超维面部转换：用于逼真和身份保持正面视图合成的全局和局部感知 GAN

Rui Huang[1,2] Shu Zhang[1,2,3] Tianyu Li[1,2] Ran He[1,2,3] [1]National Laboratory of Pattern Recognition, CASIA [2]Center for Research on Intelligent Perception and Computing, CASIA
[3]University of Chinese Academy of Sciences, Beijing, China
huangrui@cmu.edu, tianyu.lizard@gmail.com, {shu.zhang, rhe}@nlpr.ia.ac.cn

摘 要

来自单面图像的照片级真实正面视图合成在面部识别领域具有广泛的应用。尽管已经提出了数据驱动的深度学习方法来通过从充足的面部数据中寻找解决方案来解决这个问题，但是这个问题仍然存在挑战，因为它本质上是不适合的。本文提出了一种双向通路生成对抗网络（TP-GAN），用于通过同时感知全局结构和局部细节来进行逼真的正面视图合成。除了常用的全局编码器 - 解码器网络之外，还提出了四个定位于地标的补丁网络来处理局部纹理。除了新颖的架构之外，我们通过引入对抗性损失，对称性损失和识别保留损失的组合来很好地约束这个不适定问题。组合损失函数利用正面分布和预训练的判别性深度面部模型来指导身份保持对轮廓的正面视图的推断。与以往主要依赖中间特征识别的深度学习方法不同，我们的方法直接利用合成大小的身份保留图像进行人脸识别和归因估计等下游任务。实验结果表明，我们的方法不仅能够呈现出令人难以置信的感知结果，而且在大型姿势人脸识别方面也优于最先进的结果。

1. 介 绍

受益于深度学习方法的快速发展以及容易获得大量带注释的人脸图像，无约束的人脸识别技术[28,29]近年来取得了重大进展。 虽然超越了人类的表现

图 1. TP-GAN 的正面视图合成。 上半部分显示 90°剖面图像（中间）及其相应的合成和真实正面图像。 我们邀请读者猜测哪一方是我们的综合结果（请参阅第 1 节的答案）。 下半部分分别显示了 90°，75°和 45°剖面的合成正面视图.

在几个基准数据集[25]上实现，姿势变化仍然是许多实际应用场景的瓶颈。 解决姿势变化的现有方法可以分为两类。 一个类别尝试采用手工制作或学习的姿势不变特征[4,25]，而另一个类别采用合成技术从大型姿势脸部图像恢复正面视图图像，然后使用恢复的脸部图像进行人脸识别[41,42]。

对于第一类，传统方法通常利用强大的局部描述符，如 Gabor [5]，Haar [32]和 LBP [2]来解释局部失真，然后采用元学习[4,33]技术来实现姿势不变性。 相比之下，深度学习方法通常采用汇集操作来处理位置变量并使用三元组损失[25]或

*这两位作者的贡献相同

对比损失[28]，以确保非常大的类内变化的不变性。然而，由于方差和可辨性之间的权衡，这些方法不能有效地处理大的姿势情况。

对于第二类，早期对正面视图合成的努力通常利用 3D 几何变换来渲染正面视图，方法是首先将 2D 图像与一般[12]或特定身份[29,40] 3D 模型对齐。这些方法擅长对小姿势面进行归一化，但是由于严重的纹理损失，它们在大面部姿势下的性能会降低。最近，基于深度学习的方法被提出用于以数据驱动的方式恢复正面。例如，朱等人[42]建议在学习估计正面视图的同时去除身份和姿势表征。尽管它们的结果令人鼓舞，但合成图像有时缺乏精细的细节，并且在大的姿势下趋于模糊，因此它们仅使用中间特征进行面部识别。合成图像仍然不足以执行其他面部分析任务，例如取证和属性估计。

此外，从优化的角度来看，从不完全观察到的轮廓恢复正面视图是一个不适定或不明确的问题，如果没有考虑先验知识或约束，则存在多个解决这个问题的方法。因此，恢复结果的质量很大程度上依赖于训练过程中利用的先验或约束。以前的工作[15,38,41,42]通常采用成对监督，很少在训练过程中引入约束，因此它们往往会产生模糊的结果。

当人类尝试进行视图合成过程时，我们首先根据我们的先验知识和观察到的轮廓推断出正面的全局结构（或草图）。然后我们的注意力转移到所有面部细节将被填写的当地区域。受此过程的启发，我们提出了一种具有两个路径（TP-GAN）的深层架构，用于正面视图合成。这两种途径分别关注全局结构的推断和局部纹理的转换。然后将它们相应的特征图融合，用于进一步合成最终合成的过程。我们还通过将生成的面对面分布的先验知识与生成性对抗网络（GAN）相结合，使恢复过程得到很好的约束[9]。 GAN 在 2D 数据分布建模方面的出色能力显着提升了许多不适定的低水平视觉问题，如超分辨率[17]和修复[21]。特别是，从面部的对称结构中汲取灵感，提出了一种对称的损失来填充遮挡部分。此外，为了忠实地保留个体最突出的面部结构，除了像素方式的 L1 损失之外，我们在紧凑特征空间中采用感知损失[14]。结合身份保护损失对产生有用的综合结果来说至关重要，并大大提高其应用于面部分析任务的潜力。我们在图 1 的上半部分（每个元组的左侧）显示了由 TP-GAN 生成的一些样本。

我们工作的主要贡献在于三个方面：1）我们从单个图像中提出了一个类似人类的全局和局部感知 GAN 架构，用于正面视图合成，即使在非常特殊的情况下，它也可以合成照片级真实感和身份保持大姿势下的正面视图图像。 2）我们结合来自数据分布（对抗训练）的先验知识和面部的领域知识（对称性和身份保持损失）来精确地恢复将 3D 对象投影到 2D 图像空间中所固有的丢失信息。 3）我们展示了“通过生成识别”框架的可能性，并且在大的姿势下超越了最先进的识别结果。尽管已经提出了一些用于面部合成的深度学习方法，但是我们的方法是对合成面部的识别任务有效的第一次尝试.

2. 相关工作

2.1. 正面视图合成

正面视图合成，或称为面部归一化，由于其不适合的性质，是一项具有挑战性的任务。传统方法通过 2D / 3D 本地纹理变形[12,40]或统计建模[24]来解决这个问题。例如，哈斯纳等人[12]采用平均 3D 模型进行面部归一化。在[24]中提出了一种联合正面视图合成和地标定位方法，其具有约束低秩最小化模型。最近，研究人员采用卷积神经网络（CNN）进行联合表示学习和视图合成[15,38,41,42]。具体来说，Yim 等人[38]提出了一种多任务 CNN 来预测保持旋转图像的身份。朱等人[41,42]开发新颖的架构和学习目标，以在估计正面视图的同时解开身份和姿势表示。里德等人[22]建议使用玻尔兹曼机器模拟变异因子，并通过姿势流形遍历生成旋转图像。虽然如果合成的图像可以直接用于面部分析任务更方便，但是大多数先前的方法主要采用中间特征进行面部识别，因为它们不能正确地产生身份保持合成。

2.2. 生成对抗网络 (GAN)

作为深度生成模型研究的最重要改进之一[16,23]，GAN [9]引起了深度学习和计算机视觉社会的极大关注。 最小 - 最大双人游戏提供了一种简单而有效的方法来估计目标分布并生成新的图像样本[6]。 随着它

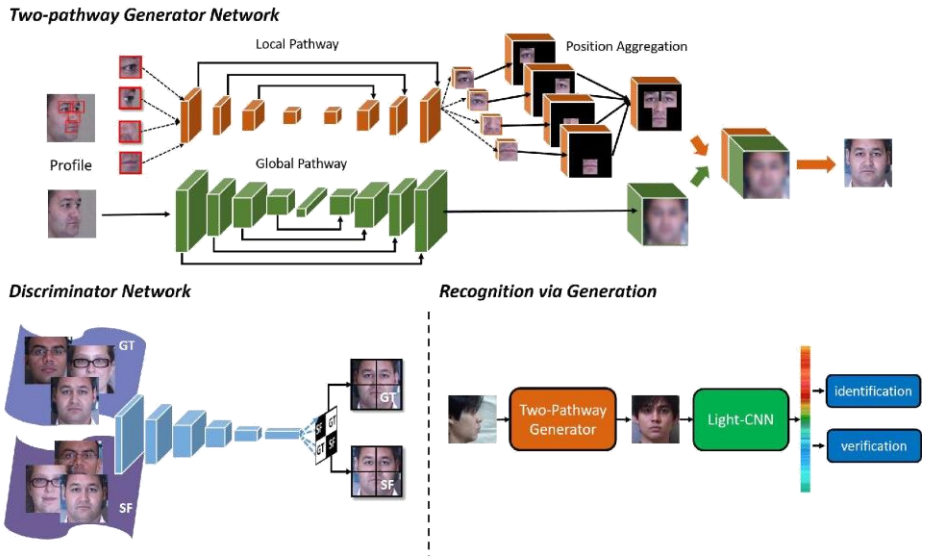

图 2. TP-GAN 的一般框架。 Generator 包含两个路径，每个路径处理全局或局部变换。 鉴别器区分合成正面（SF）视图和完全真实（GT）正面视图。 详细的网络架构可以在补充材料中找到。

用于分布建模的功能，GAN 可以鼓励生成的图像朝向真实图像移动，从而生成具有合理的高频细节的照片级真实感图像。 最近，改进的 GAN 架构，特别是有条件的 GAN [19]，已成功应用于视觉任务，如图像修复[21]，超分辨率[17]，样式转换[18]，面部属性控制等。 [26]甚至数据增加用于提升分类模型[27,39]。 GAN 的这些成功应用激励我们开发基于 GAN 的正面视图合成方法。

3. 方 法

正面视图合成的目的是从不同姿势下的面部图像（即轮廓图像 IP）恢复真实感和身份保持正面视图图像 IF。为了训练这样的网络，在训练阶段期间需要来自多个身份 y 的相应{IF，IP}对。输入 IP 和输出 IF 都来自具有 C 颜色通道的尺寸为 W×H×C 的像素空间。

我们的目标是学习一个综合函数，该函数可以从任何给定的轮廓图像中推断出相应的正面视图。具体而言，我们使用由 θ_G 参数化的双路 CNN G_{θ_G} 对合成函数建模。每个路径包含编码器和解码器，表示为 $\{G_{\theta_E^g}, G_{\theta_D^g}\}$ 和 $\{G_{\theta_E^l}, G_{\theta_D^l}\}$，其中 g 和 l 分别代表全局结构路径和局部纹理路径。在全局路径中，作为 $G_{\theta_E^g}$ 输出的瓶颈层通常用于具有交叉熵损失 Lcross-entropy 的分类任务[37]。

网络参数 G_{θ_G} 由最小化特定设计的合成损失 Lsyn 和前述的 Lcross-entropy 优化。对于具有 $\{I_n^F, I_n^P\}$ 的 N 个训练对的训练集，优化问题可以如下公式化：

$$\hat{\theta}_G = \frac{1}{N}\underset{\theta_G}{\operatorname{argmin}} \sum_{n=1}^{N} \{L_{syn}(G_{\theta_G}(I_n^P), I_n^F) + \alpha L_{cross\_entropy}(G_{\theta_E^g}(I_n^P), y_n)\} \quad (1)$$

其中α是加权参数，Lsyn 被定义为共同约束图像以驻留在所需流形中的若干损失的加权和。我们将把所有单个损失函数的详细描述推迟到 Sec3.2。

3.1. 网络架构

3.1.1 两个通路生成器

TP-GAN 的一般架构如图 2 所示。与以前的方法[15,38,41,42]不同，通常用单一网络对合成函数进行建模，我们提出的生成器 G_{θ_G} 有两个路径，一个全局网络 G_{θ^g} 处理全局结构和四个定位于地标的补丁网络 $G_{\theta_i^l}$，i∈{0,1,2,3}，它们关注四个面部地标周围的局部纹理。

我们不是第一个采用两种途径建模策略的人。实际上，这是一种非常流行的 2D / 3D 局部纹理变形[12,40]方法。类似于人类认知过程，他们通常将面部的标准化分为两个步骤，第一步是将面部全局与 2D 或 3D 模型对齐，第二步是

将局部纹理扭曲或渲染到全局结构。此外，穆罕默德等人[20]将全局参数模型与用于新颖面部合成的局部非参数模型相结合。

从轮廓图像 IP 合成正面 IF 是高度非线性的变换。由于过滤器在人脸图像的所有空间位置共享，我们认为仅使用全局网络无法学习适合旋转脸部和精确恢复局部细节的过滤器。因此，我们将传统方法中两种途径结构的成功转化为基于深度学习的框架，并引入用于正面视图合成的类人二路径生成器。

如图 2 所示，G_{θ_G} 由下采样编码器 $G_{\theta_E^g}$ 和上采样解码器 $G_{\theta_D^g}$ 组成，引入额外跳过层用于多尺度特征融合。中间的瓶颈层输出 256 维特征向量 Vid，其用于身份分类以允许身份保持合成。在这个瓶颈层，如[30]所示，我们将 100 维高斯随机噪声连接到 Vid，以模拟除姿态和身份以外的变化。

3.1.2 地标定位补丁网络

位于地标中的贴片网络 G_{θ^l} 的四个输入贴片是从四个面部标志中心裁剪的，即左眼中心，右眼中心，鼻尖和嘴中心。每个 $G_{\theta_i^l}$，i∈{0,1,2,3}学习一组单独的滤波器，用于将中心裁剪的贴片旋转到其相应的正面视图（旋转后，面部标志仍位于中心）。位于地标的补丁网络的架构也基于编码器 - 解码器结构，但它没有完全连接的瓶颈层。

为了有效地整合来自全局和本地路径的信息，我们采用直观的方法进行特征图融合。如图 2 所示，我们首先将四个局部路径的输出特征张量（多个特征图）融合到一个单一特征张量中，该特征张量与全局特征张量具有相同的空间分辨率。具体来说，我们将每个特征张量放在“模板界标位置”，然后引入最大化融合策略以减少重叠区域上的拼接伪影。然后，我们简单地连接每个路径的特征张量以产生融合特征张量，然后将其馈送到连续的卷积层以生成最终的合成输出。

3.1.3 对抗网络

为了将正面面部分布的先验知识结合到训练过程中，我们进一步引入了一个判别器 D_{θ_D}，以便根据 Goodfellow 等人的工作来区分真实的正面图像 IF 和合成的正面图像 $G_{\theta_G}(I^P)$ [9]。我们训练 D_{θ_D} 和 D。以交替的方式优化随后的最小 - 最大问题:

$$\min_{\theta_G}\max_{\theta_D}\mathbb{E}_{I^F\sim P(I^F)}\log D_{\theta_D}(I^F)+\mathbb{E}_{I^P\sim P(I^P)}\log(1-D_{\theta_D}(G_{\theta_G}(I^P))) \tag{2}$$

解决这个最小 - 最大问题将持续推动生成器的输出以匹配训练正面的目标分布，因此它促使合成图像驻留在正面的多个面中，导致照片般逼真的合成具有吸引人的高精度细节。如在[27]中，我们的 D_{θ_D} 输出 2×2 概率图而不是一个标量值。现在每个概率值都对应于某个区域而不是整个面部，并且 D_{θ_D} 可以专门关注每个语义区域。

3.2. 合成损失函数

我们工作中使用的综合损失函数是四个单独损失函数的加权和，我们将在以下部分给出详细描述.

3.2.1 像素损失

我们在多个位置采用像素方式的 L1 损耗，以便于实现多尺度图像内容的一致性:

$$L_{pixel}=\frac{1}{W\times H}\sum_{x=1}^{W}\sum_{y=1}^{H}|I_{x,y}^{pred}-I_{x,y}^{gt}| \tag{3}$$

具体地，在全局，位于地标的补丁网络及其最终融合输出的输出处测量像素损失。为了便于深入监督，我们还在 $G_{\theta_D^g}$ 的多尺度输出上添加约束。虽然这种损失会导致合成结果过于平滑，但它仍然是加速优化和卓越性能的重要组成部分。

3.2.2 对称性损失

对称性是人脸的固有特征。利用该领域知识作为先验并对合成图像施加对称约束可以有效地消除自遮挡问题，从而大大提高大型姿势案例的性能。具体来说，我们在两个空间中定义对称损失，即原始像素空间和拉普拉斯图像空间，这对于照明变化是稳健的。面部图像的对称性损失采用这种形式:

$$L_{sym}=\frac{1}{W/2\times H}\sum_{x=1}^{W/2}\sum_{y=1}^{H}|I_{x,y}^{pred}-I_{W-(x-1),y}^{pred}| \tag{4}$$

为简单起见，我们选择性地翻转输入，使得被遮挡的部分都在右侧。此外，只有 Ipred 的遮挡部分（右侧）才能获得对称性

(a) Profile (b) Ours (c) [30] (d) [38] (e) [8] (f) [40] (g) [12] (h) Frontal

图 3.在 45°（前两行）和 30°（最后一行）的姿势下与最先进的合成方法的比较

.

损失，即我们明确地将右侧拉近左侧。 Lsym 的贡献是双重的，通过鼓励对称结构产生逼真的图像，并通过提供额外的反向传播梯度来加速 TP-GAN 的收敛，从而减轻自身遮挡的影响。然而，由于光照变化或内在纹理差异，像素值在大多数时间不是严格对称的。幸运的是，局部区域内的像素差异是一致的，并且沿着所有方向的点的梯度在很大程度上保留在不同的照明下。因此，拉普拉斯空间对于光照变化更稳健并且更能指示面部结构。

3.2.3 对抗性损失

用于从合成正面图像$G_{\theta_G}(I^P)$区分真实正面图像 IF 的损失计算如下:

$$L_{adv} = \frac{1}{N}\sum_{n=1}^{N} -\log D_{\theta_D}(G_{\theta_G}(I_n^P)) \tag{5}$$

用作监督以推动合成图像驻留在正面视图图像的多个部分中。它可以防止模糊效果并产生视觉上令人愉悦的结果。

3.2.4 身份保护损失

在合成正面视图图像时保留身份是开发“通过生成识别”框架中最关键的部分。在这项工作中，我们利用最初提出的感知损失[14]来保持感知相似性，以帮助我们的模型获得身份保持能力。具体来说，我们根据 Light CNN 最后两层的激活来定义身份保留损失[35]:

$$L_{ip} = \sum_{i=1}^{2} \frac{1}{W_i \times H_i} \sum_{x=1}^{W_i}\sum_{y=1}^{H_i} |F(I^P)_{x,y}^i - F(G(I^{pred}))_{x,y}^i| \tag{6}$$

其中 Wi，Hi 表示最后一个第 i 层的空间维度。身份保持损失强制预测与紧凑的深度特征空间中的真实图像具有小的距离。由于 Light CNN 经过预先训练以对成千上万的身份进行分类，因此它可以捕获最突出的特征或面部结构以进行身份识别。因此，利用这种损失来强制执行保持正面视图合成的身份是完全可行的。

与 Ladv 一起使用时，Lip 具有更好的性能。单独使用 Lip 会使结果容易出现恼人的伪影，因为搜索局部最小的 Lip 可能会经过一条位于自然面部图像之外的路径。同时使用 Ladv 和 Lip 可以确保搜索位于该流形中并产生照片般的图像。

3.2.5 总体目标功能

最终的综合损失函数是上面定义的所有损失的加权和:

$$L_{syn} = L_{pixel} + \lambda_1 L_{sym} + \lambda_2 L_{adv} + \lambda_3 L_{ip} + \lambda_4 L_{tv} \tag{7}$$

我们还对合成结果施加了总变差正则化 Ltv [14]以减少尖峰伪像。

4. 实验

除了合成自然前视图像外，所提出的 TP-GAN 还旨在生成身份保持图像，以便通过现成的深层特征进行精确的面部分析。因此，在本节中，我们证明了我们的模型在定性合成结果并且在 4.1 和 4.2 节中给出了定量识别结果中的优点。 4.3 节展示了最终深度特征表示的可视化，以说明 TP-GAN 的有效性。最后，在 4.4 节，我们进行详细的算法评估，以证明所提出的两个路径架构和综合损失函数的优点。

实施细节 我们使用大小的彩色图像 在我们所有的实验中都有 128×128×3 的输入

图 4. TP-GAN 在不同姿势下的合成结果。从左到右，姿势分别为 90°，75°，60°，45°，30°和 15°。真实正面图像在最后一列提供。

图 5.具有挑战性的情况。面部属性，例如胡须，眼镜由 TP-GAN 保存。被遮挡的额头和脸颊被收回。

表 1.设置 1 下的视图和照明的 Rank-1 识别率（%）。对于所有剩余的表，只有标有*的方法遵循“通过生成识别”程序，而其他表格利用中间特征进行人脸识别

Method	±90°	±75°	±60°	±45°	±30°	±15°
CPF [38]	-	-	-	71.65	81.05	89.45
Hassner *et al.* * [12]	-	-	44.81	74.68	89.59	96.78
HPN [7]	29.82	47.57	61.24	72.77	78.26	84.23
FIP_40 [41]	31.37	49.10	69.75	85.54	92.98	96.30
c-CNN Forest [36]	47.26	60.66	74.38	89.02	94.05	96.97
Light CNN [35]	9.00	32.35	73.30	97.45	99.80	99.78
TP-GAN*	**64.03**	**84.10**	**92.93**	**98.58**	**99.85**	**99.78**

IP 和预测 Ipred = $G_{\theta_G}(I^P)$。我们的方法在 MultiPIE [10]上进行了评估，这是一个拥有 750,000 多个图像的大型数据集，用于在姿势，光照和表情变化下进行人脸识别。功能提取网络 Light CNN 在 MS-Celeb-1M [11]上进行了训练，并对 MultiPIE 的原始图像进行了微调。我们的网络是通过 Tensorflow [1]实现的。 TP-GAN 的训练持续一天，批量为 10，学习率为 10-4。在我们所有的实验中，我们根据经验设定：

α= 10-3，λ1= 0.3，λ2= 10-3，λ3= 3×10-3 且 λ4= 10-4。

4.1. 面部合成

以前关于正面视图合成的大部分工作都致力于在±60°的姿势范围内解决该问题。因为通常认为姿势大于 60°，很难忠实地恢复正面视图图像。但是，我们将展示给定足够的训练数据和适当的架构和损失设计，实际上可以从非常大的姿势恢复逼真的正面视图。图 4 显示了 TP-GAN 从任何姿势中恢复令人信服的保持身份的正面的能力图

(a) Ours (b)[38] (c)[8] (d)[40] (e)[12]

图 6.每个标识的六个图像（±45°内）的平均面

图 3 示出了与现有技术的脸部正面化方法的比较。请注意，大多数 TP-GAN 的竞争对手无法处理大于 45°的姿势，因此，我们仅在 30°和 45°下报告其结果。

与竞争方法相比，TP-GAN 在产生照片拟真合成的同时提供了良好的身份保持质量。得益于 Ladv 和 Lip 的先前知识的数据驱动建模，不仅整个面部结构而且局部的耳朵，脸颊和前额可以以一致的方式产生图像。此外，它还完美地保留了原始轮廓图像中的观察到的面部属性，例如，眼镜和发型，如图 5 所示。

为了进一步证明多个姿势的合成的稳定几何形状，我们在图 6 中显示了来自不同姿势的合成面部的平均图像。来自 TP-GAN 的平均面部保留了更多纹理细节并且包含更少的模糊效果，显示出稳定跨多个合成的几何形状。请注意，我们的方法不依赖于任何 3D 知识进行几何形状估计，推理是通过纯粹的数据驱动学习来完成的。

为了证明我们模型在野外人脸中具有出色的通用能力，我们使用来自 LFW [13]数据集的图像来测试仅在 Multi-PIE 上训练的 TP-GAN 模型。如图 7 所示，尽管得到的色调类似于来自 Multi-PIE 的图像，但是 TP-GAN 可以忠实地合成具有更精细细节的正面视图图像和用于 LFW 数据集中的面部的更好的全局形状，即便与顶尖方法如[12,40]相比。

(a) LFW (b) Ours (c)[40] (d)[12]

图 7. LFW 数据集的综合结果。注意 TP-GAN 是在 Mulit-PIE 上训练的。

4.2. 身份保持能力

人脸识别 为了定量地展示我们方法的身份保持能力，我们使用两种不同的设置对 MultiPIE 进行面部识别。首先通过 Light-CNN [35]提取深度特征，然后将 Rank-1 识别精度与余弦距离度量进行比较，进行实验。配置文件图像 IP 上的结果作为我们的基线，并在所有表格中用符号 Light-CNN 标记。应该注意的是，尽管已经提出了许多用于正面视图合成的深度学习方法，但是它们的合成图像都没有被证明对于识别任务是有效的。在最近一项关于面部图像的研究[34]中，作者表明直接使用 CNN 合成的高分辨率人脸图像进行识别肯定会使性能退化而不是改善它。因此，验证我们的综合结果是否可以提高识别性能（“通过生成识别”程序是否有效）具有重要意义。

在设置 1 中，我们遵循[36]中的协议，并且仅使用来自会话 1 的图像。我们包括 20 个照明下的中性表达和±90°内 11 个姿势的图像。每个测试对象使用一个具有正面视图和照明的图库图像。训练和测试集之间没有重叠。表 1 显示了我们的识别性能以及与最新技术的比较。 TP-GAN 始终在所有角度都达到最佳性能，角度越大，改进越大。与 c-CNN Forest [36]（三个模型的集合）相比，我们在大型姿势案例中实现了约 20%的性能提升。

在设置 2 中，我们遵循[38]中的协议，其中使用来自所有四个会话的神经表达图像。从第一次出现时为每个测试标识选择一个图库图像。 MultiPIE 的所有合成图像

表 2.设置 2 下的视图，照明和会话的 Rank-1 识别率（%）

Method	±90°	±75°	±60°	±45°	±30°	±15°
FIP+LDA [41]	-	-	45.9	64.1	80.7	90.7
MVP+LDA [42]	-	-	60.1	72.9	83.7	92.8
CPF [38]	-	-	61.9	79.9	88.5	95.0
DR-GAN [30]	-	-	83.2	86.2	90.1	94.0
Light CNN [35]	5.51	24.18	62.09	92.13	97.38	98.59
TP-GAN*	**64.64**	**77.43**	**87.72**	**95.38**	**98.06**	**98.68**

表 3.不同视图和照明的性别分类准确度（%）

Method	±45°	±30°	±15°
I^P_{60}	85.46	87.14	90.05
CPI* [38]	76.80	78.75	81.55
Amir *et al.* * [8]	77.65	79.70	82.05
I^P_{128}	86.22	87.70	90.46
Hassner *et al.* * [12]	83.83	84.74	87.15
TP-GAN*	**90.71**	**89.90**	**91.22**

在本文中是来自于 Setting2 下的测试身份.结果如表 2 所示。请注意，所有基于 CNN 的比较方法都具有学习中间特征的最佳性能，而我们在“通过生成识别”程序后直接使用合成图像。

性别分类 为了进一步证明我们合成图像对其他面部分析任务的潜力，我们进行了性别分类实验。本部分中的所有比较方法也遵循“通过生成识别”程序，我们直接使用它们的同义词，论文结果为性别分类。用于性别分类的 CNN 与编码器 $G_{\theta^g_E}$具有相同的结构，并且在 UMD [3]数据集的批处理 1 上训练。

我们在表 3 中报告了 Multi-PIE（设置-1）的测试性能。为了公平比较，我们在两个分辨率中显示未旋转的原始图像的结果。分别为 128×128 （I128P）和 60×60 （I60P）。TP-GAN 的合成实现了更好的分类精度，由于标准化视图，比原始配置文件图像。所有其他比较模型的性能都比基线差，这并不奇怪，因为它们的架构不是为性别分类任务而设计的。在[34]中观察到类似的现象，其中合成的高分辨率面部图像严重地使识别性能退化而不是改善它。这表明在操纵像素空间中的图像时失去 IP 的突出面部特征的高风险。

4.3. 特征可视化

我们使用 t-SNE [31]来显示二维空间上的 256 维深度特征。图 8 的左侧示出了原始轮廓图像的深特征空间。很明显，在 Light-CNN 跨越的深度特征空间中，具有大姿态（特别是 90°）的图像是不可分离的。它揭示了即使 Light-CNN 训练有数百万张图像，它仍然无法正常使用

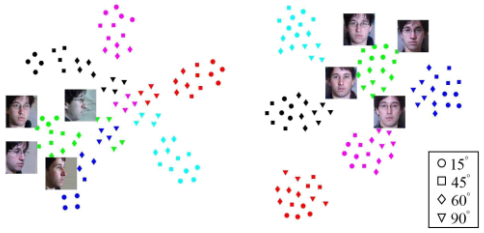

图 8.轮廓面（左）和 fontal 视图合成图像（右）的特征空间。每种颜色代表不同的标识。每个形状代表一个视图。标记一个身份的图像。

表 4.模型比较：设置 2 下的 Rank-1 识别率（%）。

Method	±90°	±75°	±60°	±45°	±30°	±15°
w/o P	44.13	66.10	80.64	92.07	96.59	98.35
w/o L_{ip}	43.23	56.55	70.99	85.87	93.43	97.06
w/o L_{adv}	62.83	76.10	85.04	92.45	96.34	98.09
w/o L_{sym}	62.47	75.71	85.23	93.13	96.50	98.47
TP-GAN	**64.64**	**77.43**	**87.72**	**95.38**	**98.06**	**98.68**

处理大型姿势人脸识别问题。在右侧，使用我们的 TP-GAN 进行正面视图合成后，生成的正面视图图像可以根据其身份轻松分类到不同的组中。

4.4. 算法分析

在本节中，我们将介绍不同的体系结构和损失函数组合，以深入了解它们在正面视图合成中的相应角色。定性视觉化结果和定量识别结果都需要进行全面比较。

我们比较了本节中 TP-GAN 的四种变体，一种用于比较体系结构，另一种用于比较目标函数。具体而言，我们训练没有局部路径（表示为 P）的网络作为第一变体。关于损失函数，我们保持双路径结构完整并且在每种情况下去除三个损失中的一个，即 Lip，Ladv 和 Lsym。

表 4 中报告了详细的识别性能。双路径架构和身份保持损失对提高识别性能贡献最大，特别是在大型姿势案例中。虽然不那么明显，但对称性损失和对抗性损失都有助于提高识别性能。图 9 说明了这些变体的感知性能。正如预期的那样，没有身份保留损失或局部路径的推断结果严重偏离了真实的外观。没有对抗性损失的合成趋于非常模糊，而没有对称性损失的结果有时会显示出不自然的不对称效应。

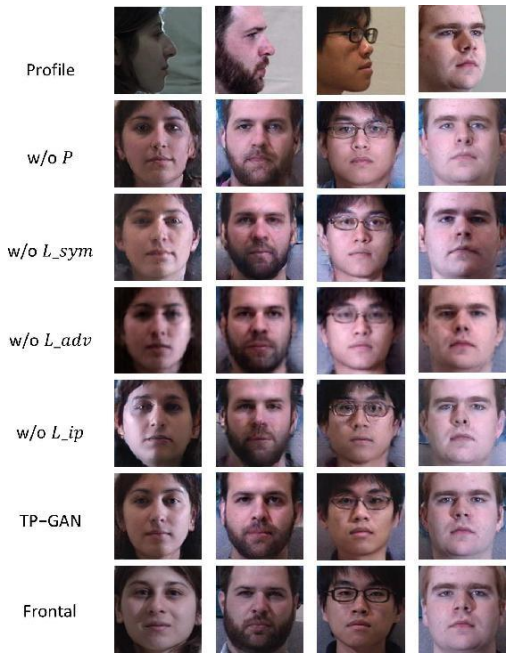

图 9.模型比较：TP-GAN 及其变体的合成结果。

5. 结论

在本文中，我们从单个图像中提出了用于正面视图合成的全局和局部感知 GAN 框架。该框架包含两个独立的路径，分别建模全局结构的平面外旋转和局部纹理的非线性变换。为了使不适定合成问题得到很好的约束，我们在训练过程中进一步引入了对抗性损失，对称性损失和身份保持损失。对抗性损失可以忠实地发现和引导综合论文驻留在正面的数据分布中。在大型姿势情况下，在减轻自遮挡效应之前，对称性损失可以明确地利用对称性。此外，身份保护损失被纳入我们的框架，因此合成结果不仅在视觉上具有吸引力，而且还易于应用于准确的人脸识别。实验结果表明，我们的方法不仅具有引人注目的感知结果，而且在大型姿势人脸识别方面也优于最先进的结果。

致 谢

这项工作部分由国家自然科学基金（批准号 61622310,61473289）和国家重点发展计划（批准号 2016YFB1001001）资助。我们感谢吴翔的有益讨论。

参考文献

[1] M. Abadi et al. Tensorflow: A system for large-scale machine learning. In *OSDI*, pages 265–283, 2016. 6

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 2006. 1

[3] A. Bansal, A. Nanduri, R. Ranjan, C. D. Castillo, and R. Chellappa. Umdfaces: An annotated face dataset for train-ing deep networks. *arXiv:1611.01484*, 2016. 7

[4] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimension-ality: High-dimensional feature and its efficient compression for face verification. In *CVPR*, 2013. 1

[5] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA*, 1985. 1

[6] E. L. Denton, S. Chintala, R. Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *NIPS*, 2015. 2

[7] C. Ding and D. Tao. Pose-invariant face recognition with homography-based normalization. *Pattern Recognition*, 66:144 – 152, 2017. 6

[8] A. Ghodrati, X. Jia, M. Pedersoli, and T. Tuytelaars. Towards automatic image editing: Learning to see another you. In *BMVC*, 2016. 5, 6, 7

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Gen-erative adversarial nets. In *NIPS*, 2014. 2, 4

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image Vision Computing*, 2010. 6

[11] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV*, 2016. 6

[12] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In *CVPR*, 2015. 2, 3, 5, 6, 7

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007. 6

[14] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, 2016. 2, 5

[15] M. Kan, S. Shan, H. Chang, and X. Chen. Stacked progressive auto-encoders (spae) for face recognition across poses. In *CVPR*, 2014. 2, 3

[16] D. P. Kingma and M. Welling. Auto-encoding variational bayes. In *ICLR*, 2014. 2

[17] C. Ledig et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 2, 3

[18] C. Li and M. Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *CVPR*, 2016. 3

[19] M. Mirza and S. Osindero. Conditional generative adversar-ial nets. *arXiv:1411.1784*, 2014. 3

[20] U. Mohammed, S. J. Prince, and J. Kautz. Visio-lization: generating novel facial images. In *TOG*, 2009. 4

[21] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, 2016. 2, 3

[22] S. Reed, K. Sohn, Y. Zhang, and H. Lee. Learning to disentangle factors of variation with manifold interaction. In *ICML*, 2014. 2

[23] D. J. Rezende, S. Mohamed, and D. Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *ICML*, 2014. 2

[24] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic. Robust statistical face frontalization. In *ICCV*, 2015. 2

[25] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A uni-fied embedding for face recognition and clustering. In *CVPR*, 2015. 1

[26] W. Shen and R. Liu. Learning residual images for face attribute manipulation. In *CVPR*, 2017. 3

[27] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. In *CVPR*, 2017. 3, 4

[28] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *CVPR*, 2014. 1, 2

[29] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verifica-tion. In *CVPR*, 2014. 1, 2

[30] L. Tran, X. Yin, and X. Liu. Disentangled representation learning gan for pose-invariant face recognition. In *CVPR*, 2017. 4, 5, 7

[31] L. van der Maaten and G. E. Hinton. Visualizing high-dimensional data using t-sne. *JMLR*, 2008. 7

[32] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, 2001. 1

[33] K. Q. Weinberger and L. K. Saul. Distance metric learn-ing for large margin nearest neighbor classification. *JMLR*, 2009. 1

[34] J. Wu, S. Ding, W. Xu, and H. Chao. Deep joint face hallucination and recognition. *arXiv:1611.08091*, 2016. 7

[35] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *arXiv:1511.02683*, 2016. 5, 6, 7

[36] C. Xiong, X. Zhao, D. Tang, K. Jayashree, S. Yan, and T. K. Kim. Conditional convolutional neural network for modality-aware face recognition. In *ICCV*, 2015. 6, 7

[37] J. Yang, S. E. Reed, M.-H. Yang, and H. Lee. Weakly-supervised disentangling with recurrent transformations for 3d view synthesis. In *NIPS*, 2015. 3

[38] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim. Rotating your face using multi-task deep neural network. In *CVPR*, 2015. 2, 3, 5, 6, 7

[39] Z. Zheng, L. Zheng, and Y. Yang. Unlabeled samples gener-ated by gan improve the person re-identification baseline in vitro. *arXiv:1701.07717*, 2017. 3

[40] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li. High-fidelity pose and expression normalization for face recognition in the wild. In *CVPR*, 2015. 2, 3, 5, 6, 7

[41] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning identity-preserving face space. In *ICCV*, 2013. 1, 2, 3, 6, 7

[42] Z. Zhu, P. Luo, X. Wang, and X. Tang. Multi-view perceptron: a deep model for learning face identity and view repre-sentations. In *NIPS*, 2014. 1, 2, 3, 7