

基于循环生成对抗网络的面部表情图像自动生成方法

袁佩瑶^{1,2} 林晶晶^{1,2} 曹林葳¹
YUAN Peiyao LIN Jingjing CAO Linwei

摘要

面部表情细节繁杂、肌肉运动组合多样,且会受到姿态、光照等外界因素的影响,现有以单个特征为引导的生成方法很难达到高一致性的表情合成效果,使得生成图像和真实图像的轮廓、纹理等方面存在较大的差异,从而影响到视觉效果以及应用的可靠性。因此,文章提出了一种基于循环生成对抗网络(CycleGAN)的面部表情图像自动生成方法。通过单元分割、归一化处理、对齐剪裁得到面部关键点,按照 FACS 标准构造结构化的 AU 标签,给生成过程提供语义引导。通过建立 AU 标签驱动的 CycleGAN 生成器,利用混合注意力机制和残差结构来完成多尺度特征融合以及自适应的表情合成,用 InstanceNorm 处理来提高生成结果的真实性、一致性。使用 PatchGAN 判别器对生成图像进行真伪判断和表情分类,可以有效地解决轮廓、纹理偏离的问题,提高图像质量,保证输出的可靠性。在 CelebA 多表情数据集上进行实验,得到的方法 FID 值为 11.2~16.5, PSNR 测点值均大于 24.55 的基线标准,比其他方法效果好,说明该方法可以提高生成的一致性、结构的真实性。该研究为解决复杂条件下面部表情生成的技术难题提供了一个有理论参考价值、工程应用潜力的解决办法。

关键词

循环生成; 对抗网络; 面部表情; 图像生成; 轮廓定位; 图像处理

doi: 10.3969/j.issn.1672-9528.2026.05.038

0 引言

面部表情图像的生成同外部环境、特定因素、面部状态等存在直接联系,常规的生成方式大多依靠定点识别,输出相应的图像,虽然可以完成任务,但受数据采集的限制,部分图像采集时会产生局部重叠,造成生成结果达不到预期标准。

近年来,众多学者开展研究,如林本旺等^[1]设计了基于组残差块生成对抗网络的面部表情生成方法。该方法在生成网络下进行关键区域特征的提取,将各组特征融入残差网络中,以各组残差块进行引导生成,达到设定的生成效果。该方法提取的关键区域特征较为单一,无法反映全局特征,导致生成图像与真实图像不一致。公正^[2]提出了虚拟数字人主播的表情生成技术。该方法根据主播基础表情设置标签,按照播报内容的变化,自适应捕捉,完成目标生成。但是多维度表情标签的设定,无法建立统一的映射,不同条件下生成

的图像难以保持一致。李韩超等^[3]设计了带虚拟边约束的面部表情基生成方法,该方法依据人脸模型分析添加顶点的连接关系,以虚拟边约束网格形变迁移算法,融合映射点实现面部表情的生成。然而在虚拟边约束在图像生成的过程中,容易出现点位偏离,致使真实图像与生成图像之间缺乏一致性。曹朗炜等^[4]设计的基于面部动作单元和 Transformer 的人脸表情生成方法。该方法利用自注意力机制关注人脸输入序列的不同部分,提取底层特征,使用判别器按照特征进行多尺度融合,强化表情图像的生成效果。但是底层特征的映射范围有限,对于表情细节位置的生成并不协同,导致生成结果不一致。在该背景下,本文结合循环生成对抗网络,设计面部表情图像自动生成方法。

1 面部表情图像循环对抗网络自动生成方法

1.1 面部表情图像生成前的 AU 标签设置与特征处理

面部表情复杂多样,不同表情由不同面部肌肉运动组合而成,直接生成特定表情图像易出现表情模糊、不准确或与预期不符的情况。为解决这些问题,进行 AU 标签设置与特征处理,具体过程如下:

首先,按照生成单元及对应的特征设置目标 AU 标签^[5]。对参考面部扫描,在 CelebA 中建立数据集,获取 landmark

1. 郑州西亚斯学院工学部 河南郑州 450000
2. 河南省智能制造数字孪生工程研究中心 河南郑州 450000
[基金项目] 河南省 2021 年民办普通高等学校学科专业建设资助项目(软件工程专业); 郑州西亚斯学院 2024 年度校级科研项目(2024XKD069)

坐标信息。将图片大小调整为统一的标准,并进行单元分割,同时开展归一化处理和对齐图像剪裁^[6]。归一化处理和对齐图像剪裁用公式表示为:

$$D = \int a^e \times \left(\sum_{i=1}^n \mu i + \eta i \right) - \frac{\sigma_1}{\delta} \times (\sigma_2 + 1) \quad (1)$$

$$J = \frac{\mu + \eta}{D} - \sigma_1 \sigma_2 \quad (2)$$

式中: D 表示图像单元归一化处理结果; a^e 表示目标区域; μ 和 η 分别表示几何和人脸归一化范围分别; i 表示单元点; δ 表示单元点距离; σ_1 和 σ_2 分别表示歪头角度和正脸角度; J 表示对齐图像剪裁结果。

完成单元性剪裁后,根据特征向量的变化规律,制定针对面部表情目标的 AU 标签^[7]。使用 FACS 定义编码标准,并以 AU 单元映射标签节点,具体如图 1 所示。

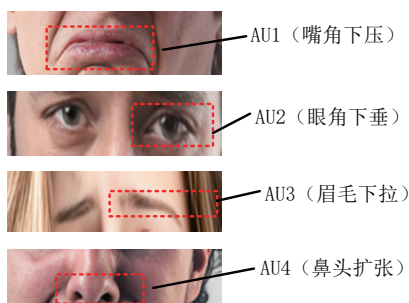


图 1 面部表情单元 AU 标签

按照图 1 设置的 AU 标签,记录悲伤状态下人脸面部表情的基础轮廓。AU 标签在生成的过程中需要按照人脸器官的位置与生成顺序设定,确保生成的图像协调,能够精准捕捉到表情即可^[8]。依照标签调取映射的节点,形成对应表情的轮廓,提取出轮廓特征值,具体用公式表示为:

$$K = \delta \sigma_2 \times \sum_{R=1}^n \varphi^e R \times |\nu| \sigma_1 \quad (3)$$

式中: K 表示特征值; φ^e 表示剪裁重叠位置; R 表示重叠点; ν 表示表情单元。

最后,将表情的特征单元导入 AU 标签之内,形成组合特征标签,依据给出的生成需求,结合标签内附带的特征输出图像,这样的图像具有更强的针对性。

1.2 AU 标签驱动下构建循环生成对抗网络表情图像生成器

在面部表情图像生成过程中, AU 标签设置与特征处理虽已初步提取表情特征,但实现高质量图像合成仍需强大的生成模型。循环生成对抗网络(CycleGAN)因能在无配对数据的情况下学习图像映射关系,成为理想选择。将 AU 标签与 CycleGAN 相结合,可利用标签的精确特征指导生成过程,使生成器更精准地生成符合目标表情的图像,充分发挥 AU 标签的规范性与 CycleGAN 的生成能力优势。

在生成机制中加入混合注意力机制组残差块,得到一个

由卷积层、残差块、反卷积层组成的结构。按照给定的表情生成要求,结合 AU 标签建立映射结构。该结构是实现多尺度特征融合的,可以同时得到全局的表情轮廓和局部的肌肉细节。以式(3)为引导,对特征进行自适应细化处理,在卷积层中分析特征的变化关系,完成表情的自适应集成处理,具体描述为:

$$F = |\omega - \omega'| \cdot \frac{1}{\psi} \sum_{O=1}^S \varphi_O^e \odot K \quad (4)$$

$$E = \mu \varphi^e - F \times \frac{1}{x} \quad (5)$$

式中: F 为特征细化结果; ω 、 ω' 为输入采样、丢失采样; O 为采样点; ψ 为映射函数; E 为图像单元集成处理结果; x 为嵌入生成点。以细化后的特征为引导,将循环生成对抗网络中的残差块、反卷积层相结合,在式(4)和 AU 标签的指导下,设计表情图像生成器,如图 2 所示。

根据图 2 原理,对生成目标做卷积处理,用残差块把细化后的特征和卷积目标传入生成器中,用 ReLU 激活函数激活初始生成的目标。由于生成过程中残差块是串联的,传递的特征只在第一层得到充分的关注,因此生成的图像不全面。因此,在得到最终生成的结果之前要对表情图像进行 InstanceNorm 处理提高生成结果的一致性、细节的真实性,使生成的目标可靠,用公式表示为:

$$Y = \varphi^e x \times \left[(\psi F + \sigma_1 \sigma_2) - y^{\mu} \right] \quad (6)$$

式中: Y 为经过 InstanceNorm 处理后的生成结果; y^{μ} 为面部表情增强的位置。此时生成的初始表情图像细节还不太清楚,还需要做进一步的验证和判断才能得到准确的目标。

1.3 PatchGAN 判别验证实现生成处理

在完成 AU(动作单元)标签驱动下构建循环生成对抗网络表情图像生成器后,虽然生成器能够生成表情图像,但仅依靠生成器自身无法保证生成图像的质量和真实性。

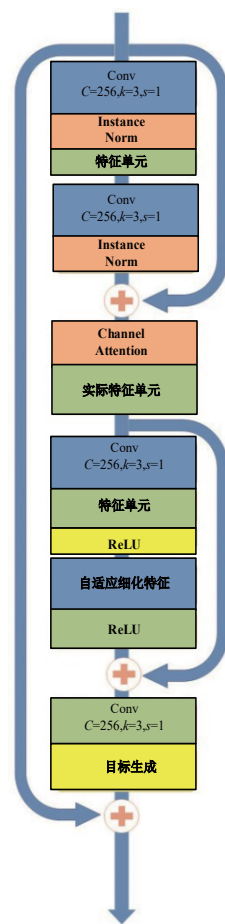


图 2 循环生成对抗网络表情图像生成器结构图

PatchGAN 判别器具有独特的结构和特性，能够对生成器输出的结果进行细致的评估，从多个方面判断生成图像的优劣，从而为生成器的优化提供有效反馈，为此，本研究引入 PatchGAN 结构作为判别器，将生成目标导入其中，判别生成图像的真伪，并分析输入图像表情的所属类别，具体用公式表示为：

$$Q = \frac{y^{iii}}{x+y} \times \left(\frac{\omega}{y} - \frac{\omega'}{y} \right) \cdot \aleph$$

(7)

式中：Q 表示判定器输出结果；y 表示 PatchGAN 判定属性； $\aleph$ 表示生成图像目标域。结合 PatchGAN 判别器的处理，验证出判别过程中的损失函数，去除损失后，输出源域中面部表情图像的生成结果，具体用公式表示为：

$$N = \omega' y \times \aleph \left(QY - \varphi^e p \right)^E$$

(8)

$$X = \sqrt{p^N} \times \frac{\kappa}{N} \left(\sigma_1 x - \sigma_2 x \right)$$

(9)

式中：N 表示损失函数；p 表示重叠域；X 表示生成结果； p^N 表示重建生成目标； κ 表示生成表情类别损失。生成的面部表情图像为经过判定的可靠图像，通过分类器进行类别划定，实现自动生成归类。

2 实验

为验证所提方法的应用效果，在文献中选定虚拟数字人主播表情生成技术、带虚拟边约束的面部表情基生成方法作为对照组，开展对比实验。

2.1 数据集与实验环境部署

为保证实验结果的真实与可靠，增加面部表情图像的还原度及清晰度，需要建立一个可控的数据集。数据集中采集的图像共 9 084 张，大小统一转换为 681 px×1 024 px 的标准。每张图像的表情单元均标注 AU 标签，便于后续识别及调度处理。此外，数据集中的图像所设定的离散表情包括开心、悲伤、愤怒、惊讶、厌恶等，较为多样化，且为保证实验的真实性，数据集中的图像均为正面照，具有明显的面部特征，可以精准识别到面部表情，图 3 为部分数据集图像的展示。

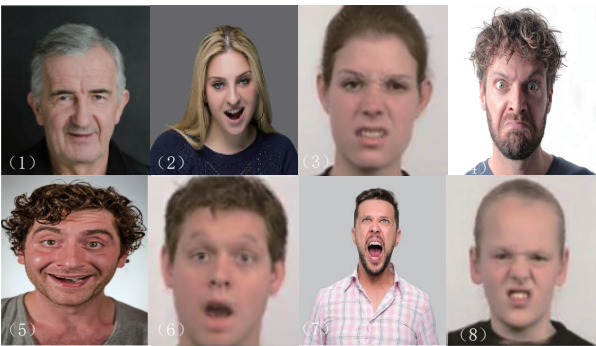


图 3 面部表情数据集部分图像展示

根据图 3 展示的局部数据集，参与当前的实验测试。为进一步完善实验环境，还需对控制条件与参数标准展开基础设置，具体如表 1 所示。

表 1 图像生成实验条件项与参数标准

实验条件项	参数标准值
迭代次数 / 次	3 000
batch size	16
图像后期剪裁尺寸	256 px×256 px
重建损失函数	9.3
类别损失函数	11.4

依据表 1 调整当前实验环境，并在测试平台中接入辅助生成图像的软件及 PyTorch 训练模型，以此来完善实验条件，展开具体测试分析。

2.2 AU 标签生成结果

数据集中随机选定 2 组图像作为生成的原始目标，基于循环生成对抗网络，按照预先设定的面部单元及轮廓，进行 AU 标签的设置，具体结果如图 4 所示。



图 4 AU 标签生成结果分析

根据图 4 结果分析可知：两组人像生成了包括开心、悲伤、厌恶、惊讶等多种表情图像。生成过程中，系统能够自动定位面部关键器官，并依序标注 AU 标签，有效凸显了不同表情下的面部特征变化，使得生成图像在结构上更为完整和生动。在循环生成对抗网络的框架下，基于 AU 标签的表情生成与特征优化同步进行，生成器与判别器的协同作用提升了表情的真实性和一致性，减少了生成图像模糊或结构偏离的问题。

2.3 面部表情生成 PSNR 值结果

上述生成的 3 组面部表情图像作为验证目标，统一选定开心图像作为生成效果验证的对象，在图像轮廓中随机标定出 6 个测点，以原始图像作为比对目标，输出 PSNR 结果，具体如图 5 所示。

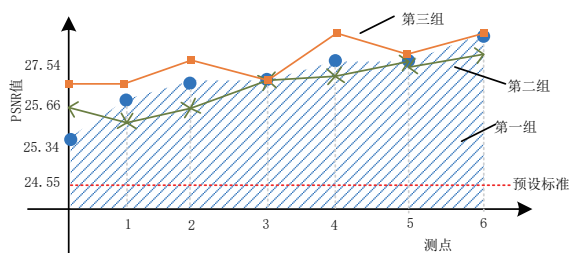


图5 面部表情图像生成 PSNR 值结果

根据图5结果分析研究：第一组图像中，6个测点输出的PSNR值均可以达到25.34左右；第二组图像的测点PSNR值结果达到25.66左右；第三组所选测点得到的PSNR值结果则在26.18左右。相较于设定的24.55生成验证标准，本文所提方法的生成效果更佳，稳定性明显较高。循环生成对抗网络的融合下，ReLU对目标的激活能够明显生成点，减少轮廓偏离，而InstanceNorm处理则基于上述标准进一步锁定生成图像的面部表情轮廓，促使图像更加清晰，表情特征也得到了增强，提升生成处理的一致性及应用性，可应用价值高。

2.4 性能结果对比分析

在预设的特征空间之中，明确原始图像与生成的图像的分布差异，通过计算FID值来量化图像分布的相似度，输出结果如图6所示。

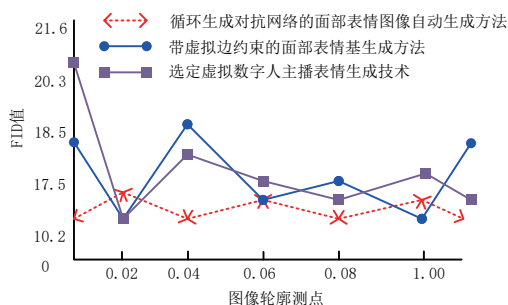


图6 图像生成 FID 值结果对比分析

由图6的结果可以看出，本研究方法的FID值分布在11.2~16.5之间，波动较小，说明生成的图像质量好且比较稳定。现有文献方法的FID值在14.5~21.4之间，显著高于本研究方法，说明其生成效果与真实图像有很大的差别。本研究利用残差块串联结构来完成多层次特征传递，使用PatchGAN判别机制对生成的结果做真假判断和类别验证，使生成的图像具有较好的结构一致性以及真实性。

3 结语

本文提出了一种使用循环生成对抗网络(CycleGAN)进行面部表情图像自动生成的方法。对原始图像进行单元分割、对齐剪裁、归一化预处理之后，系统可以得到面部关键区域

的精确定位和AU标签的结构化标注。用AU标签作为语义引导，采用多尺度特征融合和混合注意力机制来创建具有自适应细化能力的循环生成对抗网络生成器，很好地解决了真实图像和生成图像之间轮廓偏离的问题，通过InstanceNorm处理提高了生成图像的结构一致性以及质量。判别阶段加入PatchGAN结构，用图像真伪判别和表情类别验证来提高生成图像的可靠性以及身份保持能力。实验结果表明，采用本文提出的方法得到的CelebA数据集上的图像FID值稳定在11.2~16.5之间，PSNR测点值都比基线标准高24.55，生成的一致性、图像质量更好。与目前主流的方法相比，本方法在FID指标以及视觉一致性上都更好，证明了它对于提高生成图像真实性、一致性有较好的效果和先进性。

参考文献：

- [1] 林本旺, 赵光哲, 王雪平, 等. 基于组残差块生成对抗网络的面部表情生成[J]. 计算机工程与应用, 2024, 60(5): 240-249.
- [2] 公正. 虚拟数字人主播的表情生成技术研究[J]. 电视技术, 2024, 48(8): 216-218.
- [3] 李韩超, 沈成泽, 刘新国. 带虚拟边约束的面部表情基生成方法[J]. 计算机学报, 2023, 46(11): 2453-2462.
- [4] 曹朗炜, 魏润辰, 兰星, 等. 基于面部动作单元和Transformer的人脸表情生成方法[J]. 计算机与网络, 2024, 50(5): 440-447.
- [5] 陈姣, 刘小红. 基于生成对抗网络的遮挡人脸表情识别方法研究[J]. 现代计算机, 2023, 29(6): 77-79.
- [6] 程艳, 蔡壮, 吴刚, 等. 结合自注意力特征过滤分类器和双分支GAN的面部表情识别[J]. 模式识别与人工智能, 2022, 35(3): 243-253.
- [7] 潘旭宏. 基于注意力机制的特征融合面部表情识别方法[J]. 电子制作, 2025, 33(7): 27-31.
- [8] 范蕴睿. 基于PCNN的角色表情动画去毛刺生成方法[J]. 常州工学院学报, 2024, 37(1): 27-33.

【作者简介】

袁佩瑶(1996—), 女, 河南郑州人, 硕士, 助教, 研究方向: 计算机视觉。

林晶晶(1996—), 女, 河南信阳人, 硕士, 助教, 研究方向: 图像处理。

曹林葳(2003—), 男, 河南驻马店人, 本科, 研究方向: 计算机视觉。

(收稿日期: 2025-09-19 修回日期: 2026-05-09)