# D2SC-GAN：双深浅通道生成对抗网络

用于在课堂场景中识别的低分辨率面孔

**摘要：**近年来，在受限的环境中，使用卷积神经网络的人脸识别取得了相当大的成功。但是，在教室/监控场景下，如果训练和测试集分配不匹配，这些方法的性能会下降。与清晰，丰富的训练（图库）集相比，这些测试（probe）样本的质量下降，例如噪声，照明差，姿势变化，遮挡，低分辨率（LR），模糊和混叠，其中大部分由以下内容组成：在实验室环境中捕获的高分辨率（HR）人脸照片。为了应对这种情况，我们提出了一种新型的双深浅通道生成对抗网络（D2SC-GAN），该网络通过将LR降级的probe样本映射到其对应的HRgallery样对应物以执行封闭域人脸识别，来执行监督域自适应（DA）。D2SC-GAN使用包含**多分辨率逐点MSE**和**归一化卡方距离损失**函数的多成分损失函数，以及基于**Kullback-Leibler发散的损失函数**。此外，我们**提出了一种新颖的教室人脸数据集**，称为“印度教室人脸数据集”（ICFD），据我们所知，这是同类中的第一个，将有助于探索人脸识别在用于自动记录时所面临的挑战课堂条件的出勤率。

**索引**-人脸识别，人脸生成，生成对抗网络，卷积神经网络，教室FR。

## I.简介

F

ACE识别（FR）是计算机视觉（CV）和深度学习（DL）领域中一个广为人知的问题。尽管针对FR [1]，[2]，[3]的常规解决方案

[4]，[5]，[6]在受限的情况下效果很好，但是在训练（gallery）和测试（probe）数据没有相似分布的情况下（例如在教室和监视环境中），它们的表现不佳。通过采用域自适应（DA）的概念，已经进行了各种尝试[7]，[8]，[9]，[10]来解决上述问题。但是，这些方法的性能留下了很大的改进空间。尽管最近的基于DL的FR算法[11]，[12]，[13]，[14]，[15]，[16]即使在野外情况下也能达到非常高的精度，但在这种情况下它们仍无法很好地执行真实场景。

深度卷积GAN [17]（DCGAN）作为卷积体系结构首次引入，从而改善了计算机视觉（CV）应用程序中的视觉质量。BEGAN模型[18]利用一种新的均衡执行方法，结合从Wasserstein距离得出的损耗，来训练基于自动编码器的生成对抗网络。它还提供了一种新的近似收敛度量，快速稳定的训练以及较高的视觉质量。尽管早期的GAN变体缺乏分析收敛性的方法，但Wasserstein GAN [19]（WGAN）最近引入了一种损失函数，作为收敛性的一种度量。然而，在实施过程中，这是以缓慢的训练为代价的，但是却具有稳定性和更好的覆盖模式的好处[19]。Isola提出的图像到图像转换方法[20]*等。*，利用条件GAN架构从以潜伏向量输入为条件的输入图像生成清晰图像。Choi*等*人提出的StarGAN模型[10] *。*通过使用域区分符在多个域之间执行图像到图像的转换。

Zhao*等人*提出的DA-GAN模型*。*[21]，[22]使用双重代理对人脸识别任务执行人脸模拟器（3D人脸模型）输出的身份保留优化。Shen*等人*提出的FaceID-GAN模型*。*[23]利用附加的分类器作为第三玩家，以各种姿势角度合成了保存身份的面孔，以进行输入面孔验证。然而，在[21]，[22]和[23]中提出的模型通过首先从输入轮廓人脸合成正面来执行姿势不变的人脸识别，而我们提出的方法还处理了其他形式的退化，例如模糊，照明度差。和遮挡，这需要将域适应功能纳入我们提出的框架。因此，与上述工作中提出的方法相比，我们的方法本质上执行的任务不同且难度更大。

Sun*等人*提出的紧凑型网络*。*[14]，[24]使用了25个网络的集合，每个网络都在一个不同的人脸补丁上运行。Schroff*等人*提出的FaceNet模型*。*[13]基于三元组挖掘方法制定的三元组损失，使用深层的CNN直接优化嵌入（功能）本身。由Parkhi*等人*提出的DeepFace *。*[12]使用暹罗网络来训练大规模的人脸数据集。陈*等人*所做的工作*。*文献[25]中的研究表明，深度神经网络学习的非线性表示可以为数据样本之间的不同变化提供不变性。在该领域的其他重要著作是[11]，[26]，[27]，它们使用CNN进行对象识别任务。Banerjee和Das [28]提出的SMLMFKC方法采用了一种最佳的特征核组合来适应目标域。Long和Wang [29]提出的深度适应网络（DAN）通过最小化源域和目标域之间的距离来执行域自适应。Chen*等人*提出的FV\_DCNN [30] *。*将Fisher向量表示法与Deep-CNN功能相结合，可实现可靠的人脸识别和验证任务。[31]中的工作提出了一个使用暹罗网络的统一监督域自适应框架，该框架可找到一个子空间，该子空间在语义上对齐源和目标分布，并最大程度地分离其类别。Banerjee和Das [8]提出的工作采用了一种新颖的三阶段互斥训练算法来进行深域自适应，以解决降级FR的问题。文献[9]中提出的基于GAN的方法利用一种新颖的多尺度重构损失将目标域样本映射到源域以执行FR。

江*等人*的工作*。*[32]采用缺失强度插值方法，该方法使用具有局部结构先验（SRLSP）的平滑回归来执行人脸图像的超分辨率（SR）。Zhang*等人*提出的超身份卷积神经网络（SICNN）方法*。*文献[33]通过使用超身份损失生成高分辨率（HR）人脸，其身份接近于输入低分辨率人脸的真实身份，从而执行人的幻觉。李*等人*提出的方法*。*[34]利用GAN预训练方法进一步改善了现有人脸识别算法的性能。Lu*等人*提出的深度耦合ResNet（DCR）模型*。*[35]利用单个主干和两个分支网络来最小化高分辨率和低分辨率人脸图像的特征分布之间的差异。

在本文中，我们提出了一种新的基于生成对抗网络的有监督域自适应方法，称为双深浅通道生成对抗网络（D2SC-GAN），**该方法首先从LR生成HR（140 × 140 px。）gallery样的样本。 （35 ×35 px。）probe样本作为输入，然后使用辅助鉴别器（在封闭设置FR设置中）[9]，[36]进行识别。**所提出的模型利用了新颖的双通道生成器（浅-深对）网络，其中**浅通道和深通道分别捕获了LR（输入）图像的低频和高频分量，以生成相应的类似HRgallery的图像。**图像样本。**捕获两个频率范围**可确保在创建超分辨人脸样本的同时，**改善对比度并增强probe样本区分部分的特征**。为了实现这一任务，网络最小的新颖的“多分辨率重建（MRR）”损失函数，包括“多分辨率Patchwise均方误差（MR\_PMSE）”和归一化χ 2距离（NCD）损耗分量。生成器网络还将目标函数中基于“ Kullback-Leibler散度（KLD）”的损耗分量的负值最小化。根据其公式，**MR\_PMSE可以捕获probe和probe集之间的整体差异以及像素级差异**。KLD损失可确保这对生成器**在两个通道上正确捕获**输入的教室/监控人脸图像中**的不同频率分量**。最后，**NCD损失可确保在生成的样本中进行正确的色彩还原**。

此外，我们还提出了印度教室人脸数据集（ICFD），该数据集由在受控实验室（用于训练）和不受约束的教室设置（用于探查）中捕获的84位受试者的图像组成。据我们所知，该数据集尚属首次，它将帮助研究人员探索在不受限制的教室环境中进行人脸识别所面临的挑战，以达到教室出勤和搜索的目的。在其他4个现实世界的人脸数据集上进行的大量实验和消融研究表明，我们提出的D2SC-GAN模型的有效性。

## **二 PROPOSED METHOD**

所提出模型的总体架构如图1所示。**该网络由两个生成器（浅生成器和深生成器）和一个鉴别器组成**。使用两种生成器架构的动机来自[37]，它表示浅层网络固有地捕获了图像的**低频成分**，而深层网络固有地捕获了图像中的细节**（高频成分）**。因此，与使用单个生成器网络相比，这两个网络的组合在用于图像SR时可提供更好的结果。我们使用两个生成器模块，即浅生成器*G S*和深生成器*G D*来利用此思想，目的是从输入LR probe人脸重建HRgallery 的人脸图像。鉴别器*D*执行两个任务，识别输入图像是假的还是真实的（*即*生成的或真实的），并对输入图像进行分类以获得类（ID）。随后将详细说明网络的各个组成部分。

**浅生成器模型（GS）示于图2**。每个层的细节在彩色编码图中给出。所述*CONV* \_2 *d* \_ *Ñ* [ （*˚F*，*˚F*）， ＃*˚F* ]表示*Ñ*个具有二维卷积层*˚F*的数目*˚F* × *˚F*滤波器。*DENSE* \_ *N* [ *D* ]表示有*D个*节点的*第N个*线性层。RESHAPE[ （*m*，*n*），＃*N* ]操作将前一层输出重塑为形状为R *m* × *n* × *N*的张量，而*FLATTEN*操作将前一层输出转换为一维张量。*UPSAMPLING* \_2*D*操作对上一层输出执行unpolling，并在[38]中进行了详细说明。

|  |
| --- |
| 图1.提出的D2SC-GAN方法的总体架构。*我P*指的是作为输入提供给所述probe的图像*ģ小号*和*G ^ d*，其为浅和深生成器网络。对这两个网络的输出求平均值，得到*I G*，然后将其传递给鉴别器*D*，鉴别器*D*将真实（*I Gal*）与伪/合成图像（*I G*）以及生成图像的分类区分开。 |

图2 浅生成器的架构细节*ģ小号*。

深度生成器模型（*GD*）的体系结构如图3所示。该图中使用的符号已在上一段中进行了描述。**有目的地给该体系结构的明显更大的深度，因为需要利用它使用高频分量来捕获人脸图像的更精细的细节。深层和浅层网络分别有助于捕获生成图像的高频和低频分量，从而确保了所提出模型的出色重建性能。**

可能需要注意的是，建议的生成器架构模型应视为一个概念，研究人员可以探索文献[39]，[40]中可用的生成器架构的不同版本，以寻求更好的结果。经过严格的实验后，凭经验确定了此处介绍的体系结构细节。

鉴别器模型（*D*）的架构如图4所示。如图所示的*MAXPOOLING* \_2 *D* \_ *N* [ *H*，*W* ]在大小为*H* × *W*的窗口上执行第*N个*maxpool操作。

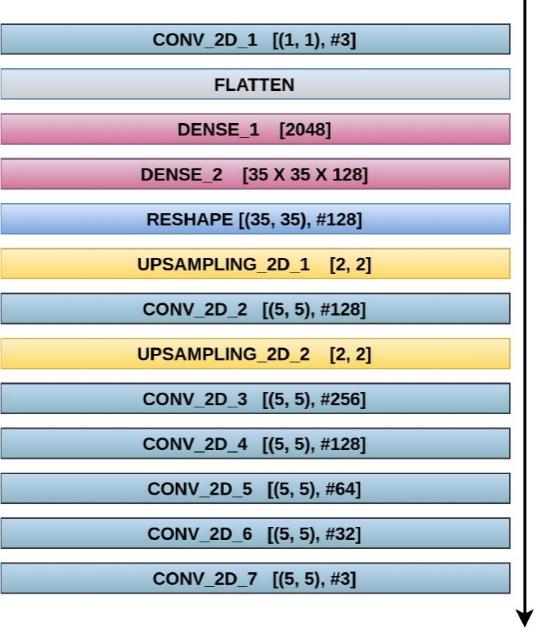


图3. 深度生成器模型*G D的*体系结构细节。

网络执行两项任务：（i）将输入图像识别为真实或伪造（对应于输出分支*D bin*），以及（ii）将输入图像分类为相应的类别（*D cls*）。

## 三，LOSS ˚FUNCTIONS FOR D2SC-GaN

D2SC-GAN模型可将训练过程中的多目标损失函数降至最低。以下各节介绍了各个损失函数，然后提出了所提出的D2SC-GAN模型的整体损失函数。

### *A.分类交叉熵损失*

分类交叉熵[41]损失函数用于训练具有S型输出的深层神经网络

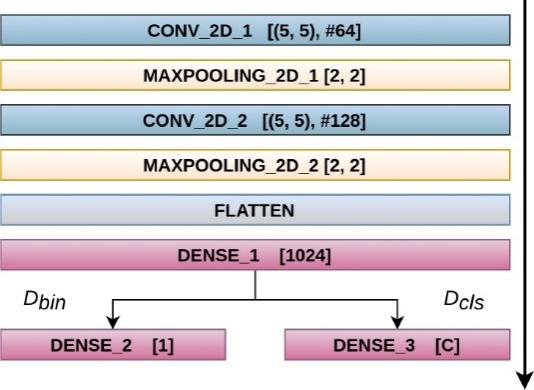


图4.辅助鉴别器模型*D的*体系结构细节。它执行从生成的（伪）图像中识别真实图像的任务，并将生成的图像分类为相应的标识。

算法1：多分辨率重建（MRR）

失利

L *ncd*：= NCD（*h*（*I G*），*h*（*I Gal*））L *mrp*：= MR\_PMSE（*I G*，*I Gal*）

function

L

*mrr*

(

*I*

*B*

*G*

,

*I*

*B*

*Gal*

)

B

←

20

;

L

*mrr*

←

0

foreach

{

*I*

*G*

,

*I*

*Gal*

}

*in*

{

*I*

*B*

*G*

,

*I*

*B*

*Gal*

}

do

*mrr*

:

=

L

*mrr*

+

+

L

2

L

*avg*

:

=

L

*mrr*

*BATCH*

\_

*SIZE*

return

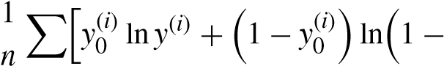
L

*avg*

L  L *NCD MRP*

分类。此损失函数定义为：

*ñ*

大号*CXE ÿ*

*我*= 1

### （1）

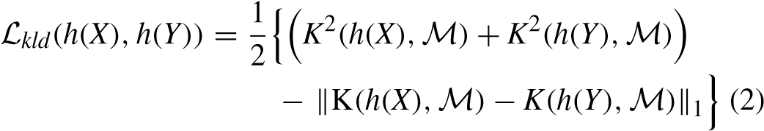
其中，*y* = { *y* （1 ），...，*y* （*n*） }是与输入图像相对应的神经网络的预测S形输出，并且y 0 = { *y* （0 1 ），...，*y* （0 *n*） }是相同输入样本的相应实际标签。

#### B.多分辨率重建（MRR）损失

我们提出了一种新颖的多分辨率重建（MRR）损失，该损失使生成器能够执行从输入的低分辨率（LR）probe样本生成类似gallery的清晰高分辨率（HR）图像的任务。该损失函数的目的是通过关**注细节和生成的人脸的整体结构**来确保人脸的精确重建。评价所提出的损失的轮廓在算法1中示出*我ģ乙* 和*我半乳糖乙*指具有批量大小所产生的和对应的地面实况gallery图像*乙*。批次的单个图像用*I G*和 *I Gal表示*分别。MR\_PMSE损失函数（稍后在III-E节中定义）通过合并新颖的多分辨率方法，**确保保留了微小的细节以及整个人脸布局。**逐通道应用NCD损失（第III-D节），可确保在生成的图像中正确显示颜色，与图库图像一致。

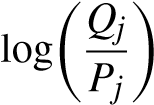
#### C.基于Kullback-Leibler散度的损失

由于拟议的架构由浅层和深层网络组成，分别捕获图像的低频和高频分量，因此我们提出了基于Kullback-Leibler散度（KLD）的损耗函数，该函数被最大化（等效于最小化损耗的负值） 。尽管流行的Jensen-Shannon散度损失可确保两个通道都学习图像的不同方面（不知道两个通道是否实际上都参与了重建任务），但建议的损失函数的性质可确保两个通道均具有相同的生成参与度从输入的LRprobe图像中提取类似HRgallery的样本，利用浅层模型和深层模型的固有特性分别捕获图像的低频分量和高频分量。损失函数为：



其中， *X*和Y表示两个图像（计算之前变换为灰度），h（。）返回64维输入图像的直方图，M =√*ħ*（*X*） *ħ*（*Y*）其中

。表示两个矢量的逐元素乘法运算符，Kullback-Leibler发散[42]度量*K*（·，·）定义为：

*大号*

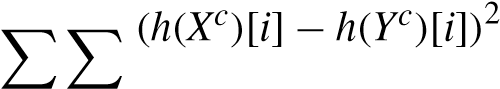
### *K P j*（3）

*j* = 1

其中，*P*和*Q*表示来自两个不同分布且向量长度*L*相同的向量。

#### D.标准化χ 2距离损失

从经验上已经发现，由MR\_PMSE损失产生的图像并不总是颜色正确的，具有肤色和头发颜色的变化，这在任何产生的图像中都是不希望的。为了解决这个问题，我们提出了标准化χ 2距离（NCD）的损失。χ 2距离[43]，[44]是用于查找直方图之间的差的公认有用的的度量。我们通过采用这种度量以产生色彩校正的人脸图像，最小化生成图像和GT图像的l2标准化逐通道直方图之间的χ 2距离。损失定义为：

1 | *C* | | *b* | L *ncd*（*h*（*X*），*h*（*Y*））= 2 | *b* || *C* | （*h*（*X c*）[ *i* ] + *h*（*Y c*）[ *i* ] ）*c* = 1 *i* = 1

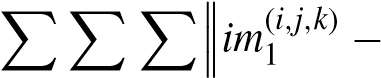
### （4）

其中，| *C* | 表示输入图像中通道的数量，*X c*和*Y c*表示两个图像*X*和 *Y*的通道*c*，*h*（·） 返回| *b* | 输入图像通道的二合二归一化直方图，并且*h*（·）[ *i* ]是|中。*b* | -bind直方图的*第i个*元素。

#### E.逐块的多分辨率MSE损失

|  |
| --- |
| 图5.来自ICFDprobe集的少量样品表现出：（a）对比度/分辨率低的照明差，（b）模糊，（c）遮挡和（d）数据集中存在的姿态变化。红色轮廓中的样本表示使用移动（三星S9）相机拍摄的人脸。该图最好用彩色查看。 |

受到[9]提出的逐块MSE（PMSE）损失的启发，我们引入了一种新颖的多分辨率逐块MSE（MR\_PMSE）损失，其中我们从生成的和地面真相图像以计算它们之间的均方误差。值得注意的是，虽然一对小补丁之间的MSE可以确保保留图像的细节，但是大补丁之间的MSE可以保留生成的人脸的整体布局。两个图像*im* 1和 *im* 2之间的MSE计算为：

| *C* | | *即时通讯*1 | | *即时通讯*1 | MSE （*im* 1，*im* 2）= *im*

*i* = 1 *j* = 1 *k* = 1

其中，| *C* | ＆| im1 | 指定输入图像的高度和宽度上的通道数和像素数（假设两者出于实验目的是相同的）。*im* （*i*，*j*，*k*）是指位于图像*im*通道*i*中位置（*j*，*k*）的像素。补充文档的附录A中详细介绍了评估MR\_PMSE以及使用的超参数值的概述。

## IV。Ø VERALL大号OSS ˚F UNCTIONS的

P ROPOSED D2SC-GaN

提议的D2SC-GAN的生成器将以下损耗函数降至最低：

*副词*

L *G*（*I P*，*I Gal*，*I G* 1 ，*I G* 2 ，*I G*）

= +

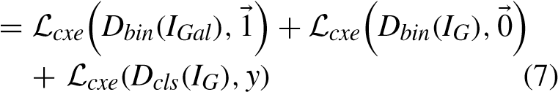
+ L *mrr*（*I G*，*I Gal*）− L *kld*（*h*（*I G* 1 ），*h*（*I G* 2 ）） （6）

其中，*我P*指示用作输入到probe样品*ģ小号*和 *G ^ d*与*y*作为对应的标签，*我加*表示数据集的库样品，*我ģ* 1 和*我ģ* 2是从两个生成器所得到的输出*ģ S*和*G D*分别（以*I P*作为输入），而它们的元素平均值由*I G*表示，其后称为增强probe。L *mrr*（·，·）在算法1中定义了L *kld*（·，·），而在方程式（2）中预先定义了L *kld*（·，·），在方程式（1）中给出了L *cxe*（·，· ）。鉴别器的二进制输出分支表示为*D bin*，而*D cls*指鉴别器*D*的分类器分支。公式（6）中的L *cxe*损失可确保生成的人脸类似于gallery图像。的大号*MRR*损失确保生成脆的高分辨率的gallery般从低分辨率probe像的图像。最后，L *kld的负数* 损耗确保了这两个生成器平等地参与捕获输入图像的不同分量，如之前在III-C节中讨论的那样。



图6.来自ICFDgallery集的少量样本显示了偏航角和俯仰角的变化。样品中的1*日* 在单独偏航行示出变型中，而那些在2*个*在偏航以及桨距角行展品变化。该图最好用彩色查看。

D2SC-GAN模型的判别器将总损耗函数最小化为：

*副词*

L *D*（*I Gal*，*I G*）

其中，损失函数中的前两项确保鉴别器从合成生成的人脸正确识别原始图像，而后一项确保在封闭设置FR设置中对生成的图像进行正确分类。

## V.我NDIAN Ç LASSROOM ˚F ACE d ATASET （ICFD）

随着深度学习的到来，以及对大型训练数据集的需求，我们已经看到了大规模人脸表情数据集的兴起。但是，据我们所知，**尚未创建任何数据集来探讨在无限制的教室/研讨会室场景中进行人脸识别时所面临的挑战**。**我们建议使用“印度教室人脸数据集”来解决此问题。**

拟议的数据集由84位受试者组成，其中63位男性和21位女性组成，具有17.8万张图像，这些图像分为两类：gallery和探测。对于gallery场景，在受控的实验室环境中以适当的照明条件捕获对象的人脸图像，而没有任何形式的现实世界干扰。受试者坐在距尼康D7200 DSLR相机约2米的距离处，并被要求移动其头部以各种俯仰和偏航角捕获图像，如图6所示。gallery样本的平均分辨率为400 × 350像素（使用标准的现成人脸检测器[45]裁剪），每个对象平均拥有近900个样本。

使用同一台数码单反相机和三星S9移动相机在志愿者指导下，在各个教室和研讨室中采集了probe样本。

|  |
| --- |
| 图7. 对于（a）ICFD，（b）FR\_SURV [46]，最上面一行的probe样本以及我们下面提出的方法产生的相应结果。  （c）TIPD [4]，（d）关键点[47]和（e）ScFace [48]数据集。红色框表示生成的人脸身份与 |

相应的probe样品。该图最好用彩色查看。

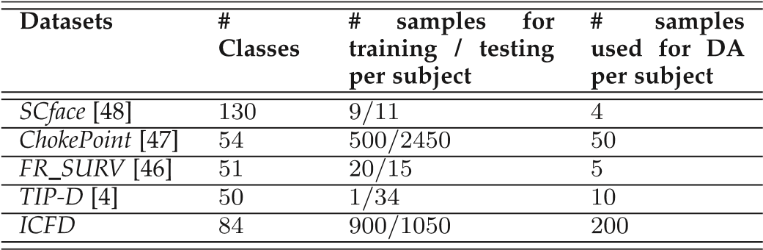
凝视着前面的dais / screen，以创建一个逼真的教室场景。使用（i）三星S9移动相机拍摄了两种类型的视频，该相机在观众面前横越了dais，并且（ii）在平移模式下在dais中央的dais中心使用了DSLR相机。观众处于适当的高度。在教室条件下获取一部分测试probe（42个受试者）的视频时，使用了两种照明条件：（i）在所有照明均打开的情况下，教室内可能达到的最大照明度；以及（ii）所有房间照明均在低照明条件下关闭电源，将空的白色幻灯片投影到观众面前的投影仪屏幕上，在这种情况下，它是唯一的光源。购置的更多细节在[49]中给出。教室/研讨室的不受约束的性质导致各种劣化的存在，例如照明和对比度差（图5（a）），模糊，这是由于所捕获图像的低分辨率以及头部运动引起的志愿者（图5（b）），部分遮挡（图5（c））和姿势变化（图5（d））在捕获的图像中显示。使用MTCNN [45]进行裁剪后，probe样本的平均分辨率为110× 80像素，每个对象平均有大约1250个样本，这些样本合并了在不同设置下捕获的所有图像。补充文件的附录B中提供了来自数据集的更多样本以及每个分区的统计信息。数据集的一部分可以从[49]中下载。

## 六。Ë XPERIMENTAL小号ETUP和[R ESULTS

实验是在装有Dual-Xeon处理器和256GB RAM，Nvidia Tesla V100（32GB）的计算机上进行的。所有实现均使用Tensorflow（1.8.0）-后端在Keras 2.1.6平台中进行了编码。模型权重全部随机初始化，需要在GPU上进行30-72小时的训练。批处理大小保持为20，并使用双三次插值法将图库和probe图像的大小调整为140 × 140和35 × 35像素

## 表一

### d ATASET P ARTITION小号战略研究FOR甲LL ˚F IVE ř EAL -W ORLD ˚F ACE d ATASETS ü SED FOR Ť下雨，Ť ESTING ，A LONG Wˉˉ ITH THE Ñ的棕土Ç LASSES IN Ë ACH d ATASET 。V ALUES d ø Ñ OT我NCLUDE Ť HAT ü SED FOR d ATA甲UGMENTATION TO牛逼RAIN D2SC-GaN



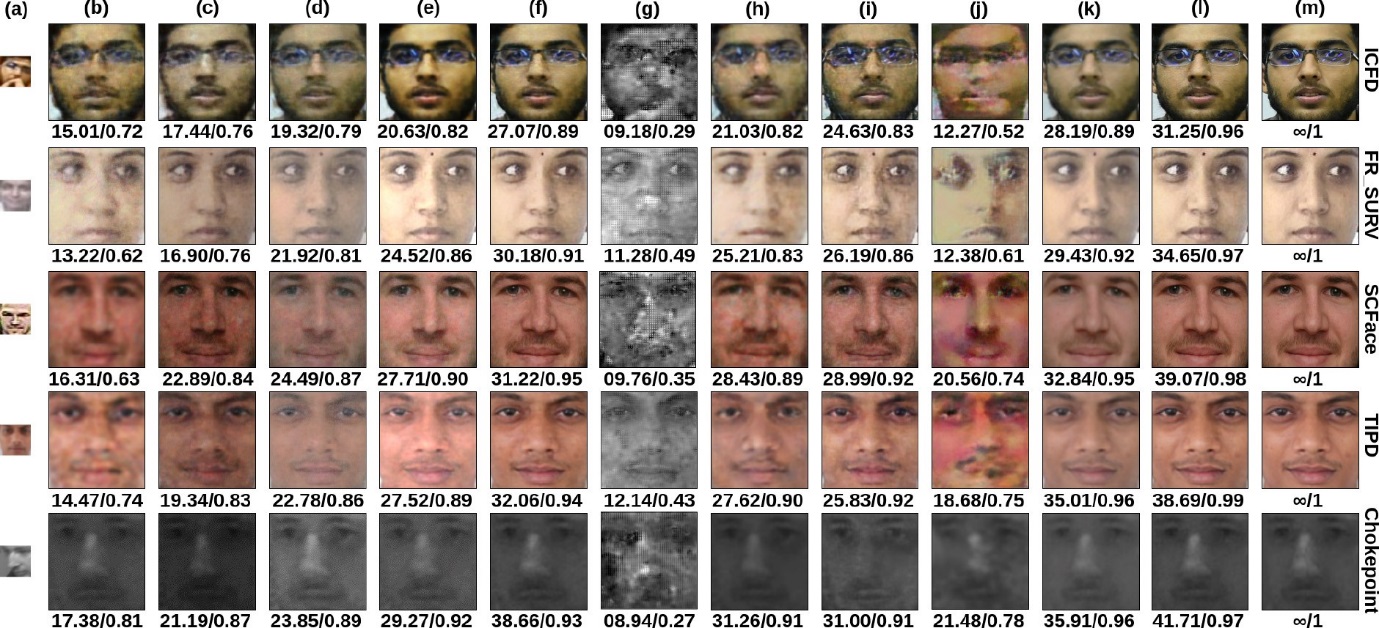
所有数据集。以下各节讨论用于实验的设置以及随后的实验结果。

### *A.实验设置*

我们对5个现实世界的人脸数据集进行了实验：扼流点[47]，Scface [48]，TIPD [50]，FR\_SURV [46]和建议的ICFD。为了展示所提出方法的优越性，我们在两种设置下进行了实验。在第一种情况下，我们展示了使用我们提出的模型的生成器作为预处理步骤优于常规（非DA）方法的优势。我们比较了由于使用我们提出的方法而不是通用probe样本生成的增强probe来训练非DA方法而导致的性能跳跃。为了进行实验，我们使用整个图库集和每个受试者15％的probe（或增强型probe）样本来训练这些模型，而将其余的probe（或增强型probe）集用于测试目的。

在第二种情况下，我们通过模型性能的定性（增强的probe生成性能）和定量的（Rank-1识别/鉴定率）比较展示了我们方法的优越性，并提供了几种最新技术监督域自适应（DA），超分辨率（SR）和降级FR（DFR）技术。正如在受监管的DA中普遍存在的那样

|  |
| --- |
| 图8.用（b）DCGAN [17]，（c）BEGAN [18]，（d）Isola*等*通过我们的方法生成的人脸比较*。*[20]，（e）StarGAN [10]，（f）LR-GAN [9]，（g）SRLSP [32]和（h）SICNN [33]，我们的方法（l）在五个真实数据集上。（a）表示输入图像，（i），（j）和（k）显示采用（i）MR\_PMSE +对抗性 +分类损失，（j）NCD +对抗性 +分类损失和（k ）时的消融研究及其生成结果）分别为MR\_PMSE + NCD +对抗+分类损失，而（m）显示对应的地面真相gallery面孔。最好看的图 |

颜色。

在实验设置[9]，[51]中，我们使用完整的图库集以及约15％的probe集（每个对象）进行域适应。我们也遵循相同的协议来训练SR和DFR模型。用于实验目的的数据拆分显示在表I中。但是，应该注意的是，我们需要成对的数据来训练D2SC-GAN，其中每对（< gallery，probe > ）都属于同一身份。为了训练D2SC-GAN，将probe样本馈送到网络，这会将新的MRR损失降至最低，从而生成类似gallery的图像。

### *B.实验结果*

我们将本节分为两部分，以展示定性和定量结果。在定性结果小节中，我们在视觉上比较所有五个数据集的生成结果及其PSNR / SSIM指标值*w*。*[R* 。*Ť*。地面真相图像。此外，我们通过展示在训练相同数量的人脸后，由每个通道以及双通道分别显示的每个通道生成的结果（人脸图像），从而展示了使用双通道生成器而不是GAN的单通道生成器的优势。时代。后来，在定量结果中，我们通过将Rank-1识别率与最新的监督DA（SDA），SR和DFR技术进行比较，显示了将本方法用作常规深度分类器的预处理步骤的优势。

*1）定性结果：*使用建议的方法为上述五个数据集生成的结果如图7所示。红色框表示通过建议的技术probe中的身份与相应生成的人脸之间的差异。

为了将我们的方法与其他最新生成模型的性能进行比较，我们将生成的结果与DCGAN [17]，BEGAN [18]，Isola*等人*的性能进行了比较*。*[20]，StarGAN [10]，LR-GAN [9]，SRLSP [32]和SICNN [33]方法。图8显示了我们的方法在五个数据集上的上述最新方法的性能。此外，当使用建议的损失函数的不同组成部分时，我们还显示了模型的生成性能。可以看出，在分别在生成的图像下方报告的PSNR / SSIM值方面，我们的方法大大优于生成任务中的所有其他方法。SR方法SRLSP [32]表现最差，这表明此类方法无法处理训练数据与测试数据之间的域差异。使用以下各项的组合时的消融研究的生成结果：（i）MR\_PMSE + 对抗性（Adv。）+分类（Cls。）损失，（ii）NCD + Adv。+ CL。损失和（iii）MR\_PMSE + NCD + Adv。+ Cls。损耗分量分别显示在图8（i），图8（j）和图8（k）列中。可以看出，在（i）的情况下，皮肤的色泽不能正确地再现在人脸。由于NCD不是判别损失函数，因此对于情况（ii）生成的人脸不是很能判别，但是可以正确捕捉人脸的肤色。然而，由于存在分类损失，已经捕获了人脸的结构。与（i）和（ii）相比，情况（iii）中损失成分的组合提供了更好的综合结果。但是，如图8（l）所示，当将情况（iii）中的KLD损失分量添加到损失中时，我们获得了最佳的生成结果。因此，我们可以得出结论，MR\_PMSE损失函数对重建具有最大的贡献，

由于源数据分布和目标数据分布之间的显着相似性，所提出的方法在Chokepoint数据集上提供了最佳性能[47]。的表现

|  |
| --- |
| 图9.使用FR\_SURV（FSV）[46]，SCface（SCF）[48]，扼流点（CP）[47]上的D2SC-GAN改进基于CNN的FR方法（无DA）的1级识别率，TIP（TIP-D）[4]和建议的ICF数据集，分别适用于下面提到的每组条形图的方法。对于每个条形图，蓝色表示使用图库和15％的probe样本训练时深层模型的性能，而红色表示通过使用从D2SC- GAN（而不是原始probe样本）以及图库 |

训练模型。该图最好用彩色查看。

我们在TIP数据集上的方法[4]是第二好的。由于数据集中存在一系列挑战（姿势变化，照明度差，模糊和遮挡），因此我们提出的ICF数据集表现最差，而在其他任何可用数据集中都找不到这些问题的组合。请注意，使用监控摄像头捕获的所有其他人脸数据集都在额叶附近。所有其他方法在ICFD上的表现特别差，突出了数据集的复杂度。

为了通过我们的方法量化人脸生成的性能，我们还显示了生成任务的平均PNSR / SSIM值，并将它们与表II中的其他最新方法进行了比较。所提出的方法表现最好，而SRLSP [32]表现最差。与LR-GAN模型相比，我们方法的性能跃升表明使用双通道模型以及针对生成任务建议的多分辨率重建损失的有效性。在表II中还报告了采用损失函数的不同分量时获得的平均PSNR / SSIM值，该表遵循我们对所提出的损失函数的不同分量的重要性的讨论得出的结论。

图10显示了我们提出的模型在使用和不使用基于KLD的损失函数的情况下的收敛速度。红线显示没有基于KLD的损失的收敛性，而蓝线显示使用基于KLD的损失函数的模型的收敛性。我们观察到，使用建议的基于KLD的损失函数时，该模型可实现更高的PNSR值，并且收敛速度更快。在表II中，我们还报告了在训练模型期间每个受试者仅使用5％和10％的probe样品时的平均PSNR / SSIM。当仅使用5％的probe样本时，我们注意到模型在所有数据集上的性能都有适度下降。但是，当使用每位受试者10％的probe样本进行训练时，我们模型的性能大大提高。

图11显示了在FR\_SURV [46]和建议的ICF数据集上使用双通道进行重构的优势。（a）列中的图像是模型的输入，

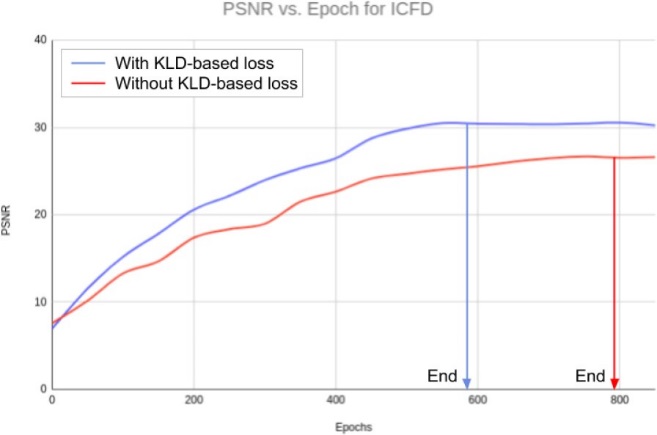


图10.我们提出的模型在ICF数据集的每个周期的平均PSNR值方面的收敛性。红色曲线表示没有基于KLD的损失函数训练的模型，而蓝色曲线表示具有基于KLD的损失训练的模型。End表示训练期间模型的收敛点。

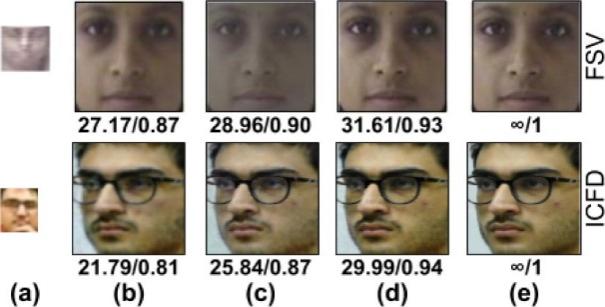


图11. FR\_SURV数据集[46]（上排）和ICFD（下排）从（a）在600个历元后使用（b）仅*G S*，（c）仅*G D*和（d）都*G*生成的人脸*S*和 *G D*（e）为真相。该图最好用彩色查看。

然后是经过600个纪元后的（b）浅（*G S*），（c）深（*G D*）和（d）双通道（*G S* + *G D*）模型生成的结果。可以看出，浅通道（*G S*）的输出保留了低频分量，而缺乏清晰度，而深通道（*G D*）的输出则保留了锐度（高频分量）。

## 表二

Ť HE甲VERAGE PSNR（P ART A）与 SSIM（P ART乙）V ALUES的小号YNTHESIZED ˚F ACES FOR THE 5 d ATASETS FR\_SURV [46]（FSV），

## TIP-d [4]，S Ç ˚F ACE [48]（SCF），C HOKEPOINT [47]（CKP）及

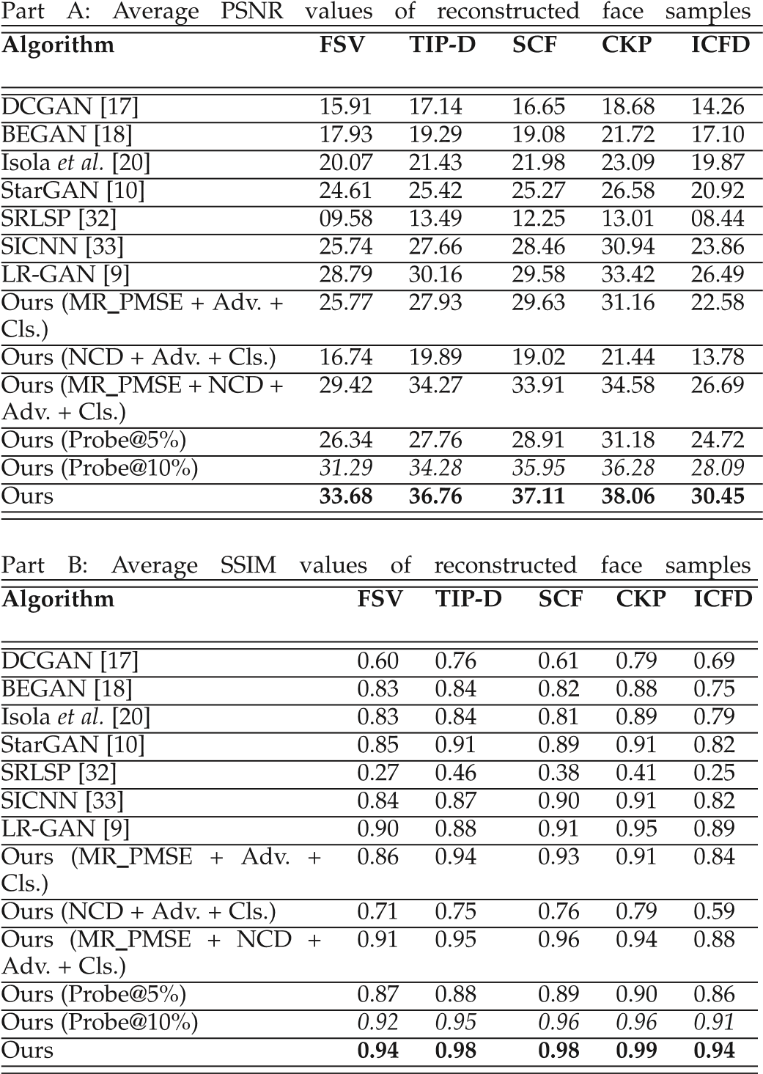
P ROPOSED ICFD，作者 ö UR P ROPOSED中号ETHOD Ç OMPARED TO

## DCGAN [17]，BEGAN [18]，I SOLA*等。*[20]，S TAR GAN [10]，

LR-GAN [9]，SRLSP [32]和SICNN [33]。甲VERAGE PSNR / SSIM

V ALUES ø BTAINED Wˉˉ HEN Ë MPLOYING d IFFERENT Ç THE作者OMPONENTS

### P ROPOSED大号OSS ˚F结和d IFFERENT P作者ERCENTAGES P ROBE小号AMPLES （P ER小号UBJECT ）U SED FOR Ť汛期甲RE甲LSO ř EPORTED 。ř ESULTS IN乙老我TALICS我NDICATES乙EST与小号的Econd -B EST P ERFORMANCES ř ESPECTIVELY



图像，但无法捕获相应gallery图像中存在的准确照明和颜色变化。双通道生成器的输出（图11（d））最接近于地面真实图像（图11（e）），显示了使用所提出的方法进行重建的有效性。

*2）定量结果：*在本节中，我们以两种方式展示我们方法的能力。首先，当使用probe/增强probe以及用于训练模型的库时，我们报告了几种最先进的非DA深度CNN方法的Rank-1识别率。其次，我们将Rank-1识别/识别率与几种最先进的监督域自适应，超分辨率（SR）以及降级FR（DFR）方法进行比较。在图9的条形图中报告了非DA方法的等级1识别率。对于每个条，蓝色表示深层模型在以下情况下的性能：

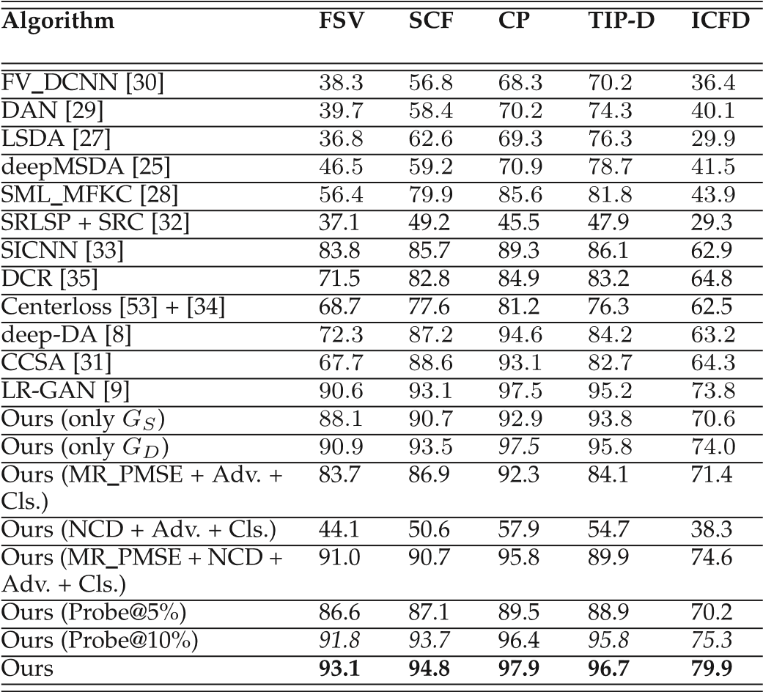
## 表III

### Ç OMPARATIVE小号作者TUDY P ERFORMANCE Wˉˉ ITH THE ř最近几个

小号UPERVISED DA-B ASED ，SR和DFR中号编制方法（W ITH AND Wˉˉ ITHOUT

CNN）IN Ť ERMS作者- [R ANK -1R ECOGNITION / I DENTIFICATION ř ATES （IN

### ％）。ř ESULTS IN乙OLD ，我TALICS和ü NDERLINE我NDICATE THE乙EST ，S的Econd -B EST与Ť HIRD -B EST P ERFORMANCES ，R ESPECTIVELY



使用库和15％的probe样本进行训练，而红色表示通过使用D2SC-GAN生成的15％的增强probe样本（而不是probe样本）与库一起训练来实现的增强模型。Inception-Resnet-V2 [52]表现最好，而AlexNet [26]表现最差，而VGG-19 [11]表现第二好。通过使用增强型probe代替原始probe样本获得的准确度显着提高（在图9中的每个条形中用红色表示），显示了使用建议方法的生成器作为识别的预处理步骤的优势任务。

表III中列出了深度和常规最先进的监督DA，SR以及DFR方法的1级识别率。对于SICNN [33]，DCR [35]和Centerloss [53] + [34]方法，我们已经报告了Rank-1识别率。我们的方法（在底行使用*G S* + *G D*时）对所有数据集表现最佳，而SRLSP [32]表现最差。当仅使用10％的probe样品时，建议的方法表现第二好。LR-GAN [9]显示了第二好的性能（不包括所提出方法的变体），而CCSA [31]紧随其后。使用*G S* + *G D*时性能的增强，而不是单独的*G S*或 *G D*，显示了双通道架构对重建任务的有效性。我们的方法在很大程度上无法胜过所有SR和DFR方法，因为这些方法并非旨在克服较大的域差异。当我们采用不同的损失成分时，我们模型的性能再次遵循了我们在第VI-B1节中对所提出的损失函数的不同成分的重要性的讨论得出的结论。有一个温和的

|  |
| --- |
| 图12.使用tSNE [54]从DA之前和之后的ICFD人脸样本中提取特征的2D可视化。（a）显示了来自gallery和probe组的样本的散布，而（b）显示了相同gallery样品的散布，但具有使用我们提出的方法的DA在（a）中的同一组probe上获得的增强probe组。来自10个对象的20个样本用于说明。不同的颜色代表不同的类别/主题。最好看的图 |

彩色。

当每个受试者使用5％的probe样品而不是15％时，我们的方法的性能下降。但是，当每类使用10％的probe样品而不是15％时，性能会大大提高。总的来说，所有方法对Chokepoint数据集的执行效果最佳，而对ICF数据集的执行效果最差，这表现出了在过去从未讨论过的不受约束的教室场景中执行人脸识别的困难。补充文档的附录C中提供了有关建议的ICF数据集的probe的各种子集的进一步实验结果。在附录D中，我们还定性和定量地报告了在两个其他具有挑战性的数据集（LFW [55]和IJB-A [56]）上，与CCSA [31]和LR-GAN [9]相比，我们模型的性能。补充文件。

为了直观地说明我们提出的方法进行域自适应的有效性背后的逻辑，我们分别对20个原始probe样本和增强probe样本的散点图以及来自ICFD的10个主题（随机选择）的2D对应图库样本进行了可视化。使用从预训练的VGGFace网络中提取的特征进行t-SNE图[54]，以进行无偏见的可视化。图12（a）中的散点图显示，尽管gallery样本在特征空间中被正确地聚类并很好地分离，但另一方面，不同类别的probe样本却存在显着的类别间重叠（彼此重叠） ），几乎没有多少类可分离性，就像相应的图库类一样。实际上，所有类别的probe样本都倾向于相对聚集在一起（一串在图12（a）中观察到， 标记），这与图库样本不同。probe样品也不接近其相应的分类图库，这解释了当使用图库和原始probe样品训练时非DA深层方法的性能差。图12（b）中的图显示了我们的方法通过执行目标到源域映射来弥合源域和目标域之间的差距的有效性。现在，增强型probe样本形成了紧凑的类聚类，靠近相应的gallery对应对象，具有较大的类间可分离性，从而为深层非DA方法的增强性能提供了基础。当使用一些增强probe样本（由建议的D2SC-GAN生成）代替原始probe样本来训练这些分类器时（见图9），

## 七。ç ONCLUSION

提出的D2SC-GAN模型使用新颖的双通道生成器体系结构，通过使用多分辨率重构组件和基于KullbackLeibler散度的组件将新的生成器损耗降至最低，从而从低分辨率probe样本中合成清晰，高分辨率（HR）的gallery样人脸图像，它在生物识别领域具有广泛的应用，并带有鉴别器以对生成的人脸进行分类。该模型还可以用作预处理技术，以将LRprobe样本转换为对应的FR和FV任务的HRgallery样人脸图像。该模型还执行封闭式人脸识别的任务。该方法通过弥合脆性通道表面和具有真实世界退化的probe之间的间隙，有效地执行了DA的任务。我们还提出了印度教室人脸数据集（ICFD），据我们所知，该数据集尚属首次，它将帮助研究人员探索在教室环境中进行FR所面临的挑战。在五个真实世界的人脸数据集上对提出的模型进行的严格实验证明了我们提出的方法的优越性。我们计划通过将无监督域自适应而不是有监督域自适应纳入我们的框架来克服开放式人脸识别的局限性，作为我们未来的工作范围。

原文

提供更好的翻译建议