

DOI:10.11784/tdxbz201801034

基于生成对抗网络的人脸图像翻译

吴华明, 刘茜瑞, 王耀宏

(天津大学数学学院, 天津 300072)

摘要: 针对人脸照片和人脸素描间的图像翻译问题, 本文基于对偶生成对抗网络模型, 对其目标函数附加两个损失函数建立新的网络模型. 通过参数优化实验不断优化本文提出的模型, 从而找到最优参数; 通过直观和量化对比实验表明本文提出的模型在人脸数据上的图像翻译效果无论在清晰度还是在保持面部特征方面是目前基于生成对抗网络的图像翻译模型中表现最优的, 并对相关 GAN 模型的稳定性进行了对比; 最后通过效果分析实验说明了所附加的损失函数的具体作用.

关键词: 生成对抗网络; 人脸数据; 图像翻译; 损失函数

中图分类号: TP391.4

文献标志码: A

文章编号: 0493-2137(2019)03-0306-09

Face Image Translation Based on Generative Adversarial Networks

Wu Huaming, Liu Qianrui, Wang Yaohong

(School of Mathematics, Tianjin University, Tianjin 300072, China)

Abstract: With regard to the problem of image translation between face photo and face sketches, a new network model was established by adding two loss functions to the objective function of the DualGAN. Through optimization experiments of the parameters, the proposed model was continuously optimized to find the optimal parameters. The qualitative and quantitative comparison experiments show that the proposed model has excellent translation performance in face data in terms of sharpness and facial features, and it is now the best among the related GAN network models. The stability of related GAN models was then compared. Finally, the effect analysis experiment clarified the specific function of the additional loss functions.

Keywords: generative adversarial networks; face data; image translation; loss functions

图像翻译是图像处理的一个重要研究领域, 图像翻译就是把两个具有不同特征的图片域进行相应转换, 包括不同域间的风格变换、物体变形、季节转换、图像增强等. 按照传统方法, 这些任务都是根据不同风格模式的图像间的内在差异分开进行处理的. 在过去的几年里, 通用的端到端的深度学习框架, 最显著的是利用全卷积网络^[1](FCNs)和条件生成对抗网络^[2](CGANs)推动了图像翻译的发展, 使得多种图像翻译问题可以得到统一的处理. 经历了针对特定领域单任务的图像翻译到多领域多任务图像翻译的转变, 从需要带标签配对的图像集到只需无标签非配对的图像集的转变, 图像翻译过程越来越简单, 功能却

越来越强大. 本文主要研究生成对抗网络(GAN)在图像翻译方面的相关模型及其在人脸数据上的应用.

Goodfellow 等^[3]在 2014 年提出 GAN 模型, GAN 模型在训练的过程中使用简单的反向传播和 dropout 算法, 不需要复杂马尔科夫链等近似推理的过程, 使用一种分布直接进行采样, 不需要预先建模, 这在一定程度上简化了训练过程. 但这种不需要建模的缺点是太过自由, 对于较大图片、较多像素的情形, 基于简单的 GAN 的方式不太可控, 生成数据与原始数据差异较大. Mirza 等^[2]提出了 CGAN 模型, 这是一个带有条件约束的 GAN 模型, 在生成器 G 和判别器 D 的建模中均引入条件变量 y , 使用额

收稿日期: 2018-01-03; 修回日期: 2018-04-28.

作者简介: 吴华明(1986—), 男, 博士, 讲师, whming@tju.edu.cn.

通信作者: 刘茜瑞, liuxirui57@tju.edu.cn.

基金项目: 国家自然科学基金资助项目(11601381).

Supported by the National Natural Science Foundation of China(No. 11601381).

外信息 y 对模型增加条件,以指导数据生成过程,随着额外信息的加入,CGAN 把纯无监督的 GAN 变成有监督的模型改善了实验效果,但实验结果仍存在不稳定、模糊、斑点等各类问题. Isola 等^[4]提出 Pix2Pix 模型,该模型是在 CGAN 目标损失函数的基础上加了一个 L1 损失,使用 U-NET 结构^[5]来作为生成器,使用一种 PatchGAN 的结构^[6]作为判别器,在图像翻译方面与 CGAN 的结果相比, Pix2Pix 可以让图像结果更加锐利更加清晰,使得输出更加接近真实图像.

由于进行有监督的学习和获得大量的带标签的配对数据可能很费时间,耗费较大的财力物力,有时甚至是不可能的,例如将日光场景转换为夜景,即使用固定摄像机,由于场景中经常有移动的物体,造成所得的配对图像有不同程度的内容差异. 随后出现了利用非配对无标签数据进行非监督学习的模型 CycleGAN、DualGAN 和 DiscoGAN. Zhu 等^[7]受循环一致思想^[8]的启发提出了 CycleGAN 模型,解决了利用非配对无标签数据解决图像间的风格转移问题. Kim 等^[9]为确保图像在不同域转化时保持图像的某些特征,比如方向角、面部特征等,提出了 DiscoGAN. Yi 等^[10]受原始的自然语言处理对偶学习方法^[11]的启发,提出 DualGAN 模型,解决如何利用无标签非配对的数据进行具有不同特点的两个域之间的图像翻译问题. 这 3 个非监督学习模型的网络模型结构很相似,都是由 2 个 GAN 结构^[12]组成. 在目标函数、生成器、判别器的构成上略有不同. 3 个模型在各自数据集上都取得了不错的实验效果,在人脸数据上的图像翻译效果在清晰度和保持面部特征方面仍有待提高.

本文针对在人脸数据上的图像翻译问题,在 DualGAN 模型的基础上进行改进,通过在 DualGAN 的目标函数的基础上附加 2 个 L1 损失函数来改善实验效果. 通过参数优化实验,寻找最优参数,优化本文提出的模型;通过直观和量化对比实验表明本文提出的模型在人脸数据上的图像翻译效果在清晰度、保持面部特征方面是目前基于生成对抗网络的图像翻译模型中表现最优的,并对相关 GAN 模型的稳定性进行了对比;最后通过效果分析实验说明了所附加的损失函数的具体作用.

1 模型的建立

1.1 模型结构

本文在人脸数据上的图像翻译任务是给定两组

人脸数据,人脸照片和人脸素描数据分别设为域 U 和域 V ,如图 1 所示,本文的模型和 DualGAN 结构一致,由两个 GAN 模型构成. 原始 GAN 生成器 $G_A:U \rightarrow V$ 根据一张人脸素描 $u \in U$ 生成一张人脸照片 $G_A(u, z)$,原始 GAN 的判别器 D_A 用来区分来自生成器 G_A 生成的假图 $G_A(u, z)$ 和来自域 V 的真图 v . 对偶 GAN 的生成器 $G_B:V \rightarrow U$ 根据一张人脸照片 $v \in V$ 生成一张人脸素描 $G_B(v, z')$,对偶 GAN 的判别器 D_B 用来区分来自生成器 G_B 生成的假图 $G_B(v, z')$ 和来自域 U 的真图 u ,生成器 G 通过训练生成尽可能真的图像去欺骗判别器 D ,判别器 D 通过训练尽可能地区分生成器生成的假图和真图. 同时最小化 4 个损失函数

$$L_1^u = \|u - G_B(G_A(u, z), z')\| \quad (1)$$

$$L_2^v = \|v - G_A(G_B(v, z'), z)\| \quad (2)$$

$$L_f = \|u - G_B(v, z')\| \quad (3)$$

$$L_b = \|v - G_A(u, z)\| \quad (4)$$

式中: L_1^u 和 L_2^v 为 2 个 DualGAN 中原有的重建损失函数; L_f 为前向损失; L_b 为后向损失.

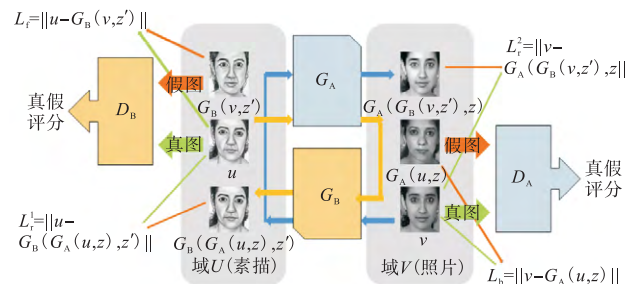


图 1 模型结构

Fig.1 Model structure

1.2 目标函数

DualGAN 并没有使用原始 GAN 的交叉熵损失函数,而是使用 WGAN^[13]所采用的损失函数. 实验表明采用 WGAN 损失函数,生成器的稳定性以及产生的样本更好. DualGAN 判别器 D_A 和 D_B 的目标函数分别为

$$l_A^d(u, v) = D_A(G_A(u, z)) - D_A(v) \quad (5)$$

$$l_B^d(u, v) = D_B(G_B(u, z)) - D_B(v) \quad (6)$$

DualGAN 的生成器的目标函数在 WGAN 的基础上加的是 L1 损失函数而非 L2,因为 L2 时常会导致图像模糊^[11, 14]. DualGAN 的最终目标函数为

$$l^E(u, v) = \lambda_u \|u - G_B(G_A(u, z), z')\| + \lambda_v \|v - G_A(G_B(v, z'), z)\| - D_A(G_B(v, z')) - D_B(G_A(u, z)) \quad (7)$$

本文模型的判别器损失函数和 DualGAN 的损失函数相同,考虑到 DualGAN 生成器的目标函数是在 u 和 $G_B(G_A(u, z), z')$ 与 v 和 $G_A(G_B(v, z'), z')$ 上做了 L1 回归, 尝试在 u 和 $G_B(v, z')$ 与 v 和 $G_A(u, z)$ 之间也加上 L1 回归, 这里要求 u 和 $G_B(v, z')$ 中的 v 是相对应的, v 和 $G_A(u, z)$ 的 u 也是相对应的, 从而本文的模型是需要配对数据的, 不是完全的非监督学习模型. 但经过实验, 发现本文的模型在人脸照片和人脸素描数据上的实验结果优于 GAN 相关的监督学习和非监督学习领域图像翻译的结果. 最终, 本文模型的生成器目标函数为

$$\begin{aligned}
 l^s(u, v) = & \lambda_u (\|u - G_B(G_A(u, z), z')\| + \\
 & \gamma \|u - G_B(v, z')\|) + \lambda_v (\|v - G_A(G_B(v, z'), z')\| + \\
 & \gamma \|v - G_A(u, z)\|) - \\
 & D_A(G_B(v, z')) - D_B(G_A(u, z)) \quad (8)
 \end{aligned}$$

式中: 通过调节参数 λ_u 的大小来控制两个生成器生成的假图 $G_B(G_A(u, z), z')$ 和 $G_B(v, z')$ 向真实图像 u 回归作用在目标函数中的权重; 通过调节参数 λ_v 的大小来控制 2 个生成器生成的假图 $G_A(G_B(v, z'), z)$ 和 $G_A(u, z)$ 向真实图像 v 回归作用在目标函数中的权重; γ 用来调节重建损失函数与本文所附加的双向损失函数在目标函数中回归作用的权重. 在 DualGAN 的实验中提出 λ_u 和 λ_v 的值取 20 比较好, 再通过反复实验发现当 λ_u 、 λ_v 取 20、 γ 取 1 时实验结果最好, 详细过程会在实验部分介绍.

1.3 网络结构

本文模型的网络结构和 DualGAN 网络结构相同, 生成器 G_A 和 G_B 有相同数量的下采样层和上采样层, 在镜像下采样层和上采样层之间使用了 U 型^[5]的跳跃连接. 这样便于低层次信息在输入输出时共享, 此外没有 U 型跳跃连接时, 所有的信息都必须经过同一个路径, 通常会导致大量高频信息的丢失, 通过 U 型跳跃连接还可以减少高频信息的丢失. 生成器中并没有提供明确的噪音 z 和 z' , 噪音是以在训练和测试阶段对生成器的几个层上应用了 dropout 的形式^[4]呈现的.

对于模型的判别器, 使用的是马尔科夫 Patch GAN 结构^[6], 这里 Patch(图像的局部碎片)和模型图像之间的像素距离的独立性, 只是在 Patch 级别, 而不是整张图片. 这样可以有效地捕捉局部的高频特征, 比如纹理风格. 但对模型全局分布方面效果没那么好. 在本文模型中正是利用判别器的 Patch GAN 结构^[6]来获取局部的高频信息, 全局和低频信息通过重建损失函数 L_1^r 和 L_2^r 损失进行获取. 这种配置有

效性在不同的翻译任务中得到验证^[11]. 这种方法的一个额外优点是它需要参数更少, 运行得更快, 并且没有输入图像的大小约束. 判别器 Patch 大小是固定在 70 像素×70 像素, 图像分辨率通常为 256 像素×256 像素, 与 Pix2Pix 相同.

1.4 训练过程

为了优化本文的网络模型, 本文模型和 DualGAN 一样采取了 WGAN 中提出的方法, 详看算法 1. 每训练判别器 n_{critic} 步, 训练 1 次生成器, 使用的是小批量的随机梯度下降法和 RMSProp 优化方法, 相比 Arjovsky 等^[13]以动量为基础的优化方法, RMSProp 更具有稳定性. 把 n_{critic} 设为 2~4, 批量大小设为 1~4, clip 参数 c 设为 0.01~0.10.

传统的 GANs 的训练需要特别注意生成器和判别器间的平衡, 因为鉴别器改进了, 交叉熵损失在局部饱和, 并可能导致梯度消失. 与传统的 GANs 不同, WGAN 的损失几乎在任何地方都是可微的, 这是一个更好的判别器. 在每次迭代中, 生成器都没有经过训练, 直到这些判别器已经接受 n_{critic} 的训练. 这样一个过程使判别器能够提供更可靠的梯度信息.

算法 1 本文模型的训练过程.

变量: 两个图像集 U 和 V , GAN A 的生成器参数为 θ_A 且判别器参数为 ω_A ; GAN B 的生成器参数为 θ_B 且判别器参数为 ω_B ; clip 参数为 c ; 批量大小为 m ; 参数 n_{critic} .

随机初始化 $\omega_i, \theta_i, i \in \{A, B\}$.

循环.

如果 $t = 1, \dots, n_{critic}$, 则

样本 $\{u^{(k)}\}_{k=1}^m \subseteq U, \{v^{(k)}\}_{k=1}^m \subseteq V$;

更新 ω_A 使 $\frac{1}{m} \sum_{k=1}^m l_A^d(u^{(k)}, v^{(k)})$ 最小化;

更新 ω_B 使 $\frac{1}{m} \sum_{k=1}^m l_B^d(u^{(k)}, v^{(k)})$ 最小化;

$\text{clip}(\omega_A, -c, c), \text{clip}(\omega_B, -c, c)$;

结束条件.

样本 $\{u^{(k)}\}_{k=1}^m \subseteq U, \{v^{(k)}\}_{k=1}^m \subseteq V$.

更新 $\theta_A, \theta_B, \frac{1}{m} \sum_{k=1}^m l^s(u^{(k)}, v^{(k)})$.

直到收敛.

2 实验及结果

2.1 参数优化实验

首先使用了 DualGAN 模型里所取的最优参数

λ_v 、 λ_r 均为 20, 对 γ 进行调参实验, 对 γ 分别取值 0、0.5、1、2, $\gamma=0$ 时即附加的双向损失的系数为 0, 此时模型实质上是原 DualGAN, 实验结果如图 2、图 3 所示, 图中 Input 表示输入图片, GT(ground truth)表示真实图片. 在人脸素描向人脸图像翻译的结果中(图 2), 与 $\gamma=1$ 时生成的图像相比, 当 $\gamma=0.5$ 时生成的图像的灰度整体变浅, $\gamma=2$ 生成的图像整体灰度变深, 甚至出现斑驳, 当 $\gamma=0$ 时生成的图像的灰度适宜, 但图像人脸的眼部以及边缘细节比较模糊. 所以

结合 GT 图像, 当 $\gamma=1$ 时生成的图像最理想. 在人脸照片向人脸素描翻译的结果中(图 3), $\gamma=0.5$ 生成的图像效果与 $\gamma=1$ 生成的图像相比相差不大, 实验结果比较理想. $\gamma=2$ 生成的图像与前两者相比相差较多, 虽然能够生成人脸素描, 但没有保持输入图像的人脸面部特征. $\gamma=0$ 生成的图像保持了人的脸部特征, 但人眼等细节处还是比较模糊. 所以结合双向图像翻译的实验结果, 综上认为 $\gamma=1$ 时, 模型实验结果最好.

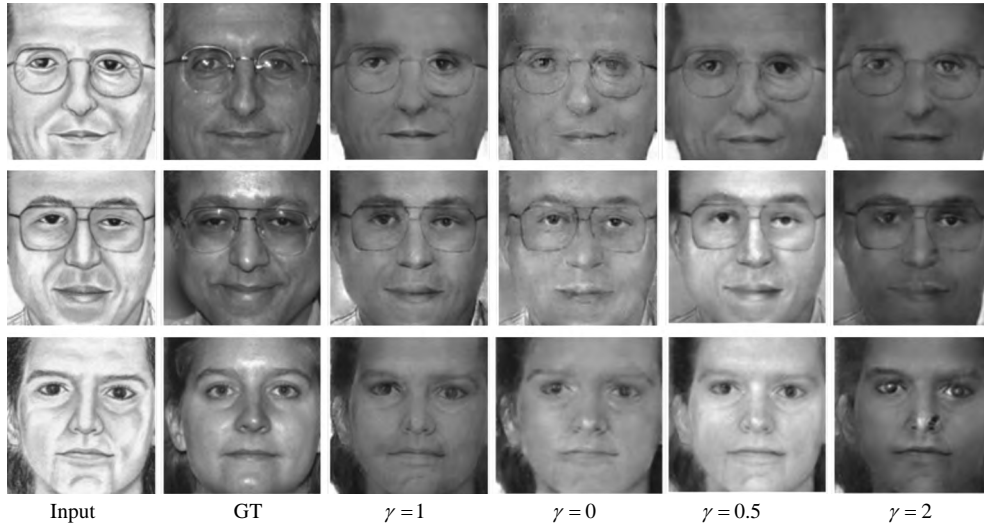


图 2 在不同参数 γ 下人脸素描向人脸照片的翻译结果
 Fig.2 Translation results from sketch to photo with various parameters γ

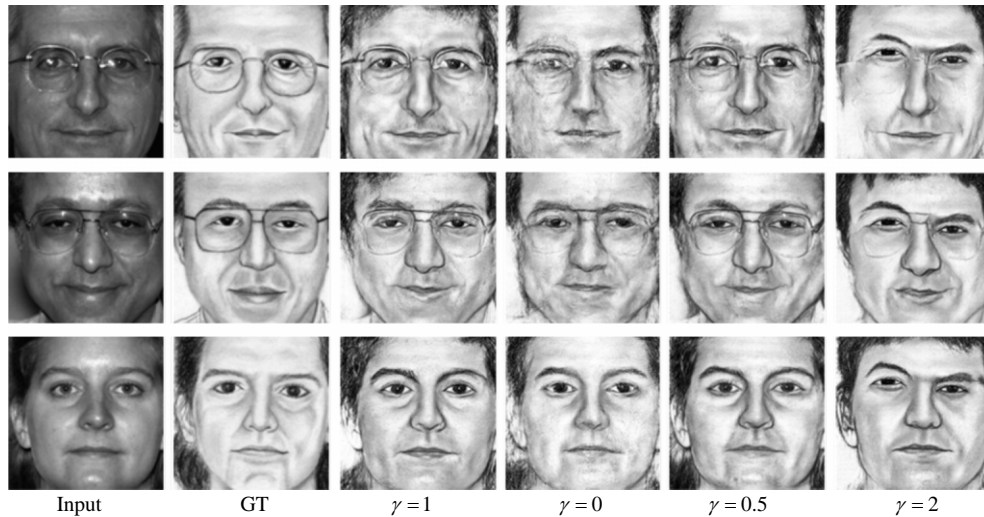


图 3 在不同参数 γ 下人脸照片向人脸素描的翻译结果
 Fig.3 Translation results from photo to sketch with various parameters γ

在确定 $\gamma=1$ 的情况下, 对 λ_v 、 λ_r 进行参数优化实验, 对 λ_v 、 λ_r 取相同值和取不同值分别进行了实验, λ_v 、 λ_r 取相同值分别取 10、20、50、100, 在人脸素描向人脸照片图像翻译的结果(如图 4 所示), 其中, λ_v 、 λ_r 取 20、50 时实验的结果比较清晰, 灰度方面

也保持得比较好; λ_v 、 λ_r 取 1 时, 图像灰度较深, 会出现黑色小斑点; λ_v 、 λ_r 取 100 时图像灰度有时深, 有时浅, 局部会有阴影出现. 在人脸照片向人脸素描图像翻译的结果(如图 5 所示), λ_v 、 λ_r 取 1、20 时实验的结果比较清晰, 灰度方面也保持得比较好; λ_v 、

λ_v 取 50 时,人脸素描会出现少量非正常阴影; λ_u 、 λ_v 取 100 时人脸素描出现严重畸形. 结合 λ_u 、 λ_v 取相同值的实验结果, λ_u 、 λ_v 取 50 时人脸素描翻译成人脸照片结果比较好, λ_u 、 λ_v 取 1 时人脸照片翻译成人脸素描结果比较好, 因此对于不同值实验, 对

$\lambda_u=50$ 、 $\lambda_v=1$ 和 $\lambda_u=1$ 、 $\lambda_v=50$ 分别在人脸数据上进行了实验, 两次图像翻译的双向结果都很差, 出现不同程度的畸形、斑点等问题. 综上分析, 认为 λ_u 、 λ_v 取 20、 γ 取 1 时, 本文的模型在人脸数据上的实验效果最好.

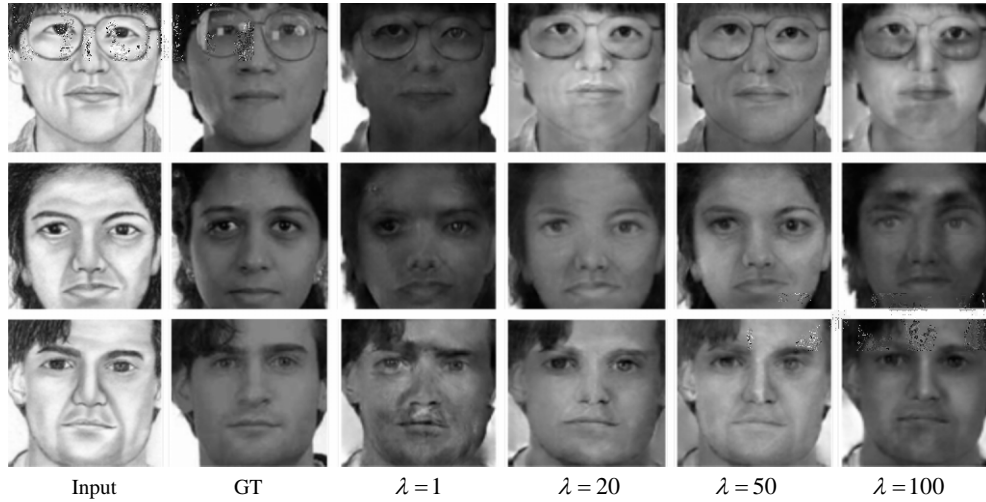


图 4 当参数 λ_u 、 λ_v 取 λ 时人脸素描向人脸照片的翻译结果
Fig.4 Translation results from sketch to photo with $\lambda_u = \lambda$, $\lambda_v = \lambda$

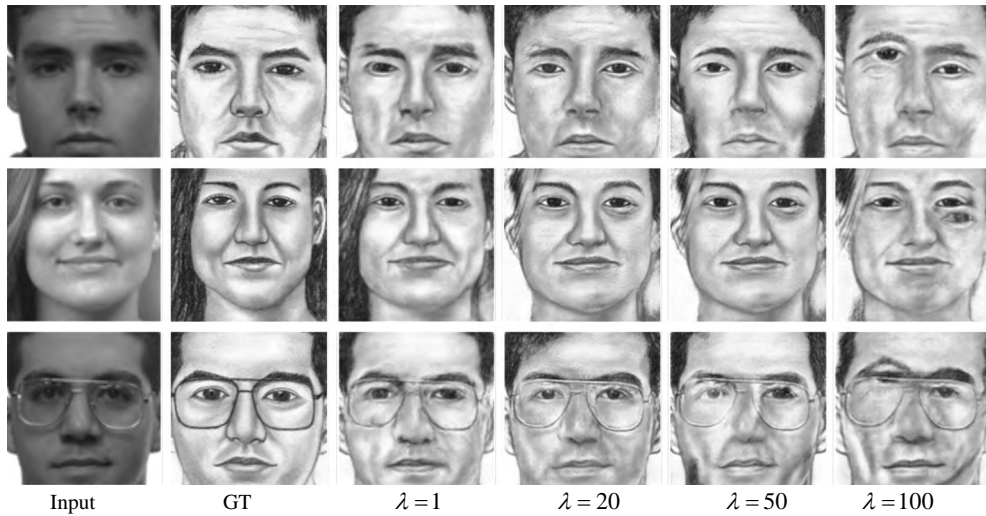


图 5 当参数 λ_u 、 λ_v 取 λ 时人脸照片向人脸素描的翻译结果
Fig.5 Translation results from photo to sketch with $\lambda_u = \lambda$, $\lambda_v = \lambda$

2.2 对比实验

2.2.1 直观对比

首先把本文的模型(这里 λ_u 、 λ_v 为 20, γ 为 1)和在图像翻译领域成果比较好的 CGAN、Pix2Pix、CycleGAN, DiscoGAN 和原始 DualGAN 模型在人脸数据上的实验结果进行了对比.

发现在人脸素描向人脸图像翻译的结果中(如图 6 所示), 监督学习的模型 CGAN 和 Pix2Pix 生成的图像比较清晰, 但有大量的人脸出现畸形和扭曲. 非监督学习的模型 CycleGAN、DiscoGAN 和原始 DualGAN 模型的结果相比较, DiscoGAN 生成的图

像的结果最差, 生成的比较好的图像有人脸的轮廓和眉目, 但图像还是比较模糊, 生成的较差的图像只能看到大概的人脸轮廓. DualGAN 生成的图片相对 DiscoGAN 的比较好, 大部分图像的人脸轮廓眉目甚至细纹都有, 但缺点是图像的某些细节比如眼睛会出现模糊不清的现象, CycleGAN 生成的图像结果最好, 保持了输入的人脸素描的轮廓、五官、细纹等特点, 但与本文的模型生成的图像进行对比, CycleGAN 生成的图像整体带着朦胧感, 给人模糊不清的感觉.

比较人脸照片向人脸素描翻译的结果(如图 7 所

示), 监督学习模型 CGAN 生成的图片比较清晰, 但在大多数人脸的鼻子和嘴巴处出现畸形和斑驳. 监督模型 Pix2Pix 生成的图片, 单从图片生成的质量来说, 生成的图像人脸轮廓和五官都比较清晰, 但跟输入图像或 GT 图像进行对比, 发现 Pix2Pix 模型并没有很好地保持输入图像的面部特征. 非监督学习的模型 CycleGAN、DiscoGAN 和原始 DualGAN 模型的结果相比较, DiscoGAN 生产的图像最差, 没有素描的线条特征, 有的甚至没有人脸的轮廓. CycleGAN

生成的图像整体保持了输入图像的特征, 但没有素描的线条特征, 图像表面好像加了一层小方格, 不具有人脸素描的特点. DualGAN 生成图像的结果相比前两者整体比较好, 既保持了输入图像的面部特征, 也具有素描的线条特征, 与 GT 图像比较接近. 但如果细致观察, 本文的模型生成的图像与 DualGAN 生成的图像相比, 本文的模型生成的图像在五官的细节上更加清晰.

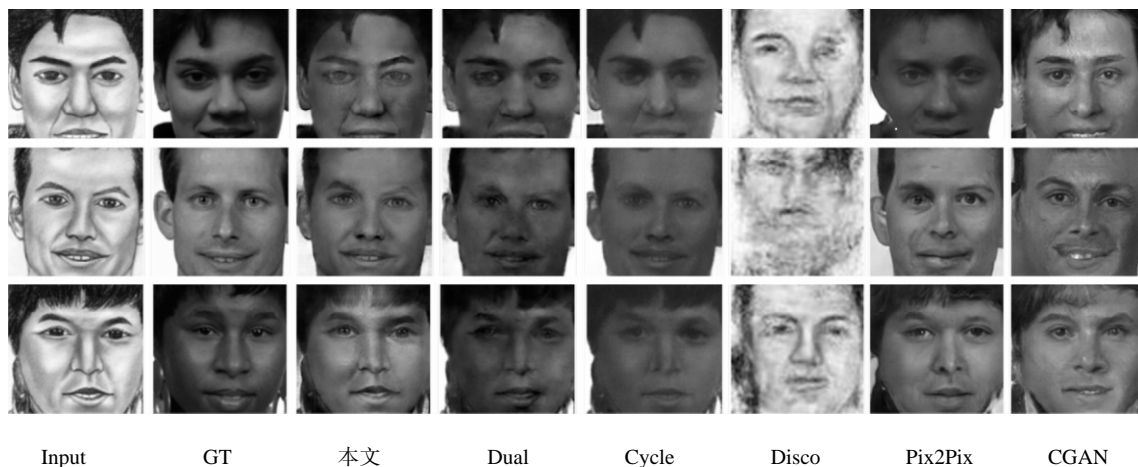


图 6 在不同 GAN 模型下人脸素描向人脸照片的翻译结果

Fig.6 Translation results from sketch to photo with various GAN models

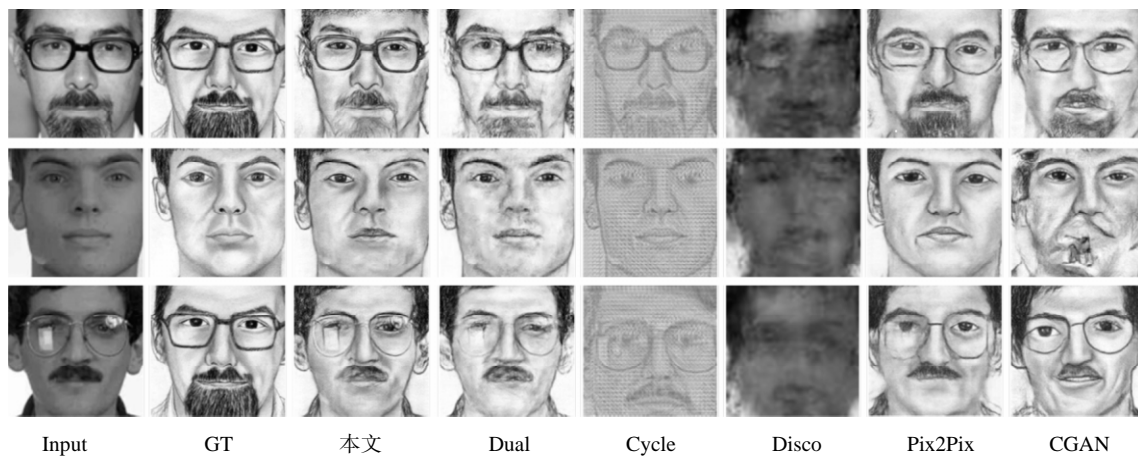


图 7 在不同 GAN 模型下人脸照片向人脸素描的翻译结果

Fig.7 Translation results from photo to sketch with various GAN models

总体上, 在人脸数据上图像翻译的实验中, 本文的模型的结果相比其他模型的结果更加好.

2.2.2 量化对比

为了进一步说明实验结果, 本文对实验结果进行了量化对比. 由于数据太多, 本文没有对所有的模型的结果进行量化对比. 本文选出在人脸数据上翻译较好的 DualGAN、Pix2Pix 和本文模型的结果与 GT 图像进行了量化对比. 具体方法: 本文在测试集中随

机抽取 24 张图, 其中前 12 张用于研究人脸素描向人脸照片的图像翻译的量化对比, 后 12 张用于研究人脸照片向人脸素描的图像的量化对比. 为研究人脸素描向人脸照片的图像翻译结果, 在 DualGAN、Pix2Pix 和本文模型的图像翻译结果中以及 GT 图像中分别找到前 12 张图相对应的人脸照片, 给出每张图的输入的人脸素描, 把来自上述 3 个模型相应的图像翻译的结果以及相应的 GT 图像随机排列, 通过根

据每个输入图像的图像特征对这 4 张随机排列的图像翻译结果进行评分,评分标准具体如下:很差为 1 分,差为 2 分,一般为 3 分,好为 4 分,很好为 5 分. 本文通过编辑电子调查问卷的方式对数据进行收集,最终得到 206 份来自不同 IP 地址的评分结果,得到 4 类来源不同的图像在前 12 张图像翻译的平均分,最后根据每类图像的 12 个平均分再算平均分. 通过对比 4 类图像翻译结果的最终平均分进行量化对比. 研究人脸照片向人脸素描的量化对比实验用后 12 张图做上述类似处理. 本文量化对比结果如表 1 所示,由表中数据可以看出,在人脸素描向人脸照片(S2P)图像翻译的结果中,4 类图像的最终的得分相差相对较大,本文模型的最终得分仅次于 GT 图像,其次为 DualGAN 模型的得分, Pix2Pix 模型的得分最低. 在人脸照片向人脸素描(P2S)的图像的翻译的结果中,4 类图像的最终得分相差相对较小,本文模型得分与 GT 图像仅差 0.07 分,其次为 DualGAN 模型的得分, Pix2Pix 模型的得分仍为最低. 综上,在 DualGAN、Pix2Pix 和本文模型在人脸数据上的图像翻译结果中,本文模型得分最高,本文图像翻译结果最好.

表 1 不同 GAN 模型图像翻译结果真实度平均分

Tab.1 Average realness scores of images translation with various GAN models

任务	真实度平均分			
	本文	DualGAN	Pix2Pix	GT
S2P	3.34	2.94	2.71	3.95
P2S	3.29	3.16	2.89	3.36

此外,通过计算这 4 类图像的真实度得分的方差,进而对不同模型的图像翻译的稳定性进行了比较. 如表 2 所示,在人脸素描向人脸照片的图像翻译结果中, Pix2Pix 模型图像翻译的稳定性最好,本文的模型真实度方差比 DualGAN 模型大 0.04,两模型的稳定性相差不大. 在人脸照片向人脸素描的图像翻译结果中, Pix2Pix 模型图像翻译的稳定性依然最好, DualGAN 模型真实度方差比本文的模型大 0.38,所以本文模型的稳定性要比 DualGAN 好.

表 2 不同 GAN 模型真实度方差

Tab.2 Variance of realness scores with various GAN models

任务	真实度方差			
	本文	DualGAN	Pix2Pix	GT
S2P	1.55	1.51	1.01	0.61
P2S	1.73	2.11	0.96	0.24

2.3 效果分析实验

为了更好地研究本文中所附加的两个损失函数 L_f 和 L_b 的具体作用,分别在 DualGAN 模型、

DiscoGAN 和 CycleGAN 模型的基础上分别进行了效果分析实验.

2.3.1 DualGAN 为基础

分别在 DualGAN 原目标函数上单独加上前向损失 L_f 和单独加上后向损失 L_b 进行双向实验,图 8 是人脸素描向人脸照片图像翻译的结果,发现加双向损失和加后向损失 L_b 的实验结果比较清晰,整体实验结果比较好,而加前向损失 L_f 的图像结果出现不同程度的白斑. 图 9 是人脸照片向人脸素描图像翻译的结果,发现加双向损失和加前向损失 L_f 的实验结果比较清晰,脸部特征方面保持得比较好,而加后向损失 L_b 的图像结果出现类似眼睛的黑斑.

综上分析,认为前向损失 L_f 改善了人脸照片向人脸素描图像翻译的结果,后向损失 L_b 改善了人脸素描向人脸照片图像翻译的结果.

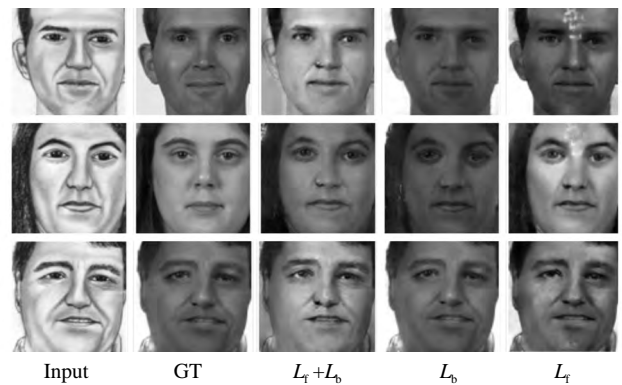


图 8 人脸素描向人脸照片翻译的结果与附加损失函数
Fig.8 Translation results from sketch to photo with additional loss functions

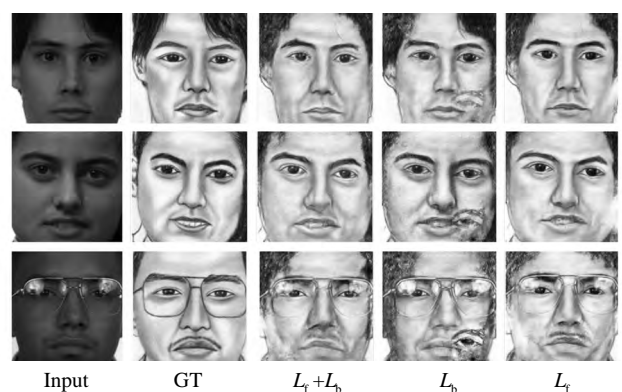


图 9 人脸照片向人脸素描翻译的结果与附加损失函数
Fig.9 Translation results from photo to sketch with additional loss functions

2.3.2 DiscoGAN 为基础

本文在 DiscoGAN 的模型基础上附加前向损失 L_f 和后向损失 L_b 对人脸图像进行了双向翻译,并与原 DiscoGAN 模型的图像翻译结果进行比较,如图 10 所示,附加损失对 DiscoGAN 模型的图像翻译改

善效果并没有像在 DualGAN 模型上的效果那么明显,但通过仔细观察,无论是在人脸素描向人脸照片的翻译结果中,还是在人脸照片向人脸素描的图像翻译结果中,附加的双向损失使所翻译图像人脸的轮廓更加清晰.



图 10 Disco和 Disco + $L_r + L_b$ 的图像翻译结果

Fig.10 Images translation results of Disco and Disco + $L_r + L_b$

2.3.3 CycleGAN 为基础

同样,在 CycleGAN 模型基础上附加前向损失 L_r 和后向损失 L_b 对人脸图像进行了双向翻译实验,但实验结果不仅没有改善,而且变得更糟糕了,如图 11 所示,原 CycleGAN 模型在人脸素描向人脸照片

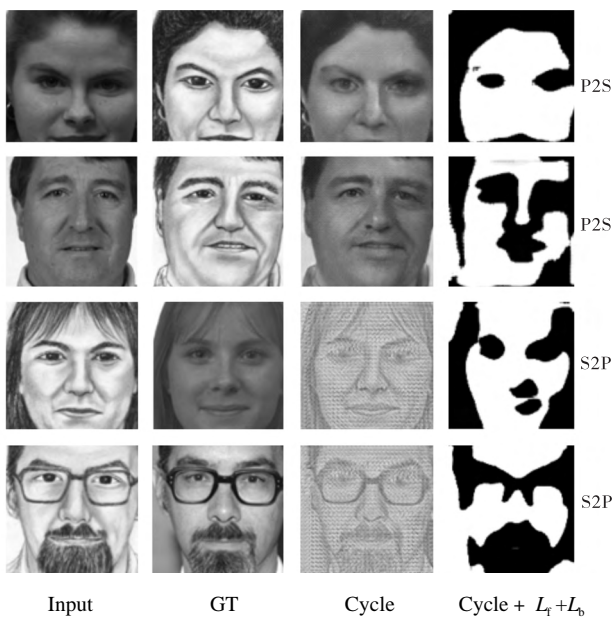


图 11 Cycle和 Cycle + $L_r + L_b$ 的图像翻译结果

Fig.11 Images translation results of Cycle and Cycle + $L_r + L_b$

图像翻译的结果,虽有一些方格纹,但人脸的特征保持比较好,通过附加双向损失,图像翻译结果失去了灰度,变成了非黑即白的轮廓图.原 CycleGAN 模型在人脸照片向人脸素描图像翻译的结果,虽没有典型的素描特征,但人脸的特征保持比较完整,通过附加双向损失,图像翻译的结果同样失去灰度变成非黑即白的轮廓图,与前面不同的是,这些轮廓图中偶尔会出现一些黄色或紫色的斑点.

在 DualGAN 模型、DiscoGAN 和 CycleGAN 模型的基础上分别进行了效果分析实验,得出结论:所附加的双向损失函数并不是可以改善所有的 GAN 相关的图像翻译模型,改善效果和相关的 GAN 模型的具体网络结构有关.

3 结 语

本文在 DualGAN 的模型基础上通过修改目标函数来改善模型,通过一系列实验说明了在人脸数据上,本文提出的模型在图像翻译清晰度和特征保持方面要优于之前的相关的 GAN 模型,并对相关 GAN 图像翻译模型的稳定性进行了比较.不足之处在于本文的模型在测试集的少量图像的翻译结果不稳定,会出现斑点.此外模型只是在单通道的人脸数据上优于各深度学习领域的图像翻译网络模型,在多通道图像上的实验效果并不是最好的,接下来工作希望进一步对模型进行改善,提高模型的稳定性,并使其在多通道图像上的实验效果也能达到最好.

参考文献:

- [1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 3431-3440.
- [2] Mirza M, Osindero S. Conditional generative adversarial nets[J]. Computer Science, 2014: 2672-2680.
- [3] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// International Conference on Neural Information Processing Systems. Kuching, Malaysia, 2014: 2672-2680.
- [4] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 5967-5976.
- [5] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmenta-

- tion[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Calcutta, India, 2015: 234-241.
- [6] Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks [C]//European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 702-716.
- [7] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[EB/OL]. <https://arxiv.org/abs/1703.10593>, 2017.
- [8] Sundaram N, Brox T, Keutzer K. Dense point trajectories by GPU-accelerated large displacement optical flow [C]//European Conference on Computer Vision. Crete, Greece, 2010: 438-451.
- [9] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[EB/OL]. <http://cn.arxiv.org/abs/1703.05192>, 2017.
- [10] Yi Z, Zhang H, Tan P, et al. DualGAN: Unsupervised dual learning for image-to-image translation[C]// IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2868-2876.
- [11] He D, Xia Y, Qin T, et al. Dual learning for machine translation[C]//Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 820-828.
- [12] Liu M Y, Tuzel O. Coupled generative adversarial networks[C]//Advances in neural information processing systems. Barcelona, Spain, 2016: 469-477.
- [13] Arjovsky M, Chintala S, Bottou L. Wasserstein gan [EB/OL]. <http://cn.arxiv.org/abs/1701.07875>, 2017.
- [14] Larsen A B L, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric [C]// International Conference on International Conference on Machine Learning. New York, USA, 2016: 1558-1566.

(责任编辑:王晓燕)