

身份保持约束下的人脸图像补全

王旭东, 卫红权, 高超, 黄瑞阳

(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘要: 人脸图像补全作为图像补全技术的一种特殊应用, 在被遮挡人脸的识别、人脸修复等问题上有不可替代的作用。现有的人脸补全算法只针对补全图像的真实性, 而未考虑其补全后的身份一致性。针对这一问题, 设计了一种基于改进的生成式对抗网络的人脸补全算法, 通过引入 SN-GAN 算法, 提高了模型训练的稳定性, 同时利用人脸识别模型对生成图像加入了身份一致性约束, 经过实验证明, 所提方法能够在生成高真实性图像时有效保持补全图像的身份一致。

关键词: 人脸补全; 身份一致性; 生成式对抗网络; 人脸识别

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2018070

Identity preserving face completion with generative adversarial networks

WANG Xudong, WEI Hongquan, GAO Chao, HUANG Ruiyang

National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China

Abstract: As a special application of image completion technology, face image completion has an irreplaceable role in the occlusion of face recognition, portrait restoration and other issues. The existing face completion algorithm only aims at complementing the authenticity of the image without considering its identity consistency after completion. A face complement algorithm based on improved generative confrontation network was designed. By introducing SN-GAN algorithm, the stability of model training was improved. At the same time, the identity recognition constraint was added to the generated image using the face recognition model. Experiments have shown that the proposed method can effectively maintain the identity of the complementary image when generating high-authenticity images.

Key words: image completion, identity preserving, generative adversarial nets(GAN), face recognition

1 引言

图像补全是指在目标区域填充特定内容, 使填充后的图像保持真实性的技术。它可以实现被遮挡、损坏物体图像的重建。尽管目前有多种方法被用于图像补全任务, 如基于纹理的补全、基

于图像库的补全等。它仍然是一个具有挑战性的问题, 因为上述补全方法未对图像的上下文信息进行有效利用, 对图像中较大面积的缺失修补效果难以令人满意。

人脸图像的补全作为图像补全的一种特殊应用, 由于其在遮挡人脸识别、人脸图像修复等方

收稿日期: 2018-06-25; 修回日期: 2018-07-26

通信作者: 王旭东, 609645296@qq.com

基金项目: 国家自然科学基金资助项目 (No.61601513)

Foundation Item: The National Natural Science Foundation of China (No.61601513)

面的应用而受到人们的关注。Mohammed 等^[1]使用人脸数据集构建修补程序库,并提出了可用于人脸补全的全局和局部参数模型。Deng 等^[2]采用基于谱图的人脸图像修复算法。然而,这些方法补全的图像真实性较低,且需要对应的人脸图像学习填充纹理特征,并不能泛化到任意的人脸补全问题。

生成对抗网络(GAN, generative adversarial net)是Goodfellow 等^[3]在2014年提出的一种无监督学习的训练方法。生成对抗网络由2部分组成:一个生成器网络和一个判别器网络。生成器网络用于生成逼真的样本,鉴别器用于从中分辨出生成样本和原样本。Radford 等^[4]利用生成对抗网络和卷积网络^[5]生成高质量的真假难辨的图像,但由于其训练困难,稳定性较差,近几年一直成为研究改进的热点。SN-GAN 是Miyato 等^[6]对Wasserstein GAN^[7]进行的改进,解决了传统GAN模型中训练困难、损失函数无法指示训练进程、生成样本缺乏多样性的问题。

利用生成对抗网络实现图像的补全可以达到更加真实的效果。Pathak 等^[8]利用GAN的方法设计了上下文补全网络,在补全网络中增加了对抗性损失,取得了较为清晰的结果。而Satoshi 等^[9]使用了全局和局部2个鉴别器作为对抗性损失,使补全图像的细节更加丰富。

身份信息是指识别确定唯一个体所需要的信息,在本文中特指人脸识别中的人脸特征。当前的方法在补全图像的真实性上都有较好的实现,但在人脸补全时未考虑图像包含的身份信息,无法保持人脸补全前后身份的一致性。本文利用生成对抗网络的方法进行人脸补全,同时利用人脸

识别模型Light-CNN^[10]对补全人脸的身份信息进行约束,在保持较高真实性的同时保留其身份信息。本文的主要工作如下。

1) 提出了一种基于SN-GAN的图像补全方法。相比传统GAN方法补全的图像更加真实,训练稳定性更好。

2) 改进的人脸补全方法利用身份一致性约束,使补全的图像能保留人脸的身份特征,补全前后的人脸身份尽可能保持一致,提高了对被遮挡人脸的识别准确率。

2 人脸补全模型

本文提出的人脸补全模型算法框架如图1所示,本节从人脸补全网络、目标函数设计以及网络训练算法3个方面对其进行阐述。

2.1 人脸补全网络

本文提出的人脸补全模型主要是基于SN-GAN模型设计的,由生成器网络(G)和鉴别器网络(D)2部分组成。生成器网络主要是将输入的带有较大空白遮罩的图像补全为完整图像。鉴别器网络主要是度量生成图像和真实图像之间的Wasserstein距离。

设输入的数据集为

$$\{(x_i, m_i) \in X, m \in M\}$$

其中, $X = \{x_1, x_2, x_3, \dots\}$ 为带有空白遮罩的人脸,大小为 $128 \times 128 \times 3$ 的矩阵。 $M = \{m_1, m_2, m_3, \dots\}$ 为图像待补全区域的遮罩,大小为 $128 \times 128 \times 3$ 的矩阵,其中,1代表该像素位置不需要补全,0代表该像素位置需要补全。

设生成器网络G产生的图像为 $N_G(x_i)$,则最终补全的图像为

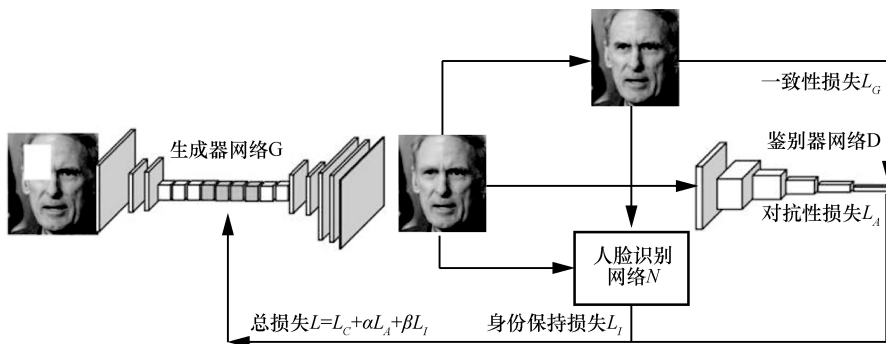


图1 人脸补全模型算法框架

$$N_G(x_i) \odot m_i + x_i \odot (1 - m_i) \quad (1)$$

其中, $\odot$ 代表矩阵元素对应相乘。

定义 2 个分布之间的 Wasserstein 距离为

$$W(P_G, P_T) = \inf_{\gamma \sim \Pi(P_G, P_T)} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (2)$$

其中, $\Pi(P_G, P_T)$ 代表 P_G 和 P_T 组合所有可能的联合分布 γ 的集合。 x, y 分别代表从联合分布 γ 中采样出来的真实样本和生成样本。 $\|x - y\|$ 代表样本 x 与 y 间的距离。 $E_{(x,y) \sim \gamma} [\|x - y\|]$ 代表分布间样本距离的期望。其期望的下确界定义为 Wasserstein 距离。

由于直接计算 Wasserstein 距离比较困难, 根据 Kantorovich-Rubinstein 对偶原理, 可以得到 Wasserstein 距离的等价形式。

$$W(P_1, P_2) = \sup_{\|f\|_{L_1} \leq 1} E_{x \sim P_1} [f(x)] - E_{x \sim P_2} [f(x)] \quad (3)$$

其中, $f(x)$ 必须满足 Lipschitz 连续条件, 即

$$|f(x_1) - f(x_2)| \leq K |x_1 - x_2| \quad (4)$$

这里用鉴别器网络近似函数 $f(x)$ 。设鉴别器网络拟合的函数为 $f_w(x)$, 权值矩阵为 W 。则需使 $f_w(x) = f(x)$, 根据 Takeru Miyato 的方法, 将权重矩阵除以其权重矩阵的最大奇异值使鉴别器网络拟合的函数 $f_w(x)$ 满足 Lipschitz 连续条件, 即

$$W_{sw}(W) = \frac{W}{\sigma(W)} \quad (5)$$

其中, $\sigma(W)$ 为权值矩阵的 L_2 范数, 也就是其最大奇异值。

此时, 生成图像分布 P_G 和真实图像分布 P_T 之间的 Wasserstein 距离 $d(P_G, P_T)$ 为

$$d(P_G, P_T) = |f_w(P_G) - f_w(P_T)| \quad (6)$$

2.2 目标函数设计

2.2.1 对抗性损失

鉴别器网络可以度量真实图像分布和待补全图像分布之间的 Wasserstein 距离。因此对抗性损失 L_A 为

$$L_A = D(G(x \odot m; \theta_g); \theta_D) \quad (7)$$

2.2.2 一致性损失

用于度量补全模型补全的图像和真实图像之间的差异。使用均方差 (MSE) 作为度量函数。一致性损失 L_C 定义为

$$L_C = \text{MSE}(x - G(x \odot m)) = \frac{1}{n} \sum_{i=1}^n |x - G(x_i \odot m_i; \theta_g)| \quad (8)$$

2.2.3 身份保持损失

用于度量补全后的人脸和真实人脸 x_i 身份之间的差异。用人脸识别网络倒数第二层——全连接层的输出作为图像包含身份信息。则身份保持损失 L_I 定义为

$$L_I = \frac{1}{n} \sum_{i=1}^n \|N_I(x_i) - N_I(G(x_i \odot m_i; \theta_g) \odot (1 - m_i) + x_i \odot m_i)\| \quad (9)$$

其中, N_I 是身份鉴别网络, 这里使用预训练好的含有 9 层卷积层的 Light-CNN 人脸识别网络作为身份鉴别网络, 由于预训练好的 Light-CNN 可以对数以万计的人脸进行分类, 它可以捕捉到人脸图像的最显著特征或面部结构。因此, 利用该网络的输出作为图像包含的人脸身份信息是完全可行的^[11]。

综上可得, 总的生成模型损失为

$$L = L_C + \alpha L_A + \beta L_I \quad (10)$$

其中, α 为对抗性损失权重, β 为身份保持损失权重。

2.3 网络训练算法

本文选择优化器为 RMSProp^[12], 不需要手动调整学习率, 可以根据一阶梯度信息自动调整学习率。相对于 SGD 具有较快的收敛速度, 对生成对抗网络模型训练时梯度不稳定的情况适应性较好。通过反向传播算法, 动态更新模型的参数, 使模型的损失函数最小化。训练过程如算法 1 所示。

算法 1 人脸补全模型训练算法

输入

X : 裁剪好的人脸图像;

M : 随机生成的随机大小的图像遮罩;

W : 鉴别器网络 D 每一层的参数矩阵;

$\sigma(W)$: 权值矩阵的最大奇异值;

η : 学习率;

K : 每训练一次生成器鉴别器训练的次数

输出

θ_G : 生成器网络的参数;

W : 鉴别器网络的参数矩阵

1) 随机初始化 θ_G , W

- 2) while θ_G , W 未收敛 do:
- 3) for $t=0, \dots, K$ do
- 4) 从人脸训练集 X 中采样 b 个样本 x_i
- 5) 从遮罩集 M 中采样 b 个样本 m_i
- 6) $g_w \leftarrow \nabla \left[\frac{1}{m} \sum_{i=1}^n f_w(x_i) - \frac{1}{m} \sum_{i=1}^n f_w(g_\theta(x_i \odot m_i)) \right]$
- 7) $w \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w)$
- 8) $w \leftarrow \frac{w}{\sigma(w)}$
- 9) end for
- 10) 从人脸训练集 X 中采样 b 个样本 x_i
- 11) 从遮罩集 M 中采样 b 个样本 m_i
- 12) $g_\theta \leftarrow \nabla \frac{1}{n} \sum_{i=1}^n |x - G(x_i \odot m; \theta_G)| - \alpha \nabla L_c(x_i, m_i; \theta_G) + \beta \nabla L_l(x_i, m_i; \theta_G)$
- 13) $\theta_G \leftarrow \theta_G - \eta \cdot \text{RMSProp}(\theta_G, g_\theta)$
- 14) end while

3 实验过程

3.1 实验设置

本文选用 LFW^[13] 人脸图像数据集。该数据集包含 13 000 多幅从网络收集的脸部图像。每张图像都含有身份信息标签。其中, 1 680 人在数据集中有 2 张或更多不同的照片。将其随机地分成训练集和测试集, 分别包含 11 700 幅和 1 300 幅不同身份的人脸图像。在训练前先采用 Gary 等^[14] 提出的人脸对齐方法, 对图像进行裁剪, 提高裁剪后图像中人脸的比例大小。将处理后的图像缩放至 128 像素×128 像素大小, 并将像素值归一化到 [0,1] 的范围。随机增加大小在 32 像素至 48 像素的空白遮罩, 每个批次包含 128 张图像。对抗性损失权重 α 设置为 0.04, 身份保持损失权重 β 设置为 0.01, 学习率设置为 0.001。

生成器模型由 17 层卷积层组成, 其中, 第 7 层到第 9 层使用了扩张率不为 1 的空洞卷积^[15]来增加卷积层的感知范围。卷积层之间采用 relu^[16] 函数作为激活函数, 并进行批归一化处理^[17]。最后由 Tanh^[18] 激活函数将取值范围映射到 [-1,1]。鉴别器由 6 层卷积层和 1 层全连接层组成。由于要度量生成图像分布和真实图像分布之间的 Wasserstein 距离, 因此在最后一层未采用 Sigmoid^[19] 激活函数改变其输出范围。为了让鉴别器的输出更加平滑, 同时屏蔽掉负值, 鉴别器最后

一层使用了 Softplus^[20] 激活函数将其映射到 $(0, +\infty)$ 。生成器和鉴别器的网络参数本文借鉴了 Satoshi 等^[9] 的模型参数, 如表 1 和表 2 所示。

表 1 生成器网络结构及参数

类型	内核大小	扩张率	步长	输出维度
卷积层	5×5	1	1	64
卷积层	3×3	1	2	128
卷积层	3×3	1	1	128
卷积层	3×3	1	2	256
卷积层	3×3	1	1	256
卷积层	3×3	1	1	256
空洞卷积层	3×3	2	1	256
空洞卷积层	3×3	4	1	256
空洞卷积层	3×3	8	1	256
空洞卷积层	3×3	16	1	256
卷积层	3×3	1	1	256
卷积层	3×3	1	1	256
卷积层	4×4	1	1/2	128
卷积层	3×3	1	1	128
卷积层	4×4	1	1/2	64
卷积层	3×3	1	1	32
卷积层	3×3	1	1	3

表 2 鉴别器网络参数

类型	内核大小	扩张率	步长	输出维度
卷积层	5×5	1	2	64
卷积层	5×5	1	2	128
卷积层	5×5	1	2	256
卷积层	5×5	1	2	512
卷积层	5×5	1	2	512
卷积层	5×5	1	2	512
全连接层	—	—	—	1

3.2 算法评估

本文使用峰值信噪比 (PSNR) 指标对所提方法补全的图像质量进行评估。峰值信噪比表示信号最大可能功率和影响它表示精度的破坏性噪声功率的比值, 常用于衡量图像的重建误差, 其定义如式(23)所示。

$$PSNR = 10 \lg \left(\frac{MAX^2}{MSE} \right) \quad (23)$$

其中, $MSE = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (Y(i, j) - X(i, j))^2 \right)$ 为图像的均方误差, MAX 表示一个像素点颜色的最大数值, 这里为 255, Y 为原始图像, X 为补全后图像, H 和 W 分别代表图像的高度和宽度。在 LFW 数据集和 CelebA^[21] 数据集上分别对所提方法和当前主流的图像补全方法^[8,9,22] 进行比较, 结果如表 3 所示。

表 3 比较不同图像补全方法的峰值信噪比

方法	LFW	CelebA
SC ^[22]	70.01	69.52
CE ^[8]	73.27	70.14
GCDGAN ^[9]	73.98	72.37
本文方法	74.31	72.12

可以看出, 所提方法在图像补全时能够保持较高的生成图像质量。图像补全的效果如图 2 所示。

为了评估所提人脸补全算法能够更好地产生用于人脸识别的人脸特征, 本文在 LFW 数据集上做了对比试验。分别对人脸图像增加随机的范围在 32 像素至 48 像素的空白遮罩, 然后分别在不进行任何处理、利用 GCDGAN 等主流方法对其进行补全、利用本文方法进行补全的情况下, 使用预先训练好的 Light-CNN 人脸分类模型进行身份辨别。准确率 (accuracy) 定义为

$$accuracy = \frac{\sum_{i=1}^n N_i}{n} \quad (24)$$

其中, n 为测试集的图片数量, N_i 当测试图像通过身份辨别模型的辨别结果与该图像在数据集中的身份标签一致, N_i 记为 1, 否则记为 0。采用不同补全方法后采用同一模型进行人脸识别, 准确率如表 4 所示。

表 4 比较不同图像补全方法在图像身份识别上的准确率

方法	准确率
不做处理	9.22%
SC	32.05%
CE	47.13%
GCDGAN	54.33%
本文方法	61.50%

4 结束语

本文通过基于 GAN 方法对人脸图像进行补全的研究, 通过增加身份一致性约束, 使补全人脸图像时可以尽可能恢复出用于人脸识别的特征, 更好地保持补全前后的身份一致, 并与其他方法的对比实验证明了其有效性。

本文设计的基于生成对抗网络的人脸图像补全算法在图像缺失信息较大时, 难以进行有效的补全, 下一步将针对此缺陷进行改进。

参考文献:

- [1] MOHAMMED U, SIMON J D P, KAUTZ J. Visio-lization: generating novel facial images[J]. ACM Transactions on Graphics 2009, 28, (3): 57.
- [2] DENG Y, DAI Q, ZHANG Z. Graph Laplace for occluded face completion and recognition.[J]. IEEE Trans Image Process, 2011,



图 2 随机局部遮挡补全效果

- 20(8):2329-2338.
- [3] GOODFELLOW I J, POUGETABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.
- [4] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. Computer Science, 2015.
- [5] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [6] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks[C]// ICML Workshop on Implicit Models. 2018.
- [7] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs[C]//Advances in Neural Information Processing Systems. 2017.
- [8] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting[C]//IEEE Computer Vision and Pattern Recognition. 2016:2536-2544.
- [9] ISHIKAWA H, ISHIKAWA H, ISHIKAWA H. Globally and locally consistent image completion[C]//ACM Transactions on Graphics (TOG). 2017: 107.
- [10] WU X, HE R, SUN Z, et al. A light CNN for deep face representation with noisy labels[J]. IEEE Transactions on Information Forensics & Security, 2017, (99):1-1.
- [11] HUANG R, ZHANG S, LI T, et al. Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis[J]. arXiv:1704.04086, 2017:2458-2467.
- [12] XU K, BA J, KIROS R, et al. Show, attend and tell: neural image caption generation with visual attention[J]. Computer Science, 2015:2048-2057.
- [13] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[R]. 2007.
- [14] HUANG G B, MATTAR M, LEE H, et al. Learning to align from scratch[C]//Advances in Neural Information Processing Systems (NIPS). 2012.
- [15] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.
- [16] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//International Conference on International Conference on Machine Learning. 2010:807-814.
- [17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. arXiv:1502.03167v3, 2015: 448-456.
- [18] LAU M M, LIM K H. Investigation of activation functions in deep belief network[C]//2017 2nd International Conference on Control and Robotics Engineering (ICCRE). 2017:201-206.
- [19] CYBENKO C. Approximation by superpositions of sigmoidal function[J]. Mathematics of Control Signals & Systems, 1989, 2.
- [20] ZHENG H, YANG Z, LIU W, et al. Improving deep neural networks using softplus units[C]// International Joint Conference on Neural Networks. 2015:1-4.
- [21] YANG S, LUO P, LOY C C, et al. From facial parts responses to face detection: a deep learning approach[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2015:3676-3684.
- [22] HUANG J B, KANG S B, AHUJA N, et al. Image completion using planar structure guidance[J]. ACM Transactions on Graphics, 2014, 33(4):1-10.

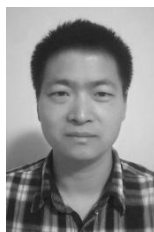
[作者简介]



王旭东 (1992-), 男, 山东青岛人, 国家数字交换系统工程技术研究中心硕士生, 主要研究方向为大数据、音视频处理。



卫红权 (1971-), 男, 河南唐河人, 国家数字交换系统工程技术研究中心副研究员, 主要研究方向为融合网络安全、可重构网络理论与技术。



高超 (1982-), 男, 河南郑州人, 博士, 信息工程大学助理研究员, 主要研究方向为计算机视觉。



黄瑞阳 (1986-), 男, 福建漳州人, 博士, 国家数字交换系统工程技术研究中心助理研究员, 主要研究方向为文本挖掘、图挖掘。