

基于生成对抗网络的短波红外-可见光人脸图像翻译

胡麟苗^{1,2,3}, 张湧^{1,3*}

¹中国科学院上海技术物理研究所, 上海 200083;

²中国科学院大学, 北京 100049;

³中国科学院红外探测与成像技术重点实验室, 上海 200083

摘要 提出一种用于短波红外人脸图像与可见光人脸图像翻译的改进 CycleGAN 框架。基于 CycleGAN 框架, 新增了损失函数计算通路并设计了新损失函数。建立数据集并通过实验调整模型参数, 改进模型在人脸图像上的翻译效果, 有效克服光谱特性不同带来的图像模态差异, 提升了图像的可观察性。在自建数据集上进行实验验证, 将所提框架与其他常用框架从主观评价、FID(Fréchet inception distance)及识别准确率三个方面进行比较。结果表明, 所提框架提升效果明显, 更好地保持了原目标的结构特征, 有效提升了图像翻译结果的可观察性和识别准确率。

关键词 图像处理; 图像翻译; 短波红外图像; 生成对抗网络; 损失函数

中图分类号 TP391

文献标志码 A

doi: 10.3788/AOS202040.0510001

Facial Image Translation in Short-Wavelength Infrared and Visible Light Based on Generative Adversarial Network

Hu Linmiao^{1,2,3}, Zhang Yong^{1,3*}

¹ Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;

² University of Chinese Academy of Sciences, Beijing 100049, China;

³ Key Laboratory of Infrared System Detection and Imaging Technology, Chinese Academy of Sciences, Shanghai 200083, China

Abstract We proposed an improved CycleGAN framework for translating short-wavelength infrared facial images and visible-light facial images. Based on the CycleGAN framework, a loss function calculation path was added and a new loss function was designed. A dataset was established, and the model parameters were adjusted based on experiments to improve the translation effect of the proposed model on the facial images. It effectively overcame the differences in images caused by different spectral characteristics so that the images could be easily recognized. The experimental verification was performed with a self-built dataset. The subjective evaluation, FID(Fréchet inception distance), and recognition accuracy were used to compare the proposed framework with several other frameworks. The results show that the improvement of the proposed framework is obvious and the structural features of the original target are better maintained, which effectively improves the observability and recognition accuracy of image translation results.

Key words image processing; image translation; short-wave infrared image; generative adversarial network; loss function

OCIS codes 100.4994; 100.4996; 150.1135

1 引言

短波红外波段是基于反射成像的红外波段, 主要依靠传感器接收物体反射的短波辐射进行成像。

与可见光和近红外波段成像系统相比, 短波红外成像系统具有更远的探测距离, 更好地穿透雾、霭、烟尘的能力^[1-2]及更好的夜间成像能力。它可以依靠大气辉光、星光等自然光源或主动式短波红外照明

收稿日期: 2019-09-23; 修回日期: 2019-10-31; 录用日期: 2019-11-09

基金项目: 国家十三五国防预研项目(Jzx2016-0404/Y72-2)、上海市现场物证重点实验室基金(2017xcwzk08)

* E-mail: zybyx@sina.com

光源进行成像。依靠短波的不可见特性,主动照明系统具有较好的隐蔽性,使得短波红外成像系统适用于安防、司法、军事等领域^[3-5]。因为光谱特性的差别,短波红外图像与目标的可见光图像有较大差异。在对目标进行观察时,短波红外图像容易发生误判、漏判的情况,通过引入图像翻译过程,输出彩色可见光图像,有助于提高观察精度与可靠性。

图像翻译是图像处理领域的一个重要研究方向,它的主要任务是在不同图像域之间对图像进行转换,包括图像超分辨^[6]、图像上色^[7]、图像填充^[8]、风格迁移^[9]及属性转换^[10]等。传统方法依据内在差异,将不同域的图像分开处理^[11]。随着深度学习的发展,尤其是全卷积网络^[12]和条件生成对抗网络^[13]的出现,图像翻译领域取得了较大的进展,人们对图像翻译问题也有了新的理解。研究者通常将图像的 latent space 分为内容空间和风格空间两部分,在图像翻译时,只需保持内容空间不变,将风格空间替换为目标域的风格空间^[14]。

Goodfellow 等^[15]在 2014 年提出生成对抗网络(GAN)模型,它能同时使用生成器和判别器进行数据生成,学习目标分布。该模型多被用于图像的生成,如飞行目标的仿真生成^[16]及 X 光图像的生成等^[17]。Isola 等^[18]借鉴了 U-Net 结构^[19],提出 Pix2Pix 框架。之后,研究者们提出了 CycleGAN^[20]、DualGAN^[21]和 DiscoGAN^[22],三者思想相近,都引入了对偶结构和循环一致性损失函数,较好地保证了图像“内容”与“风格”的一致性。之后,研究者们又提出了 UNIT^[23]、MUNIT^[14]、DRIT^[24]等多种图像翻译框架。目前,对人脸图像生成、翻译的研究主要集中在可见光领域,例如从灰度图、2D 人脸图到 3D 人脸图像的翻译^[25]和线描图到真实人脸图像^[26]的翻译等。在红外领域,多数研究集中于近红外波段^[27]和长波波段^[28-29],对应的可见光图像以灰度图为主,只有部分工作对红外图像与彩色可见光图像的翻译进行了研究^[30]。Lezama 等^[31]提出了基于卷积神经网络(CNN)的近红外-可见光的图像翻译。Song 等^[27]则引进对抗学习思想,进行了近红外-可见光的人脸图像生成和识别的相关工作。由于物质对不同波段的吸收与反射的特性不同,不同波段的红外人脸图像差异较大。例如,玻璃的短波红外透过特性与可见光类似,可以在短波红外图像中观察到眼镜后面的人眼部分,而中长红外波段透过率较低,会在眼镜处出现黑色阴影^[32-33],使用短波红外成像可以很好地规避眼镜带

来的遮挡干扰。但是受限于短波红外数据集的获取难度,短波红外-可见光图像翻译的相关研究和应用还比较少。

本文主要受 Kancharagunta 等^[26-27]启发,对 CycleGAN 进行改进,增加生成-重建计算通路,引入生成重建损失函数,并增加整体损失函数,从而减少图像翻译中的错误映射和无意义的信息添加,提升了在短波红外人脸图像翻译上的实验效果。对于图像生成结果,结合直观判断与 FID(Fréchet inception distance)进行检验,并引入人脸识别作为间接验证,综合评价了图像翻译的效果。实验结果表明,所提框架的图像翻译结果在识别准确率上提升了 26.3%。

2 CycleGAN 图像翻译框架

CycleGAN 是一种无监督的图像翻译框架。它由两个 GAN 构成对偶结构,用于两个图像域之间的图像翻译;引入循环结构,用判别器来保证翻译结果目标域特性的准确;引入循环一致性损失函数,用以保证翻译结果与输入图像目标的一致性。CycleGAN 在未配对的数据集上获得了较好的图像翻译结果。

2.1 CycleGAN 模型结构

CycleGAN 模型由两个生成器和两个判别器构成。以短波红外-可见光图像的翻译问题为例,将短波红外人脸和可见光人脸图像分别定义为域 S 和域 V 。一个生成器 $G_{SV}:S \rightarrow V$ 用于将短波红外图像 s 翻译为可见光图像 v_{syn} ,用对应的判别器 D_V 判别它接收的输入图像是来自真正的可见光域的图像 v 还是由生成器生成的伪可见光图像 v_{syn} 。另一个生成器 $G_{VS}:V \rightarrow S$ 用于将输入的可见光图像 v 翻译为短波红外图像 s_{syn} ,用对应的判别器 D_S 判断输入图像是真实的短波红外图像还是生成的伪短波红外图像。训练时,将两个生成器的输出分别作为另一个生成器的输入,形成循环结构,得到循环重建短波红外人脸图像 s_{rec} 和循环重建可见光人脸图像 v_{rec} 。在训练中,根据判别器输出和输入图像与循环重建图像的差值计算损失函数。具体结构如图 1 所示, s_{real} 、 v_{real} 为输入的短波红外、可见光图像, v_{syn} 、 s_{syn} 为生成的可见光与短波红外图像, s_{rec} 、 v_{rec} 为循环重建图像, \mathcal{L}_{cyc} 为循环一致性损失函数。生成器的目标是尽可能生成接近真实的对应域图像,而判别器的目标是尽可能区分生成器生成的图像与真实图像。

生成器与判别器的结构如图 2 所示,Conv 为卷

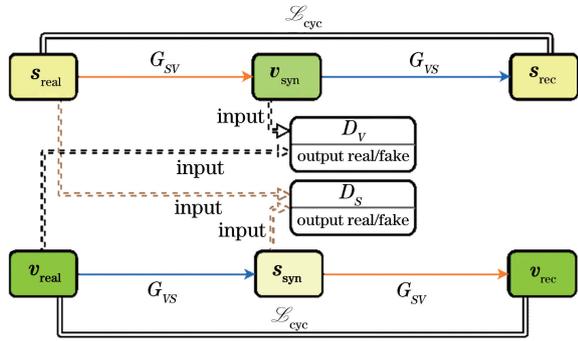


图1 CycleGAN 模型架构
Fig. 1 Structure of CycleGAN model

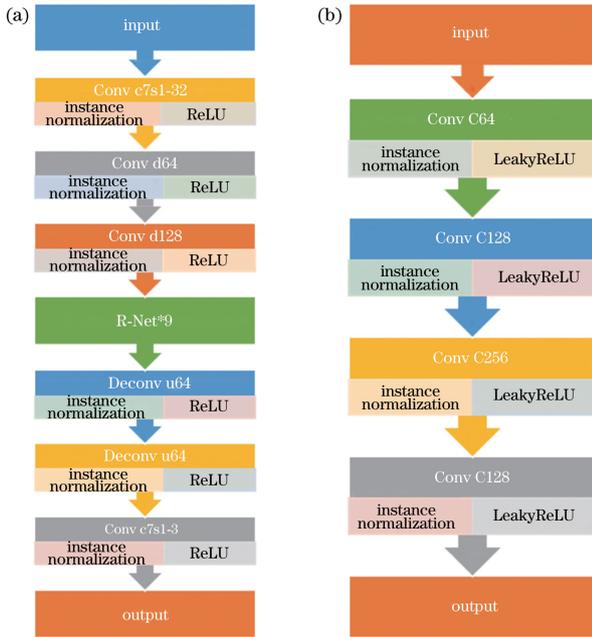


图2 生成器与判别器结构。(a)生成器;(b)判别器
Fig. 2 Structure of generator and discriminator.

(a) Generator; (b) discriminator

积层, instance normalization 表示实例正则化, ReLU 和 LeakyReLU 为激活函数, Deconv 为反卷积层。生成器由 1 个卷积核为 7×7 的卷积层 (stride 为 1)、2 个卷积核为 3×3 的卷积层 (stride 为 2)、2 个卷积核为 3×3 的反卷积层 (stride 为 0.5)、9 个残差层构成, 每个卷积层的输出会进行实例正则化操作。判别器采用 70×70 PatchGAN 结构, 由 4 个卷积核为 4×4 的卷积层 (stride 为 2) 构成, 最后由一个卷积层输出结果。

2.2 损失函数与目标函数

CycleGAN 的损失函数由生成器与判别器的对抗损失函数和循环一致性损失函数构成。生成对抗损失函数由判别器对真实样本的概率估计和判别器对生成样本的概率估计构成。在实际训练中, 采用的 LSGAN 的损失函数为

$$\mathcal{L}_{G_{SV}}(G_{SV}, D_V) = E_{s \sim P_S(s)} \{ \{ D_V[G_{SV}(s)] - 1 \}^2 \} + \frac{1}{2} \{ E_{v \sim P_V(v)} \{ [D_V(v) - 1]^2 \} + E_{s \sim P_S(s)} \{ D_V[G_{SV}(s)]^2 \} \}, \quad (1)$$

式中: $E_{s \sim P_S(s)}$ 为输入短波红外图像时的期望; $P_S(s)$ 为短波红外图像分布; $G_{SV}(s)$ 为生成器生成的可见光图像; $E_{v \sim P_V(v)}$ 为输入可见光图像时的期望; $P_V(v)$ 为可见光图像分布; 判别器 $D_V(\cdot)$ 会输出 $0 \sim 1$ 的数, 表征判断的输入图像是真实可见光图像的概率, 输入图像越接近真实图像, 输出值越接近 1。训练中, 优化生成器, 使得生成器生成的图像接近真实图像, 会使 (1) 式中的第一项趋向最小值。当输入为生成器生成的图像时, 判别器 $D_V(\cdot)$ 应当输出接近 0 的值, 优化判别器, 使得判别器在对应情况输出正确的结果, (1) 式中的第二项会趋向于最小值。实际训练中, 生成器与判别器交替优化。另一组生成器、判别器的构成与上述类似, 将 (1) 式中的生成器和判别器替换为 G_{VS} 网络中的生成器 G_{SV} 和判别器 D_S , 并将输入的短波红外图像 s 和可见光图像 v 位置互换, 可以得到 G_{VS} 网络的损失函数 $\mathcal{L}_{G_{SV}}(G_{SV}, D_V)$ 。

在图像翻译应用中, 传统 GAN 会出现翻译结果只与训练集中部分图像相似的情况, CycleGAN 通过引入循环一致性损失函数, 要求输入图像在依次通过两个生成器后得到输入图像本身, 以学习正确的映射关系。它通过计算循环生成的图片与该通道输入图片的 L_1 范数距离获得:

$$\mathcal{L}_{cyc}(G_{SV}, G_{VS}) = E_{s \sim P_S(s)} \{ \| G_{VS}[G_{SV}(s)] - s \|_1 \} + E_{v \sim P_V(v)} \{ \| G_{SV}[G_{VS}(v)] - v \|_1 \}, \quad (2)$$

式中: $\| \cdot \|_1$ 为 L_1 范数。

2.3 经典 CycleGAN 在图像翻译上的不足

CycleGAN 在短波红外-可见光人脸翻译问题上有很大的不足。第一, 该框架是无监督的, 模型会有一定概率找到错误的映射关系。虽然 CycleGAN 模型引入了循环一致性损失函数, 但这一函数是基于输入与重建图像计算的。在训练过程中, 对于同一路径上的两个生成器, 可能出现前一生成器学习了错误的映射, 第二个生成器学习了另一个错误映射以纠正前一生成器的错误映射, 导致循环一致性损失函数较小而实际学习了错误映射的情况。例如, 在图像-标签的翻译过程中, 建筑物会被标注为树木。在本问题中, 映射关系出现另一种偏差, 即将“器官”与“位置”对应, 会在“合理”位置生成与输入图像形状相似性较差的“正确”器官, 例如,

对头发与皮肤的分界线判断错误,在头发位置错误生成了皮肤,影响观察与识别。仅依靠现有的循环一致性损失函数难以解决这一问题。第二,CycleGAN对纹理变化翻译较强,对形状、位置变化的处理能力较弱。本问题中因图像来源的差异,对应图像之间存在设备、角度等原因导致的位置变化,CycleGAN会在生成图像中引入意义不明的色块以及额外的高频波纹等噪声信息,极大影响进一步的目标识别。

CycleGAN中还提出了可选的识别损失函数 $\mathcal{L}_{identity}$ 来增强图像翻译色调的准确性,对生成器输入该生成器目标域的图像,生成器的输出应与输入一致。在训练中对每个生成器均输入其目标域图像,并计算它的输出图像与输入图像的差别(L_1 距离),再将两个生成器的计算结果相加作为模型的识别损失函数。

$$\mathcal{L}_{identity}(G_{SV}, G_{VS}) = E_{s \sim P_S(s)} \| G_{VS}(s) - s \|_1 + E_{v \sim P_V(v)} \| G_{SV}(v) - v \|_1 \quad (3)$$

在实验中,加入识别损失函数后,目标的识别率有一定提升,但生成器的FID也显著提升,图像翻译质量受到影响,如表1所示,其中all表示总体识别,frontal表示正面识别。

表1 识别损失函数在Facenet上的结果

Table 1 Results of identity loss function on Facenet

Model	FID	Accuracy (all)	Accuracy (frontal)
CycleGAN	48.2173	0.137	0.146
CycleGAN+ $\mathcal{L}_{identity}$	65.0052	0.220	0.267

3 改进的 CycleGAN 结构

3.1 改进的 CycleGAN 框架结构

由于结构本身的原因,CycleGAN模型在训练中可能学习到错误的映射,导致人脸图像翻译结果一般,可辨认性较差。同时,在短波红外-可见光图像翻译任务中,短波红外图像与可见光图像由两台不同的相机采集,它们的焦距、视场、角度存在一定差异,采集的图像略有不同,同一目标同一时刻的两张图像存在一定的差异,此时仅依靠两条相对独立的训练通路难以完成较好的图像翻译。因此,本文首先引入整体损失函数,强化了对循环一致性的约束,增强整体结构信息,减少非相关信息的添加;其次引入新的损失函数计算通道,使得两条训练路径之间有更多交换信息,强化了对生成器的监督,从而减少错误映射情况,增强网络的学习能力,在位置变化时网络具有了一定的稳定性,也从另一方面减少了翻译结果中的波纹等噪声信息。

整个框架的结构如图3所示,在CycleGAN两个生成-判别通道的基础上,增加了生成重建损失函数 \mathcal{L}_{sr} 的计算通路和整体损失函数 $\mathcal{L}_{intensity}$,每次从两个域中随机选择同一目标的角度、表情等接近的两张图像作为输入图像。

3.2 增加基于图像主要信息的整体损失函数

在使用CycleGAN进行人脸图像翻译时,更关注由短波红外向可见光的图像翻译。目前有很多研究可以证明循环一致性损失函数对图像翻译有较好效果,可以减少学习错误映射的概率,基于这一思想,引入整体损失函数作为加强的约束条件,加强对

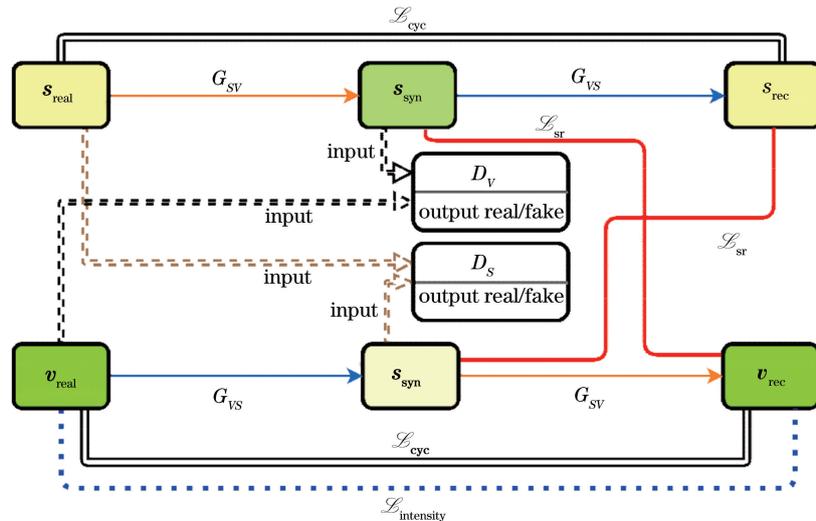


图3 改进的 CycleGAN 框架结构

Fig. 3 Structure of improved CycleGAN framework

循环一致性的约束。强化框架可保留输入图像整体特征,这有利于获得更好的翻译效果。

短波红外图像为单通道,可以视为亮度信息,而对于可见光图像,YCbCr 空间中的 Y 通道是它的亮度通道,保留了大部分的图像细节信息^[31]。两者类比,将生成图像与输入图像的 Y 通道分量间的距离作为整体损失函数,可以有效提升生成器对输入图像结构的学习能力,图像翻译结果更好。因此,对 $v \rightarrow G_{VS}(v) \rightarrow G_{SV}[G_{VS}(v)]$ 这一路径,增加了基于 Y 通道、类似于循环一致性损失函数形式的 L_1 距离以作为整体损失函数,即

$$\mathcal{L}_{intensity}(G_{SV}, G_{VS}) =$$

$$E_{v \sim P_V(v)} \|Y(v) - Y\{G_{SV}[G_{VS}(v)]\}\|_1, \quad (4)$$

式中: $Y(\cdot)$ 为计算图像 Y 通道的函数。

这一路径的首末均为可见光图像,除了结构信息外还有颜色信息,为了保留颜色信息的约束,通过增加整体损失函数来强化结构信息,保留原有的循环一致性损失函数以获得颜色、纹理上的约束。在后续的实验中可以看见,增加整体损失函数后,翻译图像的机器识别率获得了一定提高。

3.3 增加基于生成重建通道的损失函数

针对 CycleGAN 中同一路径上两个生成器可能都学习错误映射的情况,在两条训练路径之间增加损失函数计算通路,将生成器生成结果与另一路径重建结果进行比较,以此来对生成器进行进一步约束,以更好地利用 GAN 的学习能力减少错误映射情况的发生和无关信息的生成。新损失函数计算通路的引入也将不同图像来源的影响加入到生成器的优化目标中,使得网络在输入图像位置变化时具有一定的稳定性,减少了干扰信息。

生成的可见光图像 v_{syn} 与循环重建的可见光图像 v_{rec} , 生成的短波红外图像 s_{syn} 与循环重建的短波红外图像 s_{rec} , 都是同一目标在同一图像域的图像,在理想情况下它们应当保持一致。本文引入的生成重建损失函数 \mathcal{L}_{sr} 计算了生成的可见光图像 v_{syn} 与循环重建的可见光图像 v_{rec} 的距离及生成的短波红外图像 s_{syn} 与循环重建的短波红外图像 s_{rec} 的距离。短波红外图像为单通道亮度信息,因此生成的短波红外图像与循环重建的短波红外图像间直接采用 L_1 距离。对于可见光图像,与 3.2 小节分析类似,此处两图像 Y 通道之间的距离为

$$\mathcal{L}_{sr}(G_{SV}, G_{VS}) =$$

$$\|Y[G_{SV}(s)] - Y\{G_{SV}[G_{VS}(v)]\}\|_1 + \|G_{VS}(v) - G_{VS}[G_{SV}(s)]\|_1. \quad (5)$$

实验中发现,增加生成重建损失函数后,在图像翻译的结果上,无意义噪声信息和错误映射情况减少,框架的翻译效果与识别率均有所提升,当权重在 0~1 时,提升效果较为明显。

3.4 基于新损失函数的目标函数

整个框架的损失函数为 CycleGAN 损失函数、整体损失函数及生成重建损失函数的和,即

$$\mathcal{L}(G_{SV}, D_V, G_{VS}, D_S) =$$

$$\mathcal{L}_{G_{SV}}(G_{SV}, D_V) + \mathcal{L}_{G_{VS}}(G_{VS}, D_S) +$$

$$\alpha_1 \cdot \mathcal{L}_{cyc}(G_{SV}, G_{VS}) + \alpha_2 \cdot$$

$$\mathcal{L}_{intensity}(G_{SV}, G_{VS}) + \alpha_3 \cdot \mathcal{L}_{sr}(G_{SV}, G_{VS}), \quad (6)$$

式中: $\alpha_1, \alpha_2, \alpha_3$ 为各损失函数的权重。调整参数 α_1 可调节循环一致性损失函数的权重,调整参数 α_2 可调节整体损失函数的权重,调整参数 α_3 可调节生成重建损失函数的权重。在本文中,对 α_1 采用 CycleGAN 论文^[20]中的设置,调整 α_2 和 α_3 以寻求较优结果。

4 实验与分析

4.1 数据集建立

根据调研,目前还没有公开可获得的短波红外人脸图像数据集,本文的实验基于自建的短波红外-可见光数据集。自建数据集采用短波相机与可见光相机进行数据采集,并进行了前期的动态转换与图像切割。

数据采集设备为一台短波相机与一台单反相机。短波相机为 Xenics 公司的 Bobcat320,是一款采用 InGaAs 探测器的短波相机,光谱范围为 0.9~1.7 μm ,分辨率为 320×256 ,像素间距为 20 μm ,可以输出 14 bit 图像,有图片拍摄模式和录像模式可以选择。可见光数据的采集采用尼康的 D5600 单反相机,采用录像模式,可得到分辨率为 1920×1080 ,帧率为 50 frame/s 的视频。

总共采集 207 个目标数据,其中男性目标 123,女性目标 84。采集环境为半受控环境,即目标位置固定,但动作与角度多样,得到含有多种姿态、多种表情的图像。数据采集时采用视频模式,对每个姿态都采集有一定差别的多幅图像。去除质量较差图像后,每个目标有 100~200 张有效图像,再剔除部分差别较小图像,实际可用数据为每个目标 80 张左右。原始数据含有较多背景干扰,需进行裁剪和标注以获得目标区域。此次实验中,先随机选择不同性别目标各 16 个作为验证集,同时从测试集中剔除偏转过大的图像,构成“正面测试集”,以提供进一步参考。

4.2 评价标准

对于 GAN 生成的图像质量,评价方法可以分为主观评价方法和客观评价方法两类。主观评价是实验人员自身或者第三方测试者根据对生成图像的主观感受来进行评价。客观评价方法是通过选择现有指标或设计合理模型来量化评估生成图像的质量。

4.2.1 FID

选取 FID 作为图像评价指标,对翻译的图像进行质量评估。常见的图像评价指标,如 PSNR、SSIM 都是逐像素计算的,它们对于去噪、超分辨等图像像素级对齐的情况有效,而对于生成图像与参考图像未作精确配准,FID 可以更好地进行评估。FID 由 Heusel 等^[34]提出,它使用 Inception Network 神经网络的中间层提取图像特征,然后利用正态分布去模拟特征的分布,通过计算生成图像的正态分布与参考图像的正态分布的 Frechet 距离来表征两个分布的差别。FID 具有良好的判别能力,是 GAN 的优秀判别指标。FID 越小,说明生成对象的特征分布与目标特征分布越接近,生成器效果越好。

$$FID(x, g) = \|\mu_x - \mu_g\|_2^2 + \text{Tr}[\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}], \quad (7)$$

式中: $\|\cdot\|_2^2$ 为 L_2 范数的平方; $\text{Tr}[\cdot]$ 为矩阵的迹; x 和 g 分别为真实图像与生成图像; μ_x 为真实图像特征的均值; μ_g 为生成图像特征的均值; Σ_x 和 Σ_g 分别为真实图像和生成图像特征的协方差矩阵。

4.2.2 人脸识别准确率

需要注意的是,FID 是基于特征提取的评价方式,评价的是生成样本的丰富性,是对模型生成图像质量在特征层面的一种评估。它对于翻译图像结构的敏感性还未作探究,可能存在生成的图像具有类似特征但结构布局并不合理,却具有更低 FID 的情况。

为了判断图像翻译框架对人脸识别特征的还原程度,使用 Facenet 人脸识别框架对生成结果进行检测,利用识别准确率间接判断生成质量。Facenet 是由 Schroff 等^[35]提出的基于深度神经网络的人脸识别模型,它在多个公共数据集上取得较好结果。本文采用在 LFW 数据集上预训练好的准确率为 0.9965 的模型,对测试集上的真实可见光图片进行特征抽取建模后再进行人脸识别检测。

考虑到训练集中正面脸部图像所占比例较大,侧面图像尤其是大角度偏转的样本较少,除了在整个测试集进行测试外,还去除测试集中偏转超过

20°的样本,构成一个“正面测试集”,在该测试集上再次进行检测,得到正面识别率以作为另一个评估指标。

4.3 实验分析

4.3.1 预设参数

训练中固定生成器和判别器一方的权重,更新另一方,在交替中优化自身网络。判别器训练中不采用最新生成的图片,而是采用之前生成的 50 张图片。在训练中利用 Adam 进行优化求解,将 batch_size 设置为 1,在前 100 个 epoch 中学习率固定为 0.0002,采用较大的搜索空间,保证网络的学习能力,在之后的 100 个 epoch 中学习率线性衰减到 0,减少训练中的振荡情况,对权重进行更精细的调整,以保证训练的稳定。

4.3.2 参数调优

对原 CycleGAN 部分的超参数采用预设参数,将(6)式中的 α_1 设为 10,之后分别在原网络上添加整体损失函数与生成重建损失函数进行实验,对权重参数进行测试,再整体进行调优,以获得较好的实验效果。

4.3.3 整体损失函数的参数选择

实验中测试了仅增加整体损失函数的情况。首先选择一个较大范围内的权重进行测试,并对较优区间进行精细搜索,以获取最优结果。表 2 中列出了权重分别设置为 1、5、10、12.5、15、20、50 时模型在同一测试集上的表现,验证了损失函数的作用并获取了期望的权重选择方案。

从表 2 可知, α_2 为 1~20 时,整体损失函数对 FID 略有影响,但是影响不大;当 α_2 为 10~50 时,翻译结果的识别准确率相比未增加该损失函数的情况有所提升,在 10~15 范围内准确率有较大值。同时,FID 均保持相对稳定,可以认为生成结果的多样性未受影响,生成器保持了较好的泛化能力。

表 2 整体损失函数在 Facenet 上的指标分析

α_2	FID	Accuracy (all)	Accuracy (frontal)
0	48.2173	0.137	0.146
1.0	52.9253	0.121	0.119
5.0	41.6498	0.133	0.146
10.0	41.3721	0.152	0.179
12.5	43.4405	0.144	0.165
15.0	43.1670	0.161	0.175
20.0	38.4180	0.142	0.167
50.0	65.9726	0.145	0.154

4.3.4 生成重建损失函数的参数选择

实验中测试了仅增加生成重建损失函数进行优化的情况。表 3 列出了在权重 α_3 为 0.05、0.50、1、5、50 的情况下进行训练测试的结果,验证了该损失函数的效果以及较优的权重值选择方案。

从表 3 可知,在加入生成重建损失函数后,在 $\alpha_3 < 1$ 时,FID 随 α_3 增加稍稍增大,而人脸识别准确率明显提升;在权重为 0.50 时,准确率几乎是未增加该损失函数情况的 2 倍。生成重建损失函数强化了框架对目标识别特征的学习能力,减少了错误映射的情况,有效提高了翻译图像的识别率。

表 3 生成重建损失函数在 Facenet 上的指标分析
Table 3 Index analysis of synthesis-reconstruction loss function on Facenet

α_3	FID	Accuracy (all)	Accuracy (frontal)
0	48.2173	0.137	0.146
0.05	41.4332	0.168	0.214
0.50	52.5602	0.252	0.311
1.00	58.1832	0.234	0.294
5.00	87.7636	0.216	0.243
50.00	105.1097	0.034	0.029

4.3.5 联合权重选择

结合单独训练的情况,选择两者的最优权重与次优权重进行组合,并在一定范围内进行微调,以获取最优的组合结果。测试了多种权重组合的情况,表 4 为其中的一些组合,最后在 $\alpha_2 = 10, \alpha_3 = 0.50$ 时获得了较好的结果。在此参数设置下,FID 与最小值相比相差不大,而识别准确率有很大提升。

表 4 所提框架在不同参数下的结果

Table 4 Results of proposed framework with different parameters

α_2	α_3	FID	Accuracy (all)	Accuracy (frontal)
0	0	48.2173	0.137	0.146
10	0.05	43.7915	0.234	0.288
15	0.05	39.2064	0.200	0.216
8	0.50	48.6155	0.297	0.370
10	0.40	43.9396	0.260	0.325
10	0.50	47.3351	0.334	0.409
11	0.50	49.9668	0.290	0.393
15	0.50	48.2815	0.271	0.354

4.4 结果对比

4.4.1 对比框架

除了对比基础的 CycleGAN 外,还对比了 DualGAN、UNIT、MUNIT、DRIT 等图像翻译框架。

DualGAN:一种利用 WassersteinGAN 替代原始 GAN 中的交叉熵,采用对偶式 GAN 结构进行图像翻译的框架。

UNIT:一种基于共享 latent space 假设和循环损失的无监督图像翻译框架。

MUNIT:对图像内容与风格进行 latent space 共享,将输入图像的内容与从目标域抽取的风格表示进行组合,得到目标域图像。

DRIT:将 latent space 分为与图像域无关的内容空间和与图像域有关的属性空间,可以将输入图像翻译为多个输出域的图像。

在同一训练集上训练了这些网络及所提模型,参考模型均采用论文中的参数设置,其中 MUNIT 和 DRIT 将输出模式设置为 1,即仅输出一种风格的图像,之后在同一测试集上对这些框架进行测试。

4.4.2 主观分析

不同框架的图像翻译结果如图 4 所示。从图 4 可以看到:MUNIT 和 DualGAN 存在过拟合现象,“记忆”了部分图像,将已有图像“套用”到了测试集上;CycleGAN 头发与脸部皮肤分界不清晰,会有错误的颜色纹理情况;UNIT 的部分图像整体轮廓出现了较大改变,但在有偏转的情况下结果较优;DRIT 的结果接近于对图片进行了模糊处理的结果,整体不够清晰,部分位置出现无意义波纹。

所提框架在进行图像翻译后,脸部轮廓与参考图像基本接近,未出现变形情况。所提框架可以较为准确地区分头发与脸部皮肤,部分框架仅保留部分头发的纹理、形状,但颜色与脸部肤色相同。在所提框架生成的脸部图像中,额外添加的色块、纹理相对较少,各器官基本准确还原了参考图像。

4.4.3 量化分析

量化比较结果如表 5 所示,所提框架生成的图像的 FID 相较于其他框架更低,与 CycleGAN 和 DualGAN 接近,可以认为翻译结果的质量较好。同时,所提框架在 Facenet 上的人脸识别准确率优于其他框架,总体识别准确率与正面识别准确率对比框架中表现最好的 DRIT 高 0.115 左右,与 CycleGAN 相比分别提高了 0.197 和 0.263。所提改进框架有效提高了人脸图像翻译过程中的识别准确率。

在实验中发现,当目标存在一定偏转(即使角度不大)或者有眼镜干扰时,所提框架虽然能在一定程度上翻译图像且效果优于多数的对比框架,但失真很明显。主要原因是训练样本是以正面图像为主,

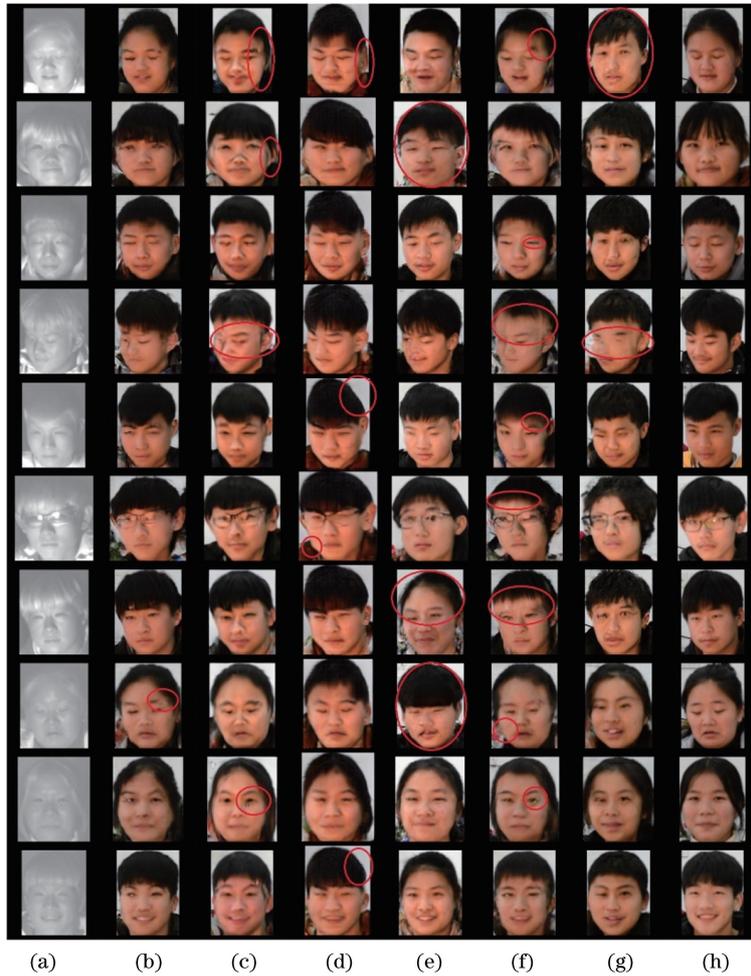


图 4 不同框架生成的图像比较。(a) 短波图像；(b) 所提框架；(c) DRIT；(d) UNIT；(e) MUNIT；(f) CycleGAN；(g) DualGAN；(h) 参考图像

Fig. 4 Comparison of generated images by different frameworks. (a) Short-wave images; (b) proposed framework; (c) DRIT; (d) UNIT; (e) MUNIT; (f) CycleGAN; (g) DualGAN; (h) reference images

表 5 不同框架翻译结果的指标分析

Table 5 Index analysis of results generated by different frameworks

Framework	FID	Accuracy (all)	Accuracy (frontal)
Real image	16.7590	0.832	
Proposed framework	47.3351	0.334	0.409
CycleGAN	48.2173	0.137	0.146
DualGAN	46.7556	0.078	0.086
UNIT	50.9268	0.166	0.214
MUNIT	69.1422	0.112	0.129
DRIT	52.6906	0.222	0.290

偏转角度偏少；同时多数目标无眼镜遮挡，遮挡眼镜的款式、形状样本不足。“正面测试集”的识别准确率均优于总测试集的表现。

5 结 论

提出用于短波红外-可见光人脸图像翻译的改进 CycleGAN 框架,利用深度学习学习图像的跨波段映射关系。对原网络增加损失函数计算通路,引入生成重建损失函数与整体损失函数,在没有过多影响训练速度的情况下,减少图像翻译结果中的错误信息,改善了错误映射的情况,在保持目标框架的同时,减少了轮廓畸变和异常纹理。在现有人脸识别框架下,总体、正面人脸识别准确率分别提升了 19.7% 和 26.3%。接下来的工作将围绕多姿态识别和模型简化展开,以增强实际复杂场景下的性能,进一步提高识别准确率。

参 考 文 献

[1] Dutta A K. Imaging beyond human vision [C]//8th

- International Conference on Electrical and Computer Engineering, December 20-22, 2014, Dhaka, Bangladesh. New York: IEEE, 2014: 224-229.
- [2] Cao Z C, Schmid N A, Bourlai T. Composite multilobe descriptors for cross-spectral recognition of full and partial face[J]. *Optical Engineering*, 2016, 55(8): 083107.
- [3] Cai Y, Hu X. Short wave infrared imaging technology and its defence application[J]. *Infrared and Laser Engineering*, 2006, 35(6): 643-647.
蔡毅, 胡旭. 短波红外成像技术及其军事应用[J]. *红外与激光工程*, 2006, 35(6): 643-647.
- [4] Kalka N D, Bourlai T, Cukic B, et al. Cross-spectral face recognition in heterogeneous environments: a case study on matching visible to short-wave infrared imagery[C]//2011 International Joint Conference on Biometrics (IJCB), October 11-13, 2011, Washington, DC, USA. New York: IEEE, 2011: 12494407.
- [5] Hu S W, Short N, Riggan B S, et al. Heterogeneous face recognition: recent advances in infrared-to-visible matching [C]//2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), May 30-June 3, 2017, Washington, DC, USA. New York: IEEE, 2017: 883-890.
- [6] Shao B T, Tang X Y, Jin L, et al. Single frame infrared image super-resolution algorithm based on generative adversarial nets[J]. *Journal of Infrared and Millimeter Waves*, 2018, 37(4): 427-432.
邵保泰, 汤心溢, 金璐, 等. 基于生成对抗网络的单帧红外图像超分辨率算法[J]. *红外与毫米波学报*, 2018, 37(4): 427-432.
- [7] Zhang R, Isola P, Efros A A. Colorful image colorization[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9907: 649-666.
- [8] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: feature learning by inpainting [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 2536-2544.
- [9] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 2414-2423.
- [10] Laffont P Y, Ren Z L, Tao X F, et al. Transient attributes for high-level understanding and editing of outdoor scenes[J]. *ACM Transactions on Graphics*, 2014, 33(4): 149.
- [11] Wu H M, Liu Q R, Wang Y H. Face image translation based on generative adversarial networks [J]. *Journal of Tianjin University (Science and Technology)*, 2019, 52(3): 306-314.
吴华明, 刘茜瑞, 王耀宏. 基于生成对抗网络的人脸图像翻译[J]. *天津大学学报(自然科学与工程技术版)*, 2019, 52(3): 306-314.
- [12] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3431-3440.
- [13] Mirza M, Osindero S. Conditional generative adversarial nets[J/OL]. (2014-11-06) [2019-09-22]. <https://arxiv.xileisou.top/abs/1411.1784>.
- [14] Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11207: 179-196.
- [15] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, December 8-13, 2014, Montreal, Quebec, Canada. New York: Curran Associates, 2014: 2672-2680.
- [16] Xie J R, Li F M, Wei H, et al. Infrared target simulation method based on generative adversarial neural networks[J]. *Acta Optica Sinica*, 2019, 39(3): 0311002.
谢江荣, 李范鸣, 卫红, 等. 基于生成对抗式神经网络的红外目标仿真方法[J]. *光学学报*, 2019, 39(3): 0311002.
- [17] Liu K, Wang D, Rong M X. X-ray image classification algorithm based on semi-supervised generative adversarial networks [J]. *Acta Optica Sinica*, 2019, 39(8): 0810003.
刘坤, 王典, 荣梦学. 基于半监督生成对抗网络 X 光图像分类算法 [J]. *光学学报*, 2019, 39(8): 0810003.
- [18] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5967-5976.
- [19] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [M]//Navab N, Hornegger J, Wells W, et al. *Medical image computing and computer-*

- assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [20] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2242-2251.
- [21] Yi Z L, Zhang H, Tan P, et al. DualGAN: unsupervised dual learning for image-to-image translation [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2868-2876.
- [22] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks [J]. Proceedings of the 34th International Conference on Machine Learning, 2017, 70: 1857-1865.
- [23] Liu M Y, Breuel T, Kautz J. Unsupervised image-to-image translation networks [C]//Advances in Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. New York: Curran Associates, 2017: 700-708.
- [24] Lee H Y, Tseng H Y, Huang J B, et al. Diverse image-to-image translation via disentangled representations [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11205: 36-52.
- [25] Zhang W, Shu Z, Samaras D, et al. Improving heterogeneous face recognition with conditional adversarial networks [J/OL]. (2017-09-13) [2019-09-22]. <https://arxiv.xilesou.top/abs/1709.02848>.
- [26] Kancharagunta K B, Dubey S R. CSGAN: cyclic-synthesized generative adversarial networks for image-to-image transformation [J/OL]. (2019-01-11) [2019-09-22]. <https://arxiv.xilesou.top/abs/1901.03554>.
- [27] Song L, Zhang M, Wu X, et al. Adversarial discriminative heterogeneous face recognition [C]//Thirty-Second AAAI Conference on Artificial Intelligence, February 2-7, 2018, New Orleans, Louisiana, USA. Menlo Park: AAAI, 2018: 7355-7362.
- [28] Wang Z L, Chen Z Z, Wu F. Thermal to visible facial image translation using generative adversarial networks [J]. IEEE Signal Processing Letters, 2018, 25(8): 1161-1165.
- [29] Riggan B S, Short N J, Hu S W. Thermal to visible synthesis of face images using multiple regions [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE, 2018: 30-38.
- [30] Zhang T, Wiliem A, Yang S Q, et al. TV-GAN: generative adversarial network based thermal to visible face recognition [C]//2018 International Conference on Biometrics (ICB), February 20-23, 2018, Gold Coast, QLD, Australia. New York: IEEE, 2018: 174-181.
- [31] Lezama J, Qiu Q, Sapiro G. Not afraid of the dark: NIR-VIS face recognition via cross-spectral hallucination and low-rank embedding [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6807-6816.
- [32] Ghiass R S, Arandjelović O, Bendada H, et al. Infrared face recognition: a literature review [C]//The 2013 International Joint Conference on Neural Networks (IJCNN), August 4-9, 2013, Dallas, TX, USA. New York: IEEE, 2013: 14027413.
- [33] Sun L, Zheng Z W. Thermal-to-visible face alignment on edge map [J]. IEEE Access, 2017, 5: 11215-11227.
- [34] 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, December 4-9, 2017, Long Beach, CA, USA. New York: Curran Associates, 2017: 6626-6637.
- [35] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 815-823.