智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于生成对抗网络的人脸口罩图像合成

姜义， 吕荣镇， 刘明珠， 韩闯

引用本文:
姜义, 吕荣镇, 刘明珠, 等. 基于生成对抗网络的人脸口罩图像合成[J]. 智能系统学报, 2021, 16(6): 1073–1080.
JIANG Yi, LYU Rongzhen, LIU Mingzhu, et al. Masked face image synthesis based on a generative adversarial network[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1073–1080.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202012010

DOI: 10.11992/tis.202012010
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20210830.1306.004.html

基于生成对抗网络的人脸口罩图像合成

姜义, 吕荣镇, 刘明珠, 韩闯
(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080)

摘 要: 为了解决现阶段缺乏被口罩遮挡的人脸数据集的问题, 本文提出了基于生成对抗网络与空间变换网络相结合生成戴口罩的人脸图像的方法。本文的方法以生成对抗网络为基础, 结合了多尺度卷积核对图像进行不同尺度的特征提取, 并引入了沃瑟斯坦散度作为度量真实样本和合成样本之间的距离, 并以此来优化生成器的性能。实验表明, 所提方法能够在没有对原始图像进行任何标注的情况下有效地对人脸图像进行口罩佩戴, 且合成的图像具有较高的真实性。

关键词: 深度学习; 生成对抗网络; 空间变换; 卷积神经网络; 图像融合; 口罩; 人脸数据集; 人脸识别
中图分类号: TP391 **文献标志码**: A **文章编号**: 1673−4785(2021)06−1073−08

中文引用格式: 姜义, 吕荣镇, 刘明珠, 等. 基于生成对抗网络的人脸口罩图像合成 [J]. 智能系统学报, 2021, 16(6): 1073–1080.
英文引用格式: JIANG Yi, LYU Rongzhen, LIU Mingzhu, et al. Masked face image synthesis based on a generative adversarial network[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1073–1080.

Masked face image synthesis based on a generative adversarial network

JIANG Yi, LYU Rongzhen, LIU Mingzhu, HAN Chuang
(School of Measurement-Control Technology and Communications Engineering, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: This paper proposes a method for generating masked face images using a generative adversarial network (GAN) and spatial transformer networks. The proposed method is used to solve the present problem of lacking face datasets of people wearing masks. Based on the GAN, the proposed method introduces a multiscale convolution kernel to extract image characteristics in various dimensions. This method introduces the Wasserstein divergence to measure the distance between an authentic specimen and a synthetic specimen so that generator's performance can be optimized. Experiments show that the proposed method can add a mask to a face image effectively without any annotations on the original image, and the synthesized image has high fidelity.

Keywords: deep learning; generative adversarial networks; spatial transformation; convolution neural network; image fusion; face mask; human face dataset; face recognition

Coronavirus disease 2019(COVID-19) 虽然在我国已经得到了很好的控制, 但仍然在全球一些地区蔓延。COVID-19 是指 2019 年开始流行的新型冠状病毒感染导致的肺炎, 是一种急性呼吸道传染病[1]。导致该肺炎的病毒可以通过呼吸道飞沫在人群中进行大范围的传播。此外, 病毒感染者接触过的物体也可能残留病毒, 人们可能通过接触这些物体导致感染。所以戴口罩出行和在公共场所保持社交距离成为了阻止疫情传播的重要方法。同时由于该病毒具有接触传染的特性, 在公共场合使用指纹或掌纹等接触式的身份识别方式同样存在安全风险。人脸识别系统由于能够避免不必要的接触因而比其他识别方式安全得多。在口罩成为生活必需品时, 也对现有的人脸识别系统提出了挑战。目前的基于深度学习的人脸识别方法[2-3], 在面对无遮挡物的人脸识别上取得了很好的识别率, 但是在大面积遮挡的人脸面前已经不再能够准确识别身份了[4]。其主要原因在于训

收稿日期: 2020−12−03. **网络出版日期**: 2021−08−30.
基金项目: 国家自然科学基金项目 (61601149); 黑龙江省科学基金项目 (QC2017074); 黑龙江省普通本科高等学校青年创新人才培养计划项目 (UNPYSCT-2018199).
通信作者: 姜义. E-mail: jasonj@hrbust.edu.cn.

练人脸识别神经网络模型时，没有使用戴口罩的人脸数据进行训练。所以，为了提高人脸识别系统对口罩遮挡人脸的识别率，需要一个具有大量样本的戴口罩的人脸数据集。在目前该类型数据集缺乏的情况下，为了更好地训练神经网络对戴口罩人脸进行识别，本文通过给现有的人脸识别数据集中的人脸图像戴口罩的方式解决该问题。

Anwar 等[5] 采用基于 dlib 的面部检测器来识别人脸和口罩上的 6 个关键点，然后将口罩上的关键点与人脸上的关键点对应，最后将口罩图片进行拉伸等变换后贴图在人脸图片上的相应位置得到戴口罩的人脸图片。Cabani 等[6] 采用的技术与 Anwar 相似，不同的是采用了 12 个人脸的关键点和 12 个口罩关键点来对人脸图片进行口罩佩戴。Anwar 与 Cabani 的方法虽然较为简单易用，但是得到的佩戴口罩人脸图片真实性不高，有明显的人工痕迹。

本文认为给人脸图片戴上口罩本质上是一个图像融合问题。图像融合是用特定算法将两幅或多幅图像融合成一幅新的图像。在融合的过程中通过利用图像的相关性和互补性，使得融合后的图像达到想要的效果。随着卷积神经网络的出现，图像融合取得了显著的进展。其中大多数方法都是通过学习低维的自然图像子空间的编码，并由此对像素进行预测并约束图像可能的外观，最终生成融合后的图像。

生成对抗网络 (generative adversarial networks, GAN)[7-8] 就是一个强大的图像生成网络。它包含两个相互竞争和博弈的神经网络模型，生成器 (generator) 和判别器 (discriminator)。生成器将噪声作为输入并生成图片，判别器则接收生成器产生的数据和对应的真实数据，训练得到能正确区分生成数据与原始数据的分类器。这种能从随机噪声生成图像的方式使得生成对抗网络备受关注。GAN 除了无监督的训练方法，还能通过给定标签得到特定的图像，例如，CGAN[9] 就是通过引入一个额外信息进行半监督的图像生成。Radford 等[10] 则是将 GAN 网络和卷积神经网络进行结合，得到了一个更稳定的深度卷积生成对抗网络 (deep convolutional generative adversarial network, DCGAN)，而且极大地提高了生成图像的质量。此外，Arjovsky 等[11] 提出的 WGAN 网络 (wasserstein GAN) 将计算生成的图像数据分布与真实的图像数据分布之间的 Jensen–Shannon 距离 (简称 JS 距离) 改为 Wasserstein 距离 (W 距离)。W 距离帮助 WGAN 解决了原始 GAN 网络的模式坍塌问题，使得生成的样本更加多样化，而且使得训练更加稳定。WGAN 变体还有 WGAN-GP[12] 和 WGAN-div[13] 等，进一步提高了生成的多样性和图片质量。

1　相关技术

本文提出的方法结合了空间变换网络和使用金字塔卷积的 WGAN-div 生成对抗网络。

1.1　空间变换网络

空间变换网络 (spatial transformer networks, STNs)[14] 模块主要由 3 个部分组成：本地化网络 (localization network)、参数采样网络 (parameterized sampling grid) 和图像采样 (image sampling)。本地化网络的输入是原始的图片，输出是一个变换参数 p，它映射的是输入图片和理想图片的坐标关系。参数采样网络则是对特征图像进行仿射变换，通过变换参数和输入特征图的坐标位置，得到对应的特征关系。而图像采样是经过前两个网络得到的特征关系对原图像进行变换以得到期望的图像。其主要思想是对输入的图像进行空间变换，输出一张变换后的理想图像。本文将采用空间变换网络将口罩图像进行变换以使其符合人脸轮廓，从而得到逼真的戴口罩人脸图像。本文采用空间变换网络的一个主要原因是可以不用提前对口罩图像进行控制点标注，提高了算法的实用性。为了达到这个目标，本文将使用生成对抗网络来优化空间变换网络的参数 p。

1.2　金字塔卷积网络

卷积网络在计算机视觉中得到了广泛的应用[15-16]。但是卷积网络的实际感受野比理论上的要小很多；且池化、卷积步长等下采样方案都会产生信息的损耗，进而影响模型的性能。Duta 等[17] 提出的金字塔卷积 (pyramidal convolution, PyConv) 可以在多个滤波器尺度对输入进行处理。PyConv 包含一个核金字塔，每一层包含不同类型的滤波器，每个滤波器的大小和深度都不同，以此来提取不同尺度的图像特征。

PyConv 采用了金字塔结构的卷积，包含了不同深度和尺度的卷积核，能够同时提取不同尺度的特征。PyConv 的结构如图 1 所示，它包含了一个由 n 层不同尺寸卷积核构成的金字塔，能够在不提升计算复杂度和参数数量的基础上采用多尺度核对输入进行处理，每一层的核包含不同的空间尺度，卷积核尺度越大，深度越低。由于 PyConv 在不同层使用不同深度的卷积核，需要将输入特征划分为不同的组并独立地进行卷积计算，称之为组卷积。假设 PyConv 的输入通道数为 C，每一

层的卷积核大小为 $K_1^2,K_2^2,\cdots,K_n^2$，C_i 对应输入的特征维度，而对应的输出特征维度为 $C_{o1},C_{o2},\cdots,C_{on}$。$C_o$ 是输出特征维度的总和，C_o 和参数量 N 如式 (1)、式 (2) 所示。

$$C_{o1}+C_{o2}+\cdots+C_{on}=C_o \tag{1}$$

$$N=K_n^2\times C_{on}\times\frac{C_i}{\left(\frac{K_n^2}{K_1^2}\right)}+\cdots+K_4^2\times C_{o4}\times\frac{C_i}{\left(\frac{K_4^2}{K_1^2}\right)}+\cdots+K_1^2\times C_{o1} \tag{2}$$

在构建过程中，PyConv 的每一层的通道数应当是相同的，这也要求输入通道数必须是 2 的指数次幂。

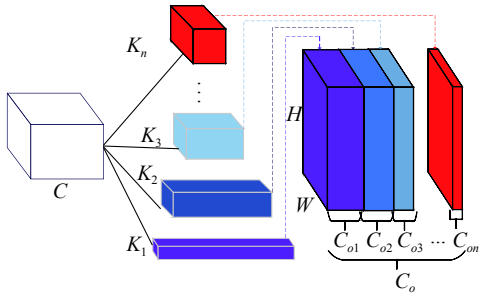

图 1　PyConv 结构

Fig. 1　Structure of PyConv

1.3　Wasserstein GAN 网络

生成对抗网络由于在训练中需要达到纳什均衡，一直存在着训练困难以及不稳定的问题。不稳定的问题也会导致模式坍塌，造成样本生成缺乏多样性，即使增加训练时间也很难改善。而 Wasserstein GAN(WGAN) 较好地解决了训练不稳定的问题，不再需要小心地对生成器和判别器的训练程度进行平衡，确保了生成结果的多样性。原始 GAN 采用的 JS 距离衡量的是两个分布之间的差异，通过将 JS 散度作为优化目标最终得到优化的生成网络。但是这只能在两个分布有重叠部分时才成立，如果原始图像和生成的图像在分布上没有重叠部分或重叠部分可忽略不计，则对应的 JS 散度就是一个固定值，这样无论如何训练都无法得到优化的生成器。WGAN 则采用 W 距离，其定义如式 (3)：

$$\mathrm{EM}(p_r,p_f(Z))=\inf_{\gamma\in\prod(p_r,p_f(Z))}E_{(x,y)\sim\gamma}[\|x-y\|] \tag{3}$$

式中：p_r、$p_f(Z)$ 分别表示真实样本分布和生成样本的分布；$\gamma\in\prod(p_r,p_f(Z))$ 中，γ 表示联合分布，后面表示联合分布的集合；$\|x-y\|$ 则是样本 x、y 之间的距离。W 距离是计算所有联合分布中能够对期望值取到的下界。为了最小化 W 距离，分别为生成器和判别器设计了两个损失函数 G 和 D，如式 (4)、(5) 所示：

$$G=-E_{x\sim P_f}[D(x)] \tag{4}$$

$$D=E_{x\sim p_f}[D(x)]-E_{x\sim P_r}[D(x)] \tag{5}$$

2　人脸图像与口罩图像的合成

本文的目标是对输入的未戴口罩人脸图像 I_F、口罩图像 I_M 以及掩膜 M 的情况下进行图像合成。对口罩图像进行空间变换，校正其视角、位置及方向，使得合成的照片更加自然，合成的过程表示如式 (6) 所示：

$$I_{\text{composite}}=I_M\odot M+I_F\odot(1-M) \tag{6}$$

2.1　网络结构设计

相较于通过将标注好的口罩图像通过关键点定位叠加到识别出的人脸上的方法，本文提出一种使用金字塔卷积改进的 WGAN-div 神经网络与空间变换网络相结合的图像合成模型 (Py-WGAN-div)，该模型在训练时不需要对待合成的人脸及口罩图像做任何提前标注。模型以生成对抗网络为主体，并分别对生成器和判别器的神经网络部分进行了改进，其结构如图 2 所示。

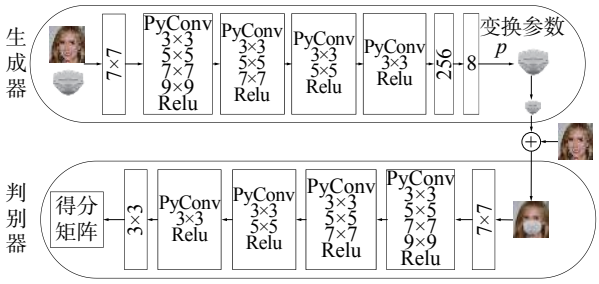

图 2　网络整体结构

Fig. 2　Overall neural network structure

在生成器部分，原始的生成对抗网络是通过一个随机噪声来生成新的图像。但是直接生成的图像会有很多问题，例如生成的人脸分辨率低、口罩错误的被作为肤色合成到人脸上。而本文的目的是能构建成对的人脸数据 (包含未戴口罩与戴口罩的人脸图像对)，而不是产生随机的人脸。因此本文方法中的生成器生成的是一组更新的变形参数 Δp (且该变形参数随着优化的进行不断进行更新)。修正的变换参数如式 (7) 所示：

$$\begin{aligned}\Delta p_i&=G_i(I_M(p_{i-1}),I_F)\\ p_i&=p_{i-1}+\Delta p_i\end{aligned} \tag{7}$$

式中：I_F 为未佩戴口罩的人脸图像；I_M 为通过变形参数变形后的口罩图片；p_{i-1} 表示上一次的变形参数；G_i 表示生成器。

多尺度卷积与标准单一卷积相比，能在没有额外参数的情况下，扩大卷积核的感受野，并且由于使用不同大小的卷积核而获得不同的空间分

辨率和深度。卷积核随着尺寸的减小深度加深，这样不同尺寸的卷积核能带来互补的信息，有助于获取更丰富的特征。

在生成器网络的结构设计上，本文采用了 Py-Conv 的网络结构。首先将输入的正常人脸图像和口罩图像进行通道数的叠加，再通过一个 7×7 的大卷积核提取特征。使用 7×7 的大卷积核的目的是尽可能地保留原始图片的信息且减少计算量。之后采用了 4 个去掉了批标准化层的 Py-Conv 卷基层。在 PyConv 中卷积层中去除批量归一化层的目的是减少计算的复杂度以提升训练效率。生成器中第一个 PyConv 层用了 4 个不同尺寸的卷积核 (3×3、5×5、7×7、9×9) 以获取不同尺度的特征来增强模型的特征提取能力。生成器的输出是一个 8 维的向量，该向量作为空间变换网络的参数使用。生成器的网络结构如表 1 与图 2 所示，其中 s 为步长。

表 1　Py-WGAN-div 网络结构

Table 1　Py-WGAN-div network structure

输出尺寸/像素×像素	Py-WGAN-div
72×72	7×7, 64, s=2
36×36	3×3, 32, s=2 5×5, 32, s=2 7×7, 32, s=2 9×9, 32, s=2
18×18	3×3, 64, s=2 5×5, 64, s=2 7×7, 128, s=2 3×3, 256, s=2
9×9	5×5, 256, s=2
5×5	3×3, 1 024, s=2
1×1	256
1×1	8

判别器的输入数据是由真实的佩戴口罩的人脸图像和合成的戴口罩图像构成。而合成图像则是通过空间变换网络生成的变形后的口罩图像和未遮挡人脸图像进图像融合产生的，空间变换网络的参数则是由生成器生成而来的。通过判别器网络对图像进行判别后，输出一个分值来表示图像合成的质量。判别器网络没有使用全连接层，而是通过一个 3×3 的卷积得到 5×5×1 的矩阵来计算得分。由图 3 所示的训练流程图可知，本文通过不断地优化变换参数的值来优化对口罩进行的投射变换，并最终获得较好的合成图像。

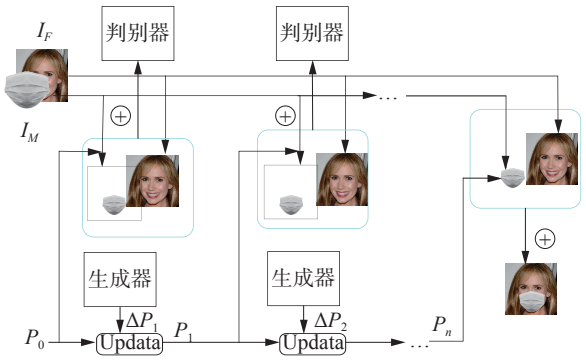

图 3　网络训练流程图

Fig. 3　Training flow diagram

2.2　损失函数与超参数设置

本文将采用 WGAN-div 的目标函数作为优化指标。WGAN 虽然相比 GAN 网络有了很大程度的优化，但是在训练过程中仍然表现出收敛速度慢和训练困难的问题[18]。其主要原因是：在处理利普希茨连续条件限制条件时直接使用了权重裁减，大部分权重都在±0.01。而判别器希望能尽可能地拉大真假样本之间的差距。之后其他研究者提出的 WGAN-GP[12] 和 SNGAN[19] 分别通过了梯度惩罚和谱归一化的方法实现了 Lipschitz 约束。而且 WGAN 不能使用基于动量的优化算法，例如 Adam。WGAN-div 则是提出了式 (8) 所示的 W 散度来真正缩小两个分布之间的距离损失函数：

$$W_{k,p}(P_r,P_f)=\max_{D\in C^1}E_{x\sim P_r}[D(x)]-E_{x\sim P_f}[D(x)]-kE_{x\sim\text{random}}[(\|\nabla_x D(x)\|^p)] \tag{8}$$

式中：k 和 p 的设置则是根据经验选取，通常设置 $k=2$，$p=6$；C^1 指一阶连续函数族测度；P_r 与 P_f 分别表示真实样本集合和合成样本集合；$D(x)$ 和 $G(x)$ 分别表示判别器和生成器损失函数。

通过式 (8) 可以得到生成器的损失函数 $G(x)$ 和判别器的损失函数 $D(x)$ 分别为式 (9) 和式 (10)：

$$G=\max_{D\in C^1}E_{x\sim P_r}[D(x)]-E_{x\sim P_f}[D(G(z))] \tag{9}$$

$$D=\max_{D\in C^1}E_{x\sim P_r}[D(x)]-E_{x\sim P_f}[D(x)]-kE_{x\sim\text{random}}[(\|\nabla_x D(x)\|^p)] \tag{10}$$

本文使用的超参数设置为：生成器和判别器的学习率都是 0.000 01，空间变换次数为 5 次，总迭代次数为 30 万次，每 2 万次迭代后学习率衰减，批量大小为 20，优化算法采用 Adam。

3　实验与结果分析

3.1　数据集与实验环境

本文中用于训练的图片数据挑选自武汉大学发布的人脸口罩数据集[20]、使用网络爬虫从互联网上抓取的图片以及从其他研究者[21] 合成的人脸口罩数据集中挑选的部分图片。在对选取的图

片中进行随机平移、旋转以及缩放后, 获得总计158 462 张戴口罩的人脸图片作为数据集。将其中142 618 张图片 (约 90%) 作为训练集进行判别器的训练, 其余则作为测试集。数据集中图片尺寸统一缩放至 144 像素×144 像素。并手工制作了20 张类型、各种花色的口罩图片,口罩图片的尺寸同样是 144 像素×144 像素, 且口罩基本位于图片的中心, 如图 4 所示。本文采用的实验环境配置如表 2 所示。

图 4　口罩图片
Fig. 4　Masks used in experiments

表 2　实验环境配置
Table 2　Experiment configuration

实验环境	参数
处理器	Xeon E3-1285L
内存	32 GB
GPU	GTX 1080Ti
操作系统	Ubuntu Linux 18.04
编程框架	Tensorflow1.14

3.2　实验结果分析

实验使用基于 Py-WGAN-div 的生成对抗网络对训练集进行训练, 在训练时随机从训练集中选取人脸图片和口罩图片。图 5 显示了使用本文方法进行训练时, 每 5 万次迭代并更新口罩的变换参数后合成的图片。从图 5 中可以看出, 口罩位置随着训练的进行逐渐变得更加贴合面部, 最终得到了较真实的人脸佩戴口罩图像。

图 5　训练过程
Fig. 5　Training process example

在进行算法对比时, 本文选取了关键点匹配算法[5]、基于 GAN、DCGAN、和 WGAN 的算法进行对比。根据本文实验, 原始的生成对抗网络算法在人脸口罩合成上效果很差, 口罩几乎无法覆盖正确的位置。因此本文将对比组算法中的 GAN、DCGAN、WGAN 的生成器和判别器之间增加了空变换网络 (spatial transformer network, STN) 以更合理地进行对比。图 6 是不同算法对应不同口罩和人脸合成的效果对比图。在对比时选取了 5 种风格不同的口罩, 有最常见的蓝色外科口罩、KN95 口罩、粉色、方格以及斑点花纹的口罩。对比实验还选取了 4种不同肤色的人脸及背景, 包括各种肤色与背景颜色。针对不同人脸、不同口罩以及不同算法进行了对比, 结果如图 6 所示。从图 6(a)、(b) 可以看出, 在人脸姿态比较好的时候, 各种算法都能较好地将口罩合成到人脸图像中。其中, 基于关键点匹配的算法[5] 和本文的算法效果最好, 但本文的算法产生的图像更加自然和逼真。从图 6 中可以清楚地看到, 基于 GAN 和 DCGAN 的算法生成的图片效果相对比较差, 口罩会遮住眼睛或者完全超过人脸的轮廓; 而 WGAN 的方法效果虽然比基于 GAN 和 DCGAN 的算法更好, 但合成的口罩不能很好地贴合人脸轮廓。

(a) 样本 1

(b) 样本 2

(c) 样本 3

(d) 样本 4

图 6　不同算法结果对比

Fig. 6　Comparisons between various methods

图 6(c) 中的人脸往右偏，除本文方法外的其他方法的结果中，口罩只能比较好地贴合左半边脸，右半边脸的口罩则会过大。此时本文的算法虽然也不十分理想，但是能够基本贴合人脸轮廓，相对更好一些。而对于图 6(d) 中低头的人脸，除了基于关键点的算法和本文的算法外，其他算法获得的人脸口罩图像都有较大失真，表现在不能覆盖下巴和口鼻，或者覆盖了不该覆盖的区域，相较之下，在图像中的人脸姿态不是正面面向镜头时，本文算法获得的人脸口罩图像仍然是更好的。综上得出，在图像中的人脸姿态没有正面面向摄像头时，所有的算法得到的戴口罩人脸图像都有所欠缺，但是本文的算法在人脸口罩合成的效果上明显优于其他算法，基本能够贴合人脸的轮廓，没有遮挡不该遮挡的部位，并且在细节上更加真实。

为了更客观地比较不同算法的合成效果，本文采用了 IS Score(inception score)[22-23]、结构相似性 (structural similarity) 和深度特征度量图像相似度 (learned perceptual image patch similarity)[24-25] 3 个指标来客观评价各种不同 GAN 模型在口罩合成上的效果。

IS 评价方法将生成的图片送入训练好的 Inception 分类模型中。该 Inception 分类模型的输出是一个 1 000 维的标签，该标签的每一个维度表示了输入图像属于某个分类的概率。如果训练结果较好，结果会比较集中。结果如表 3 所示，虽然 GAN 和 DCGAN 网络生成的戴口罩图片与希望的结果相差甚远，但它们的 IS 分数却比本文算法更高。出现这个现象是因为，虽然 IS 能够作为图像合成质量的一个指标，但该指标无法真正反映合成图像中的细节，例如：口罩是否正确地覆盖了人的嘴巴和鼻子，口罩覆盖的区域是否过大而导致面部信息的丢失，脸部的其他部位是否有被保留等。因此本文还对生成的图片进行了人工评判。人工评判的方式为将 100 组不同的人脸图像分别给 20 个人进行评分，每组图像包含了人脸的原始图像、佩戴口罩的类型以及两种方法合成后的图片，并对合成后的图像进行是否真实的判别，判别结果如表 3。由判别结果可知，本文方法合成的图像更加真实。

表 3　性能对比 1

Table 3　Performance comparison 1

方法	IS	人工评价/%
GAN	2.567	0
DCGAN	2.433	23
WGAN	2.272	71
本文方法	2.326	77

此外，本文还采取了两种相对客观的评价方法，结构相似性 (structural similarity, SSIM) 和深度特征度量图像相似度 (learned perceptual image patch similarity, LPIPS)[24]，来对生成的图像进行评价。SSIM 是一种参考的图像质量评估指标，通过对图像的亮度、对比度和结构 3 个方面对图像的相似度进行比较度量。而深度特征度量图像相似度则是使用由预训练的神经网络提取的特征图来量化两幅图像之间的感知差异，两幅图越相似则距离越近。SSIM 和 LPIPS 指标的对比结果如表 4 所示。

从表 4 中可以看出，本文算法相比与对比算法在结构相似度上更高。而深度特征度量图像相似度非常小，说明了合成的戴口罩图像与真实的戴口罩的人脸距离很接近，充分证明了本文算法的有效性。

表 4　性能对比 2
Table 4　Performance comparison 2

方法	SSIM	LPIPS
GAN	0.634	0.170
DCGAN	0.748	0.119
WGAN	0.898	0.058
本文方法	0.921	0.010

4　结束语

本文提出了一种生成对抗网络与空间变换网络相结合的给人脸图像佩戴口罩的方法, 并且在设计上采用了由生成对抗网络生成空间变换网络的变换参数, 而不是直接生成人脸与口罩融合后的图像的特殊设计。在设计神经网络时使用了多尺度卷积的方法, 使生成器能更好地提取特征。在训练时采用了 W 距离作为衡量两个不同样本之间距离的计算, 克服了生成对抗网络训练难且容易出现模式坍塌的问题。相比于其他方法, 本文方法在合成的图像更加逼真, 口罩也更贴合人脸。

实验结果显示, 在人脸和口罩都无任何标记的情况下, 该神经网络模型可以学习到相应的变换参数并合成高质量的人脸戴口罩图像。实验结果证实, 融合后的人脸图像不失真且很好的保留了面部特征, 同时也将口罩覆盖到了人脸正确的位置。在研究过程中也发现, 在人脸图像由于角度问题只有半张脸可见的情况下本文方法效果不完美的问题。因此如何在任意角度对人脸图片上不失真地进行口罩合成将是进一步的研究方向, 进一步将利用本文制作的戴口罩人脸数据集进行口罩遮挡的面部识别研究。

参考文献:

[1] 国家卫生健康委员会.新型冠状病毒肺炎诊疗方案 (试行第八版)[EB/OL].(2020-08-18)[2020-12-08]http://www.gov.cn/zhengce/zhengceku/2020-08/19/content_5535757.htm.
National Health Commission of the People's Republic of China. Diagnosis and treatment protocol for novel coronavirus pneumonia (trial version 8) [EB/OL].(2020-08-18)[2020-12-01]http://www.gov.cn/zhengce/zhengceku/2020-08/19/content_5535757.htm.

[2] TAIGMAN Y, YANG Ming, RANZATO M A, et al. DeepFace: closing the gap to human-level performance in face verification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1701−1708.

[3] 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. Boston, USA, 2015: 815−823.

[4] 李小薪, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习 [J]. 计算机学报, 2018, 41(1): 177–207.
LI Xiaoxin, LIANG Ronghua. A review for face recognition with occlusion: from subspace regression to deep learning[J]. Chinese journal of computers, 2018, 41(1): 177–207.

[5] ANWAR A, RAYCHOWDHURY A. Masked face recognition for secure authentication[EB/OL].(2020-08-25)[2020-12-01] https://arxiv.org/abs/2008.11104.

[6] CABANI A, HAMMOUNDI K, BENHABILES H, et al. Masked-Face-Net—a dataset of correctly/incorrectly masked face images in the context of covid-19[EB/OL].(2020-08-18)[2020-12-01] https://arxiv.org/abs/2008.08016.

[7] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada, 2014: 2672−2680.

[8] 胡铭菲,刘建伟,左信.深度生成模型综述 [EB/OL].（2021-10-28）[2021-10-30] https://doi.org/10.16383/j.aas.c190866.
HU Mingfei, LIU Jianwei, ZUO Xin. Survey on deep generative model[EB/OL].（2021-10-28）[2021-10-30] https://doi.org/10.16383/j.aas.c190866.

[9] MIRZA M, OSINDERO S. Conditional generative adversarial nets[EB/OL].(2014-11-06)[2020-12-01] https://arxiv.org/abs/1411.1784.

[10] RADFORD A, METZ L, CHINTALA S. Unsupervised repress-enttation learning with deep convolutional generative adversarial networks[EB/OL]. (2016-01-07)[2020-12-01] https://arxiv.org/abs/1511.06434.

[11] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN [EB/OL].(2017-12-06)[2020-12-01] https://arxiv.org/abs/1701.07875.

[12] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA, 2017: 5769−5779.

[13] WU Jiqing, HUANG Zhiwu, THOMA J, et al. Wasserstein divergence for GANs[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 673−688.

[14] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[C]//Proceedings of the

28th International Conference on Neural Information Processing Systems. Montreal, Canada, 2015: 2017−2025.

[15] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1−9.

[16] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6): 1229–1251.
ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. Chinese journal of computers, 2017, 40(6): 1229–1251.

[17] DUTA I C, LIU L, ZHU F, et al. Pyramidal convolution: rethinking convolutional neural networks for vis-ual recognition[EB/OL].(2020-06-20)[2020-12-01] https://arxiv.org/abs/2006.11538.

[18] ARJOVSKY M, BOTTOU L. Towards principled methods for training generative adversarial networks [EB/OL]. (2017-01-17)[2020-12-01] https://arxiv.org/abs/1701.04862.

[19] MIYATO T, KATAOKA T, KOYAMA M, et al. Spectral normalization for generative adversarial networks [C]//6th International Conference on Learning Representations. Vancouver, Canada, 2018.

[20] WANG Z, WANG G, HUANG B, et al. Masked face recognition dataset and application[EB/OL].(2020-03-23)[2020-12-01] https://arxiv.org/abs/2003.09093.

[21] LIU Ziwei, LUO Ping, WANG Xiaogang, et al. Deep learning face attributes in the wild[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 3730−3738.

[22] HAN Xintong, WU Zuxuan, WU Zhe, et al. VITON: an image-based virtual try-on network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7543−7552.

[23] HAN Xintong, WU Zuxuan, HUANG Weilin, et al. Compatible and diverse fashion image inpainting [EB/OL]. (2019-04-24)[2020-12-01] https://arxiv.org/abs/1902.01096.

[24] PANDEY N, SAVAKIS A. Poly-GAN: multi-conditioned GAN for fashion synthesis[J]. Neurocomputing, 2020, 414: 356–364.

[25] DONG Haoye, LIANG Xiaodan, SHEN Xiaohui, et al. Towards multi-pose guided virtual try-on network [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South), 2019: 9025−9034.

作者简介：

姜义，讲师，主要研究方向为人工智能、传感器网络、分布式系统。

吕荣镇，硕士研究生，主要研究方向为人工智能。

刘明珠，副教授，主要研究方向为通信与信息系统。发表学术论文 10 余篇。