时更新 1 次，每更新 5 次判别器时更新 1 次生成器。部分损失函数权重与教师网络保持一致，类别损失函数的权重为 1，重构损失权重为 10，梯度惩罚权重为 10，生成器和判别器蒸馏损失均为 1。

为了衡量网络的轻量化，本文以网络参数量和浮点数运算次数（GFLOPs）作为评价指标。此外，图像风格转换任务常用指标包括图像信噪比（PSNR）、原图像和生成图像的结构相似性（SSIM）、生成图像质量弗雷歇初始距离（FID）。在 SRGAN[8]中，由于 PSNR 定义在像素级别的图像区别上，不能很好地表示图像的高维细节，本文只在消融实验中进行了 PSNR 指标的对比，在算法对比实验中不再对比这个指标，只

给出 FID 和 SSIM 的数据。

PSNR、SSIM 和 FID 指标计算公式分别为

$$\mathrm{PSNR}=10*\log_{10}(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}) \quad (10)$$

$$\mathrm{SSIM}(X,Y)=\frac{(2u_Xu_Y+C_1)*(2\sigma_{XY}+C_2)}{(u_X^2+u_Y^2+C_1)*(\sigma_X^2+\sigma_Y^2+C_2)} \quad (11)$$

$$\mathrm{FID}=\|u_r-u_g\|^2+\mathrm{tr}(\sum\nolimits_r+\sum\nolimits_g-2(\sum\nolimits_r\sum\nolimits_g)^{1/2}) \quad (12)$$

式中：MAX 为图像像素理论最大值；MSE 为两张图像的均方误差；u 表示均值；σ 表示方差；C 表示常数；tr 表示矩阵对角线上元素的综合，即矩阵论中成为矩阵的迹；r 和 g 表示真实的图片和生成的图片；Σ 是协方差矩阵。

3.3 实验结果

实验在 CelebA 数据集上生成的效果如图 6 所示，其中第一列为原始图像，后面 5 列依次对原图像属性进行更改，分别为黑发、金发、棕发、性别以及年龄。图中第 2 行第 3 列肤色有些变化，可能是在变换黑色头发时，对黑色肤色产生了影响。

图 6 学生网络在数据集 CelebA 上的效果示意图

Fig.6 Diagram of result of stuStarGAN on CelebA

在 Fer2013 数据集上的生成的效果如图 7 所示，图中第 1 列为原始图像，第 1 列到第 7 列分别表示该数据集中存在的 7 中表情，分别是 neutral、anger、fear、disgust、happy、sad、surprised。学生模型依然能很好地改变图像的表情属性，而不改变其他部分。

图 7 学生网络在数据集 Fer2013 上的效果示意图

Fig. 7 Diagram of result of stuStarGAN on Fer2013

模型在 CelebA 数据集上进行多属性风格图像生成的结果如图 8 所示，第 1 列表示原图像，第 2 列改变头发颜色以及性别，第 3 列改变了头发颜色以及年龄两个属性。结果表明生成器能对图像进行单属性和多属性转换。

图 8 多属性的改变示意图

Fig. 8 Diagram of change of multi attribute

3.4 消融实验

为了确保学生网络在轻量化之后仍然能有很好的图像质量，本文以教师网络为基准模型进行消融实验，在评价指标上评估各个模块或损失函数对性能的影响，结果见表 1，其中 U 表示使用 U-net 的 skip-connection，CL 表示使用内容损失，DP 表示将普通卷积更换为深度可分离卷积，DL 表示蒸馏生成器同时新增判别器蒸馏损失。

表 1 在 CelebA 数据集上不同模块的对比

Tab. 1 Comparison of different module on CelebA

模型	PSNR	SSIM	参数量/M	GFLOPs	FID
StarGAN	22.324	0.878	8.43	2.23	40.438
+KD	21.943	0.867	3.12	1.31	43.132
+U	22.714	0.886	3.12	1.31	39.093
+DL	22.753	0.885	3.12	1.31	38.256
+CL	22.712	**0.887**	3.12	1.31	37.543
4+DP	22.821	0.884	**1.55**	**0.31**	37.923
本文模型	**22.853**	0.885	**1.55**	**0.31**	**36.112**

在表 1 中，第 1 行为教师网络，KD 表示知识蒸馏，第 2 行+KD 表示只对教师网络进行知识蒸馏的结果，以后的每一行都是在前面的基础上进一步改进学生网络，如第 3 行是在第 2 行的基础上采用 skip-connection。比较第 1 行和第 2 行，只对 StarGAN 蒸馏，模型虽然显著降低参数量，但性能却有下降。学生网络在使用 skip-connection 之后效果有改善，表明模型的底层信息正确地传

递给了上层神经元。

比较第 3 行和第 4 行，其中 DL 表示对学生判别器采取的蒸馏损失，也就是直接比较教师和学生判别器的输出，包括真假概率输出和属性类别输出，可以看出使用 DL 后确实提高了学生网络的效果。

第 5 行在前面基础上采用内容损失，SSIM 进一步提高，说明生成图像很好地保留了原图像的结构信息，而 PSNR 相较第 4 行有明显下降。推测其原因主要在于引入内容损失后，为了保证图像主题结构信息不变，生成图像在其余部分引入噪声，使得图像信噪比下降，即 PSNR 降低。第 6 行 4+DP 是指在第 4 行的基础上引入 DP，结果表明 CL 确实影响了 PSNR 指标，然而引入 DP 之后 PSNR 有明显提升，但是 SSIM 相较第 4 行有所下降。比较第 4 行和最后一行，虽然 SSIM 有些微的降低，但是深度可分离卷积成倍地降低了网络的参数量和计算量，并提升了网络模型的 FID。因此，虽然 PSNR 和 SSIM 有些微变化，FID 进一步提高的情况，本文选择了参数量和计算量更少的模型作为最后的学生网络。从第 5 列浮点数运算次数，表明本文模型有更少的运算量，说明模型确实降低了网络的计算量。

表 1 中参数量只包括生成器参数，因为判别器只在训练阶段起作用，最终也只需是部署生成器。

3.5　对比实验

在 CelebA 以及 Fer2013 两个数据集上进行对比实验，在 CelebA 数据集上，本实验与近年来的一些在图像翻译领域的先进模型进行比较，包括 pix2pix、CycleGAN、StarGAN、UE-StarGAN 等结果见表 2。

表 2　不同算法在 CelebA 数据集上的性能比较

Tab. 2 Comparison of different methods on CelebA

模型	SSIM	FID	参数量/M
pix2pix[13]	0.767	39.7	54.4*10
CycleGAN[10]	0.749	38.4	52.6*10
StarGAN[14]	0.788	40.4	53.2
UE-starGAN[27]	0.881	-	53.2
本文模型	**0.885**	**36.1**	**28.4**

从表 2 中可以看出，在 CelebA 数据集上，在 SSIM 和 FID 两个指标上，学生网络都有较好的性能。由于本文主要设计能生成多属性风格图像的轻量化学生网络，在保证生成质量的前提下降低模型复杂度，因此关于参数量，是在考虑了判别器参数的情况，依然以小于最少参数量的 50%，其中 CycleGAN 的参数量中的 *10 是因为 CycleGAN 的 1 个生成器只能转换原图像的属性值，而 StarGAN 可以 1 个生成器转换多种属性值。

该学生模型也在 Fer2013 数据量上进行对比实验，实验按照数据集中 7 种人脸属性对每张图像变换，并于其他模型比较，结果见表 3。

表 3　不同算法在 Fer2013 数据集上的性能比较

Tab. 3 Comparison of different methods on Fer2013

模型	SSIM	FID	参数量/M	GFLOPs
pix2pix[13]	0.834		54.4*14	18.15*14
CycleGAN[10]	0.859		52.6*14	19.22*14
StarGAN[14]	0.886	163	53.2	18.74
UE-starGAN[27]	0.866		53.2	18.74
本文模型	**0.896**	**159**	**20.4**	**8.92**

在表 3 中，从第 2 列可以看到本文所提出的学生网络可以保证很高的结构相似性，并且参数量和 GFLOPs 有较大地降低，相较其他模型性能接近的情况下，计算量更少，结构复杂度更低，表明了该学生模型的有效性。

4 结 语

本文基于 StarGAN 设计了 1 个能生成多属性风格变化的轻量化网络 stuStarGAN。模型首先应用知识蒸馏技术降低教师网络参数量；然后为了确保生成图片质量，采用 skip-connection 提供跨模块的连接，使用内容损失确保生成图像和原始图像的内容信息一致，在蒸馏生成器的同时新增判别器蒸馏损失以提高生成器性能；并将普通卷积替换为深度可分离卷积，进一步降低参数量并提高图片生成质量；最后将模型在两个数据集进行实验给出了单属性和多属性生成的效果；并与其他模型进行比较，在保证生成图片质量的基础上极大低降低了参数量和计算量，可以应用于实际场景采集数据不足和数据分布不均需要扩充数据集以及某些实时应用场景中，如监控行人数据集样本分布不均、社交网络更改头像保护隐私、游戏角色变化头像等具体场景中。由于学生网络生成的图像多样性不足，在后续研究中会考虑并继续完善模型，做好效率和精度之间的平衡。

参考文献：

[1]　谭明奎，许守恺，张书海，等. 深度对抗视觉生成综

述[J]. 中国图象图形学报，2021，26(12):2751-2766.

[2] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial networks [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1406.2661v1.

[3] MIRZA M，OSINDERO S. Conditional generative adversarial nets [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1411.1784v1.

[4] RADFORD A，METZ L，CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1511.06434v2.

[5] ARJOVSKY M，CHINTALA S，BOTTOU L. Wasserstein GAN [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1701.07875v3.

[6] GULRAJANI I，AHMED F，ARJOVSKY M，et al. Improved training of wasserstein GANs [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1704.00028v3.

[7] ZHANG H，GOODFELLOW I J，METAXAS D，et al. Self-attention generative adversarial networks [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1805.08318v2.

[8] LEDIG C.，THEIS L，HUSZAR F，et al. Photorealistic single image super resolution using a generative adversarial network[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2017:4681-4690.

[9] IIZUKA S，SIMO-SERRA E，ISHIKAWA H，et al. Globally and locally consistent image completion[J]. ACM Transactions on graphics，2017，36(4):1-14.

[10] ZHU J Y，PARK T，ISOLA P，et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision (ICCV). New York：IEEE，2017:2223-2232.

[11] YI Z L，ZHANG H，TAN P，et al. DualGAN: Unsupervised dual learning for image-to-image translation[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision (ICCV). New York：IEEE，2017:2849-2857.

[12] YANG S，JIANG L M，LIU Z W，et al. Pastiche master: exemplar-based high-resolution portrait style transfer[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2022: 7693-7702.

[13] ISOLA P，ZHU J Y，ZHOU T H，et al. Image-to-image translation with conditional adversarial networks[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2017:1125-1134.

[14] CHOI Y，CHOI M J，KIM M Y，et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation[C]// IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2018:8789–8797.

[15] MAO Q，LEE H Y，TSENG H Y，et al. Mode seeking generative adversarial networks for diverse image synthesis[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2019:1429-1437.

[16] CHOI Y，UH Y J，YOO J，et al. StarGAN v2: diverse image synthesis for multiple domains[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2020:8188–8197.

[17] AGUINALDO A，CHIANG P Y，GAIN A，et al. Compressing GANs using knowledge distillation [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1902.00159v1.

[18] CHEN H，WANG Y H，SHU H，et al. Distilling portable generative adversarial networks for image translation [EB/OL]. [2023-03-25]. https://arxiv.org/abs/2003.03519v1.

[19] HINTON G.，VINYALS O，DEAN J. Distilling the knowledge in a neural network [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1503.02531v1.

[20] ROMERO A，BALLAS N，BENGIO Y，et al. FitNets: hints for thin deep nets [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1412.6550v4.

[21] ZAGORUYKO S，KOMODAKIS N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1612.03928v3.

[22] TIAN Y L，KRISHNAN D，ISOLA P. Contrastive representation distillation[C]// International Conference on Learning Representations (ICLR)，2020:1-19.

[23] LIN M，CHEN Q，YAN S C. Network in Network [EB/OL]. [2023-03-25].https://arxiv.org/abs/1312.4400v3.

[24] YANG G X，FEI N Y，DING M Y，et al. L2M-GAN: learning to manipulate latent space semantics for facial attribute editing[C]// IEEE. Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2021: 2951–2960.

[25] HE K M，ZHANG X Y，REN S Q，et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2016: 770–778.

[26] HOWARD A G，ZHU M L，CHEN B，et al. MobileNets: efficient convolutional neural networks for mobile vision applications. [EB/OL]. [2023-03-25]. https://arxiv.org/abs/1704.04861v1.

[27] 许新征，常建英，丁世飞. 基于 StarGAN 和类别编码器的图像风格转换 [J]. 软件学报，2022，33(4): 1516-1526.

责任编辑：郎婧