野外低分辨率人脸识别的比较与新技术

裴丽1 洛雷托·普里托2 多明哥·梅里2 帕特里克·弗林1

1圣母大学计算机科学与工程系

2计算机科学系

智利天主教大学

*摘要—*尽管近年来人脸识别系统取得了令人瞩目的性能，但低分辨率人脸识别任务仍然具有挑战性，尤其是在非理想条件下捕获低分辨率人脸时，这在基于监视的应用程序中很常见。在这种情况下捕获的人脸通常会被模糊，不均匀的光照和非正面的人脸姿势所污染。在本文中，我们使用野外在低质量条件下捕获的数据来分析人脸识别技术。我们对实际监视应用中两个最重要的应用提供了实验结果的全面分析，并展示了处理这两种表现出令人鼓舞性能的案例的实用方法。做出了以下三个贡献：*（i）*我们进行实验以评估用于低分辨率人脸识别的超分辨率方法；*（ii）*我们研究了各种公开人脸数据集上的人脸重新识别，包括真实的监视和大规模数据集的低分辨率子集，提出了几种基于深度学习的方法的基线结果，并通过引入生成对抗网络（GAN）进行了改进）预训练方法和全卷积架构；和*（ⅲ）*我们通过采用一个国家的最先进的监督判别学习方法探索低分辨率人脸识别。对SCface和UCCSface数据集的具有挑战性的部分进行评估。

1. 导论

近年来，我们见证了人脸识别系统性能的巨大提高，尤其是在采用不同的专门设计的深度学习架构时。一些基于深度学习模型的最新方法，包括[66]，[60]，[51]，[67]，[77]和[42]，在公众人脸数据集上的准确率超过99％如LFW [21]。尽管这些算法可以有效地处理具有明显姿势变化的人脸，但是这些人脸通常需要较大的面积。另外，需要诸如人脸正面化和人脸对准之类的预处理技术。这些为HR人脸图像数据设计的过程无法直接应用于低质量的人脸图像。越来越多的公共场所使用监视系统，在检测到的人脸处于低分辨率（LR）的环境中，创建具有挑战性的人脸识别用例。我们称此为*低分辨率人脸识别*（LRFR）问题。用于在监视场景中捕获的图像的实用人脸识别系统可以处理人脸识别任务（使用监视列表）和人脸重新识别任务（其中对象与监视系统中的先前外观匹配）。

arXiv:1805.11529v2 [cs.CV] 28 Mar 2019

这项工作部分基于[37]，[36]中发布的先前工作，但提出了许多新的贡献。在原始工作[37]中，我们专注于轻型深度神经网络的原型来进行LR人脸重新识别，并表明LR人脸图像可用于人的重新识别。我们采用了来自真实监控系统的数据，并进行了广泛的人员重新识别研究。我们将在更大范围内解决LR人脸识别任务，其中包括对使用超分辨率技术，人脸识别和人脸重新识别的受控人脸图像和视频质量人脸图像之间的识别性能差距进行定性探索。我们还基于[37]专注于提高性能，并使用更多的数据集（包括SCface，UCCSface和MegaFace Challenge 2 LR子集）扩展了我们的评估。

在第二部分中，我们提供了一些新近发布的真实监视数据集的文献综述和摘要描述。在第三节中，我们提出了以下新方法和实验：第三节B提出了几种通过评估最新的超分辨率方法获得的基准人脸识别结果[40]，[82]。还介绍了使用来自两个流行人脸数据集（AR [46]和YouTube Faces（YTF）[78]）的LR人脸输入的超分辨率增强人脸识别技术的性能摘要，并说明了人脸识别之间的性能差距在受控环境中捕获并识别野外捕获的人脸。在第III-C节中，我们提出了用于LR人脸识别的中心正则化方法。这项工作的目的是学习一个公共的特征空间，该特征空间将同一对象的LR和HR人脸图像在特征空间中定位得尽可能靠近，而无需生成数千对训练。通过对两个不同的数据集进行综合评估，我们能够显示在使用LR输入时在开放式和封闭式人脸识别之间的性能差距。在第III-D节中，我们首先基于名为VBOLO的真实监视数据集总结了先前工作中的发现[37]。设计和评估了用于LR监视人脸图像的不同深度架构，并将其作为基线结果，并通过引入完全卷积空间金字塔池（SPP）对其进行了改进。为了概括性能在更大数据规模上的能力，我们将从在线收集的视频或图像中的几个监视和LR数据集中进一步研究人脸重新识别。另外，我们提出了一种新颖的基于DCGAN的预训练方法，以进一步提高性能。在第四节中，提供了结束语。

1. 相应工作

在本节中，我们介绍有关LR图像的人脸识别和人脸重新识别的相关工作，并描述一些可以在此研究领域中使用的重要数据集。

# A.低分辨率人脸识别

LRFR任务是一般人脸识别问题的子集。用于此任务的最有用的应用场景之一是使用监视图像进行识别。在这种情况下，会从具有较大支架的摄像头捕获人脸，这些摄像头位于头顶上方，有时甚至在充满挑战的照明条件下。即使可以通过受过专门训练的人脸检测器检测到人脸，由于缺乏嵌入图像本身的信息，因此很难构造鲁棒的特征表示。尽管HR人脸识别系统已在多个数据集上取得了近乎完美的性能，但LRFR仍然是一个具有挑战性的问题。文献中解决这个问题的方法主要有两个主题。一些技术采用超分辨率（SR）和去模糊技术来将输入LR人脸尺寸增加到高分辨率（HR）人脸识别技术可以很好工作的程度。另一种流行的方法是为LR和HR人脸图像学习统一的特征空间，在该空间中特征向量距离作为匹配得分发挥其典型作用。

SR技术被广泛用于将LR图像转换为质量更好的图像。因此，人脸SR方法是恢复LR人脸图像以进行识别的直观方法。HenningsYeomans等。[19]包括先前提取的人脸特征作为SR结果的拟合度量，并从重建和识别的角度执行SR。杨等。[82]提出了一种通用字典联合字典训练方法，该方法同时使用了LR和HR图像补丁。邹等。[93]设计了一个线性回归模型，其中包含两个元素（新数据和判别约束）以学习映射。江等。[23]提出了一种通过多层局部约束迭代邻居嵌入的从粗到细的人脸SR方法。Kolouri等。[31]介绍了一种单帧SR技术，该技术使用了基于传输的解决方案。Wang等。

解决LR人脸识别问题的另一种方法是寻找一个统一的特征空间，以保留不同分辨率的人脸之间的接近度。Li等。[35]提出了一种耦合映射方法，将具有不同分辨率的人脸图像投影到统一的特征空间中。Biswas等。[3]将LR和HR人脸同时嵌入到公共空间中，以使它们在变换后的空间中的距离近似于同一对象的两个HR人脸图像之间的距离。任等人。[58]使用HR和LR对应的耦合映射策略来学习投影方向以及利用判别信息。Shekhar等。[61]提出了鲁棒的字典学习的LR人脸识别共享共同的稀疏代码。邱等人的技术。[54]学习了一个领域自适应词典来处理在源域和目标域中捕获的两个人脸的匹配。Li等。[36]提出了几种浅层网络结构来学习LR和HR图像之间的潜在空间，并在新的监视数据集上评估了所提出的方法。还有其他一些方法，例如[56]和[20]，它们直接探索更鲁棒的功能来提高模糊和退化的人脸图像中的识别率。另外，类似于[17]，[49]和[50]中提出的方法，可以使用去模糊技术来恢复上采样的LR图像。还有其他一些方法，例如[56]和[20]，它们直接探索更鲁棒的功能来提高模糊和退化的人脸图像中的识别率。另外，类似于[17]，[49]和[50]中提出的方法，可以使用去模糊技术来恢复上采样的LR图像。还有其他一些方法，例如[56]和[20]，它们直接探索更鲁棒的功能来提高模糊和退化的人脸图像中的识别率。另外，类似于[17]，[49]和[50]中提出的方法，可以使用去模糊技术来恢复上采样的LR图像。

在最近三年中，已经开发出了基于深度学习的低质量人脸图像的新颖人脸识别方法。最相关的方法如下。在[74]中，提出了部分耦合网络用于无监督的超分辨率预训练。分类器是通过在不同的数据集上进行微调以获得特定域同时进行超分辨率和识别而获得的。在[25]，[26]中，提出了一种注意力模型，该模型通过以各种百分比的模糊度对图像进行模糊处理来转移训练过程中网络的注意力，从而实现性别识别。在[47]中，提出了三种模糊处理技术来恢复由于*镶嵌*而退化的人脸图像。（像素化）和模糊过程。在[4]中，提出了一个多任务深度模型来同时学习人脸超分辨率和人脸界标定位。使用生成对抗网络（GAN）[12]，[33]来训练人脸超分辨率子网（请参阅[8]中关于人脸超分辨率的不同版本GAN的比较）。在[22]中，受传统的小波启发，该小波可以在不同级别上描述图像的上下文和纹理信息，提出了一种深层架构。在[7]中，提出了一种网络，该网络包含用于恢复粗糙HR图像的粗糙超分辨率网络。这是在端到端训练和测试之前第一个利用人脸几何形状的深脸超分辨率网络。在[10]中，提出了一种使用基于Resnet的非最大缓冲架构对人脸图像进行模糊处理的网络。在[85]中，提出了一种基于上采样网络和判别网络的人脸幻觉方法。该方法包括带有其他人脸属性信息的特征图。在[62]中，利用人脸的全局语义先验来恢复模糊的人脸图像。在[44]中，设计了一个新的分支网络，可以将其添加到干线网络，以将不同分辨率的探测图像匹配到图库图像。

在所有这些方法中，我们看到，当处理在不受约束的环境（*例如*监视，取证等）中拍摄的更具挑战性的图像时，自动人脸识别远非完美。上述大多数方法都是在低质量版本的受约束人脸上进行评估的数据集，例如Multi-PIE，FERET或FRGC。通过直接对原始图像进行下采样或模糊处理来创建图像。然而，当在不受约束的环境中捕获人脸时，LR人脸识别问题成为一个挑战。在这种情况下，需要更广泛地研究LR人脸识别问题。

# B.人脸重新识别

在此，我们简要介绍人物重新识别（ReID）和我们进行人脸重新识别的动机。典型的端到端ReID系统通常包括以下步骤：人员检测，预处理，特征提取和匹配。它被广泛地用于监视目的，按照给定的一对图像进行建模。目的是查找图像是否来自同一个人。通常，图片是从监视网络中的不同摄像机捕获的。解决ReID问题的传统方法着重于两个主要部分：特征提取和用于匹配的相似度计算。我们可以将大多数现有方法分为两类：采用深度学习的方法和不采用深度学习的方法。对于非深度学习方法，研究提出并使用了不同的手工特征，例如对称驱动的局部特征（SDALF）[2]，颜色直方图[80]，[89]，颜色名称[83]，局部二进制模式[28]，补丁集合摘要功能[90]，度量学习方法[80]，[28]，[30]，[38]，[45]，[39]，[80]，[24]，[6]，[88]和各种这些的组合。几种深度学习方法使用“暹罗”深度卷积神经网络[84]，[70]，[71]，[41]进行特征提取和度量学习，同时提供新颖的端到端解决方案。艾哈迈德（Ahmed）等人。[1]根据Yi等人的想法提出了一种新的深度学习框架。[84]，其中两个新颖的层被用来通过整合基于中层特征的局部关系来计算跨输入邻域差异。他们还表明，从头和颈部获得的特征可能是重新识别人的重要线索。Wu等。[79]通过使用更深的架构和新的优化方法，基于艾哈迈德的思想提高了性能。还设计了其他深层网络结构，例如[69]和[65]，这些结构也可以有效解决旧版ReID数据集上的ReID问题。Qui等。[54]尝试通过使用域自适应方法来调和不同的人脸姿势来执行人脸ReID；然而，他们的实验是在Multi-PIE [15]数据集上进行的，其中人脸图像具有可控制的姿势和照明。尽管在公共区域部署了越来越多的监视摄像机，视频帧的质量通常较低，并且捕获到帧中的人员处于不受控制的姿势和照明条件下。因此，普通人的重新识别可能是一项艰巨的任务。如[5]所示，身体和步态可能在识别LR视频帧中的目标上起作用，但是，遮盖目标会导致人类水平的识别性能急剧下降。同样，文献[36]中的人脸遮盖实验也证明了人脸可能是身份识别中必不可少的一部分。这些工作有助于将LRFR问题作为重新识别问题的一部分。掩盖目标会使人类的识别性能急剧下降。同样，文献[36]中的人脸遮盖实验也证明了人脸可能是身份识别中必不可少的一部分。这些工作有助于将LRFR问题作为重新识别问题的一部分。掩盖目标会使人类的识别性能急剧下降。同样，文献[36]中的人脸遮盖实验也证明了人脸可能是身份识别中必不可少的一部分。这些工作有助于将LRFR问题作为重新识别问题的一部分。

# C.数据集

有一些良好的监视数据集，也有一些大型的无约束人脸数据集，其中包含适合训练和测试LRFR系统的自然LR人脸。

用于研究的大多数LR人脸图像都是通过对在受控环境中收集的标准人脸识别数据集进行下采样而生成的。我们选择AR数据集来研究在不受控制的场景下的LRFR任务以及其他不受约束的LR人脸数据集，以进行更多探索。它用于说明从高质量的人脸图像人工生成的LR人脸图像与在不受限制的场景（例如监视摄像机网络）中直接收集的LR人脸图像之间的不同数据分布。

1. *AR人脸：* AR数据库[46]中的图像是从100名受试者（50名女性和50名男性）拍摄的，这些受试者的人脸表情，光照条件不同，并且在严格控制的条件下被太阳镜或围巾遮挡。在我们的工作中，这些图像用于估计受控环境中对齐的人脸图像上的人脸识别方法的基线性能。
2. *MegaFace Challenge 2 LR子集：*MegaFace挑战2（MF2）训练数据集[48]是最大的（按身份数量计）公开可用的人脸识别数据集，具有470万张人脸图像和超过672,000个身份。MF2数据集是通过在Flickr [68]的图像上运行Dlib [29]人脸检测器而获得的，在130,154个不同的Flickr帐户中产生了4000万张未标记的人脸。使用群集算法执行自动身份标记。我们从带有紧定边界框的MegaFace Challenge 2训练集中进行了子集选择，以生成此数据集的LR子集。为每个身份收集小于50x50像素的人脸，然后我们用少于五个的可用图像来消除身份。这种子集选择方法总共产生了6,700个身份和85,344张人脸图像。与原始数据集处理一样，提取过程确实会产生一些非人脸图像。不会对此子集执行进一步的数据清理。
3. *YouTubeFaces：* YouTubeFaces数据库[78]用于研究视频中人脸识别不受约束的问题。数据集包含1,595个不同人的3,425个视频。每个主题平均提供2.15个视频。
4. *SCface：* SCface [14]是使用五个不同质量的视频监控摄像机在不受控制的室内环境中收集的人脸静态图像的数据库。该数据库包含130位受试者的4,160张静态图像（在可见光和红外光谱中）。我们选择HR和LR可见人脸子集进行训练和测试。
5. *UCCSface：*不受限制的大学生（UCCS）数据集[59]包含从科罗拉多大学科罗拉多斯普林斯分校的18百万像素摄像机捕获的HR图像，捕获了从人行道100到150米对峙的校园人行道上的人，每秒一帧。数据集包含超过70,000个手工裁剪的人脸区域。身份被手动分配给许多人脸。我们使用已分配身份的数据库子集（总共180个身份）。尽管数据是由高清相机捕获的，但由于间距较大，人脸区域很小，并且包含大量噪点和模糊度。
6. *VBOLO人脸：*此数据集[36]，[37]（以前称为EBOLO）是在各个阶段的多个会话中收集的

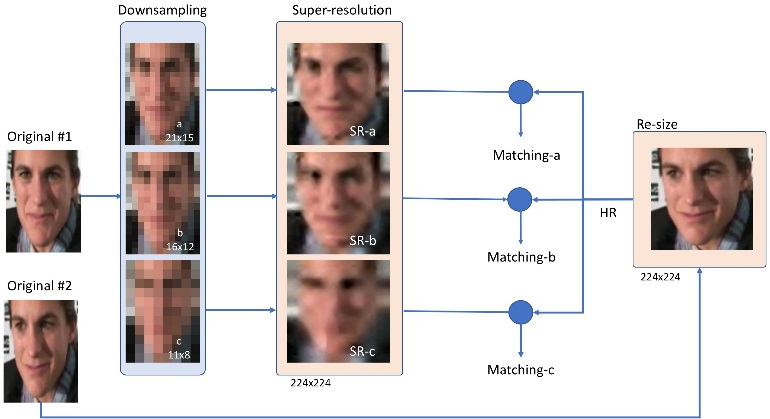


图1：超分辨率实验：针对五个不同的LR尺寸，使用五种不同的超分辨率算法将＃01（用于LR的目的）和＃02（用于HR的目的）的两个图像进行匹配。

隧道，桥梁和走廊等公共交通设施内的检查站。这些捕获环境包括不同的摄像机安装高度和俯角，照明，背景，分辨率，行人姿势和干扰物。该数据集为人脸ReID问题提供了一个很好的方案。该数据集使用一小组已知人员（“演员”）和未知人员（“干扰者”）进出监视摄像机的视野。演员在相机视野中的每个外观之间随机更换衣服。与典型的基于身体的ReID数据集（每个受试者只有很少的图像）相比，VBOLO数据集具有连续视频帧中每个受试者的大量注释，从而模仿了监视跟踪和检测的真实情况。*i）*人脸大小发生明显变化，并且姿势发生明显变化；和*ⅱ）*的两个摄像机供给探针和库的图像可以具有不同的分辨率和观点。在VBOLO中，我们采用了在这项研究中在两个不同的位置（分别表示为Station 1和Station 2）捕获的视频。每个集合都有九个演员，每个都有九个出场。

三， 方法理论和实验

在本节中，我们介绍了有关LRFR的四组方法和实验：*A）*超分辨率技术，*B）*虚拟和真实LR的比较，*C）*人脸识别和*D）人*脸ReId。

# A.超分辨率技术

*1）说明：*为了探索受限和不受约束的LR人脸识别性能之间的差距，我们使用AR [46]和YouTube Faces（YTF）[78]数据集设计了一个小型超分辨率（SR）实验。在此实验中，其想法是评估两个人脸图像（LR图像和HR图像）的匹配性能。

在数据选择过程中，从每个主题的数据集中选择两个图像或视频帧。如图1所示，“original＃1”和“original＃2”分别用于LR和HR。第一个降采样为三种LR尺寸中的每一种：（a）21 × 15，（b）16 × 12和（c）11 × 8像素。第二个被上采样到匹配器所需的224 × 224像素输入大小。在两种情况下，都使用双三次插值法[13]。图像或帧的选择如下。

• **AR数据集**：原始的LR和HR人脸图像分别取自每个被摄对象的人脸图像从＃01到＃14。由于AR数据集中有100个对象，因此我们为此实验获得了100对LR–HR人脸图像。

• **YouTube人脸数据集**：原始的LR和HR人脸图像选则为最早和最新的帧，其中包含每个主题可用的最新视频片段中的正面人脸图像。通过对每帧应用朱的方法[92]并选择朱的方法返回的姿态为0的人脸图像，来进行正面人脸的检测。但是，只有1,463个对象在上一个视频剪辑中有两个合适的帧。因此，我们有1,463对LR-HR对人脸图像可用。

在选择图像和帧并缩放到目标HR和LR大小之后，然后可以通过以下每种方法将来自该过程的每个LR图像放大到HR图像的相同大小（224 × 224像素）：

1. 双三次插值[13]，（

（b）SCN [40]，

（c）稀疏表示超分辨率（ScSR）[82]，

（d）LapSRN [32]和

（e）SRGAN [34]。

另外，我们在实验中加入了“直接方法”，其中原始图像＃1和＃2在不使用双三次插值进行下采样的情况下进行了比较，去实现了匹配器所需的224 × 224像素大小。该实验设计如图1所示。

***2）实验和结果****：*实验方法如图1所示。[51]中的VGG-face训练网络用于为HR图像和所有15个SR图像生成特征向量。对于涉及一个HR和15个不同的放大SR图像之一的每个匹配和不匹配对，获得匹配分数。余弦距离用作匹配分数。

AR和YouTube Faces数据集的累积匹配特征曲线可以在图2和3看到。可以预期，性能随着分辨率的降低而降低：Direct方法（无下采样）比LR图像获得更好的结果。此外，21 × 15 LR图像的性能优于16 × 12 LR图像，而这些图像也优于11 × 8 LR图像。

为了显示使用低质量图像的性能下降，我们进行了另一个实验。由于YouTube视频帧中的原始人脸质量可能较低，因此我们不仅评估了所有1,463对匹配的性能，而且评估了质量最高的500对和质量最低的500对的性能。我们分别将这两个新子集称为HQ-YT和LQ-YT。为了测量原始人脸图像的质量，我们使用基于图像小波变换的高频系数和低频系数之比的分数[52]。低分值表示低质量。当更改HQ-YT和LQ-YT中的下采样目标分辨率时，我们会评估rank 10的性能。

0

20

40

60

80

100

Rank

0

20

40

60

80

100

Cumulative score

Bicubic - 21x15

Bicubic - 16x12

Bicubic - 11x8

SCN - 21x15

SCN - 16x12

SCN - 11x8

ScSR - 21x15

ScSR - 16x12

ScSR - 11x8

SRGAN - 21x15

SRGAN - 16x12

SRGAN - 11x8

LapSRN - 21x15

LapSRN - 16x12

LapSRN - 11x8

Direct method

图2：AR的累积匹配特性。

0

20

40

60

80

100

Cumulative score

Bicubic - 21x15

Bicubic - 16x12

Bicubic - 11x8

SCN - 21x15

SCN - 16x12

SCN - 11x8

ScSR - 21x15

ScSR - 16x12

ScSR - 11x8

SRGAN - 21x15

SRGAN - 16x12

SRGAN - 11x8

LapSRN - 21x15

LapSRN - 16x12

LapSRN - 11x8

Direct method

02004006008001000100012001400

秩

图3：YouTube Faces的累积匹配特征。

在表I中可以看到，高质量的图像比低质量的图像具有更好的性能。

图2和图3的结果证明，对于这些数据集和这些算法实现，在相同的分辨率下，稀疏表示超分辨率（ScSR）和双三次插值始终优于深度学习超分辨率方法。深度学习方法的性能较差是由于在严重降级的图像中引入了伪像，如图1的示例所示。

# B.虚拟（合成）和真实LR图像上的FR之间的比较

*1）说明：*为了探究虚拟（合成）LR和真实LR之间的差距，我们设计了一个YouTube Faces和SCface数据集的实验，通过匹配HR图像和放大的LR图像（使用双三次插值或SRGAN）来评估性能差距。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 子集 |  | （一个） | （b） | （C） |
|  |  | 21 × 15 | 16 × 12 | 11 × 8 |
| 总部YT： | 双三次 | 80.8％ | 53.2％ | 15.6％ |
|  | SCN | 69.6％ | 35.2％ | 9.6％ |
|  | 社会责任 | 87.2％ | 66.4％ | 26.2％ |
|  | LapSRN | 78.4％ | 56.4％ | 20.4％ |
|  | SRGAN | 75.4％ | 53.4％ | 16.2％ |
| LQ-YT： | 双三次 | 81.4％ | 45.8％ | 10.8％ |
|  | SCN | 64.4％ | 27.2％ | 7.2％ |
|  | 社会责任 | 87.2％ | 61.0％ | 14.6％ |
|  | LapSRN | 64.7％ | 38.9％ | 9.8％ |
|  | SRGAN | 70.9％ | 40.5％ | 12.4％ |

LR数据集（如第III-A节所述，根据YouTube Faces计算得出）和真实的LR数据集（SCface）。使用的升级方法是传统方法（bicubic）和深度学习方法（SRGAN）。

• **YouTube Faces数据集**：使用与超分辨率实验相同的图像选择。然后，使用双三次插值将LR图像降采样为16 × 12。

• **SCface**：对于数据集中的130个对象中的每一个，都选择相机3以4.20米的对距拍摄的HR人脸照片和LR作为画廊图像和探测图像。图像也将调整为16 × 12。然后，使用SRGAN和双三次插值输出224 × 224像素的人脸图像。

***2）实验和结果：***实验方法类似于超分辨率实验。我们使用训练过的VGG-face网络[51]为数据集的HR和LR图像生成特征向量。余弦距离用作分数。可以从4中看到来自Youtube Faces和SCface以及这两种方法的100个随机受试者的累积匹配特征曲线。

决定使用100个随机对象以使比较更公平。如果使用所有对象才能在rank 1中获得100％的表现，则YouTube Faces需要正确获取1463个主题，而SCface仅需要130个主题。

图4显示，虚拟（人造的）LR上超分辨率方法的性能始终比真实LR好得多。因此，在其他方面来表征真实LR图像上的系统性能时，应使用真实LR人脸图像，而不是通过简单地对HR人脸图像进行下采样获得的合成LR人脸图像。

# C.低分辨率人脸识别

在本节中，我们重点介绍LR人脸识别。我们首先关注跨分辨率的人脸识别，该识别方法适用于以下情况：当所注册的人脸图像主要是在具有不受控制的姿势和光照条件的监控摄像机捕获的HR和LR人脸的受控场景中收集的。这是一项具有挑战性的识别任务，在很大程度上取决于良好的分辨率不变表示。

0

20

40

60

80

100

Rank

0

20

40

60

80

100

Cumulative score

SCface - Bicubic - 16x12

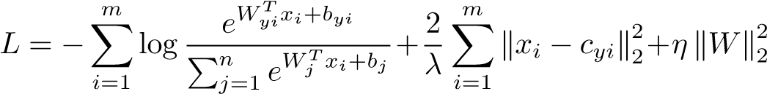
SCface - SRGAN - 16x12

YouTube faces - Bicubic - 16x12

YouTube faces - SRGAN - 16x12

图4：虚拟和真实LR实验的累积匹配特征图

1. *说明：*现有的大多数交叉分辨率匹配方法将学习一个表示LR和HR人脸的统一空间。这需要在训练过程中精心设计的人脸配对挖掘策略，既耗时又对性能敏感。为了充分探究HR和LR域之间的内在联系，我们决定平等地包含HR和LR人脸图像，期望学习一个公共特征空间，该空间可以将同一主题中的LR和HR人脸聚类。尽管分辨率不同，但仍保持较低的类间相似度。我们基于[77]提出了一种方法，该方法采用了新的正则项来进一步将同一主题的特征强制聚类，从而产生更好的判别特征。为了进一步稳定训练过程并减少模型在较小数据集上的过拟合，我们还采用了L2正则化项。损失函数为：



其中 *x i* 代表具有不同分辨率的人脸图像的batches，而*c yi*是每个类别的中心，它们在训练时会更新。

1. *实验和结果：*在以前的大多数研究中，由于缺乏从野外收集到的LR图像，因此将HR数据集（如Multi-PIE [15]和FERET [53]）用于HR人脸的下采样图像来创建LR副本。由于此技术创建的合成LR图像与实际监视图像之间存在差异，因此该方法在应用于现实世界时会受到限制。两种实验协议定义如下：

* *Close-set人脸识别*通常被视为分类问题，并且在采用深度学习方法时，可直接从分类层预测标签。但是，当用于评估的主题没有出现在训练阶段时，分类结构往往不够灵活。在现实生活中，这种无主题的训练和测试协议更加实用。
* 当一个人出现在系统中时，open-set*人脸识别*需要1到N匹配。但是，这还要求系统正确拒绝未注册系统数据库的个人。这与大多数监视系统的工作方式非常相似。随机出现在场景中的个人可能会或可能不在人脸数据库中。在这种情况下，如果该人不在数据库中，则系统必须正确拒绝该探测，并正确识别该人是否在数据库中。（？）

对两个监控质量数据集（SCface和UCCSface）进行了评估，我们认为这两个数据集更具挑战性，而且更接近于实际情况。对于网络体系结构，我们采用具有七个权重层的较浅网络，称为conv（32）-conv（32）-maxpooling-conv（64）-conv（64）-maxpooling-conv（128）-conv（128）- maxpooling-flatten-dense。卷积核大小保持为3。对于中心损失，我们使用512维的密集层，对于large margin softmax损失[43]，addictive margin softmax [72]和L2-constrained-softmax损失[57]，我们使用维度为256的层作为密集层。对于中心损失，我们将alpha设置为0.9。对于L2约束的softmax损失，alpha设置为5。对addictive margin softmax损失，我们将margin设置为1，对于addictive margin softmax损失，我们将margin设置为0.35。

•SCface数据集：我们使用的SCface数据集的130个对象中的每一个都有一个高清相机拍摄的人脸HR人脸照片和15个LR人脸图像，这些图像由五个在三个不同的对峙距离处放置的可见光相机捕获（ 1m，2.6m和4.2m）。总共产生2080张脸。我们按照[81]中定义的两种不同协议进行了两组实验。我们以subject不相交的方式分别将训练集和测试集分为80和50个主题。对于实验1，我们将HR图像用作图库图像，并将在三个间隔中捕获的HR图像用作gallery图像。在实验2中，我们选择距离1m的人脸图像作为gallery图像，并选择2.6m和4.2m相距的人脸图像作为probe图像。其他设置与实验1相同。所有的HR和LR图像都调整为64x64大小，以呈现给网络进行培训和测试。我们使用匹配分数的余弦距离进行匹配，II中报告了rank-1率。与相同协议下的最新技术相比，我们在实验1中的等级1速率提高了近5％，在实验2中的等级提高了9％。当图库图像和探针图像的分辨率级别差异较大时，由我们的模型更加鲁棒。当采用从我们的模型中提取的特征时，探针图像的大小从1m间距变为4.2m间距时，其他方法的性能会迅速下降。与相同协议下的最新技术相比，我们在实验1中的等级1速率提高了近5％，在实验2中的等级提高了9％。当图库图像和探针图像的分辨率级别差异较大时，由我们的模型生成的功能会更强大。当采用从我们的模型中提取的特征时，探针图像的大小从1m间距变为4.2m间距时，其他方法的性能会迅速下降。

* **UCCSface数据集**：UCCSface是另一个设计为接近真实监视设置的数据集。我们遵循[73]和[59]中提供的实验设置，并研究封闭式和开放式方案。对于封闭式评估，使用了180个受试者，对于开放式评估，我们将我们的结果与[59]中表II中报告的性能进行了比较：表II：实验1：具有HD和三个间隔距离的SCface上的Rank-1率

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | HD-1m | 高清-2.6m | HD-470万 |
| SCface [14] | 6.18％ | 6.18％ | 1.82％ |
| CLPM [35] | 3.08％ | 4.32％ | 3.46％ |
| SSR [82] | 18.09％ | 13.2％ | 7.04％ |
| CSCDN [75] | 18.97％ | 13.58％ | 6.99％ |
| 共同国家评估[76] | 20.69％ | 14.85％ | 9.79％ |
| DCA [16] | 25.53％ | 18.44％ | 12.19％ |
| C-RSDA [11] | 18.46％ | 18.08％ | 15.77％ |
| 中心损失[77]（我们的） | 31.71％ | 20.80％ | 20.40％ |
| LMsoftmax [43]（我们的） | 18.00％ | 16.00％ | 14.00％ |
| AMsoftmax [72]（我们的） | 18.4％ | 20.8％ | 14.80％ |
| L2softmax [57]（我们的） | 16.8％ | 18.8％ | 9.2％ |

表III：实验2：间距为1.0m和2.6m的SCface上的Rank-1速率

|  |  |
| --- | --- |
| 方法 | 1.0m-2.6m |
| CLPM [35] | 29.12％ |
| SDA [91] | 40.08％ |
| CMFA [64] | 39.56％ |
| 耦合映射方法[63] | 43.24％ |
| LMCM [87] | 60.40％ |
| 中心损失[77]（我们的） | 69.60％ |
| LMSoftmax [43]（我们的） | 40.4％ |
| AMSoftmax [72]（我们的） | 46.8％ |
| L2softmax [57]（我们的） | 42.8％ |

定义的开放度为14.11％。当查看封闭式评估的结果时，在相同训练下，我们的方法在1级精度上比UCCS基线高出近20％，在等级1率上也优于[73]中的DNN方法高35％。和评估协议。对于开放式评估，与UCCS面基准结果相比，我们达到了73.6％的准确性。

表IV：UCCSface数据集的1级比率

|  |  |  |
| --- | --- | --- |
| 决议案 | 方法 | 等级1率 |
| 原始与原始： | UCCSface | 78.00％ |
|  | 中心损失[77]（我们的） | 95.10％ |
|  | LMSoftmax [43]（我们的） | 65.8％ |
|  | AMSoftmax [72]（我们的） | 60.6％ |
|  | L2Softmax [57]（我们的） | 86.50％ |
| 80 × 80与16 × 16： | DNN | 59.03％ |
| （HR与LR） | 中心损失[77]（我们的） | 93.40％ |
|  | LMSoftmax [43]（我们的） | 64.9％ |
|  | AMsoftmax [72]（我们的） | 58.6％ |
|  | L2Softmax [57]（我们的） | 85％ |

# D.低分辨率人脸重识别

在本节中，我们探索LR人脸重新识别并在不受限制的环境中捕获的几个数据集上对其进行评估。我们使用VBOLO数据集进行深入研究，并使用SCface，UCCSface和MegaFace Challenge 2 LR子集进行其他主题探索。

*1）实验与结果：*

1. *使用选定的深层网络的Actor-Disjoint实验：*我们探索文献，并受到最新技术的补丁匹配方法的启发，这些方法可能会表现出对自动或手动注释中小的未对准的稳健性以及对不同有效分辨率的稳健性。我们还期望网络能够处理由相同或不同相机以不同的主体间隔拍摄的人脸图像的匹配情况。我们利用基本DNN架构开发了四种最先进的人脸匹配方法，并通过完全卷积的结构对其进行了增强，以减少数据集的过度拟合。此外，为了让网络更好地适应分辨率的变化，我们采用了一个空间金字塔池（SPP）层，希望了解区别特征以及在监视摄像机中捕获的不同大小的LR人脸之间的映射。
2. *匹配协议：*我们为此数据集制定了两种不同的匹配协议。第一个协议旨在将同一台摄像机中的人物在不同的外观之间进行匹配（通常称为“单摄像机ReID”）。此实验旨在测试视频中外貌不同的人，而外表不同的人的人脸识别性能。第二协议匹配从不同相机位置获取的人的图像。该协议旨在评估ReID模型的综合性能，该模型同时包含单摄像机和多摄像机人ReID。为了增加匹配的复杂度，我们在协议中添加了干扰项以获得更多的不匹配对。

我们遵循两种匹配协议，在我们的数据集中训练并测试了几种最新的深度学习补丁匹配架构。演员ID不区分训练和测试集，以便模仿这样的现实：目标不太可能出现在训练集中。其中六个演员用于培训，三个用于测试。使用下面的随机对采样程序对每组实验进行五次，并对结果取平均值。

1. *训练配对策略：在配对训练中*创建配对并对生成的配对进行采样是预处理的关键步骤。由于我们每次出现时平均每个人大约有200张人脸，因此正负配对的数量非常不平衡。除了使用不同外观的人脸创建的对之外，我们还决定在训练数据中添加来自同一外观的人脸对，以增加训练对的数量。我们表示训练集合作为 *Ť*，并且在特定的面 *Ť*如*吨IJF*，其中*我* ∈1 *... N*是演员的ID， *Ĵ* ∈1*，...，M*是出现次数，并*f*表示帧号。我们首先随机打乱所有人脸，以破坏帧的时间连续性，以避免在时间上彼此接近的帧中更频繁地获得正脸对。对于每一个的面对*Ť IJF*固定*我*和*Ĵ*，我们通过随机地选择创建一个积极的一对*Ť IJ* 0 *˚F* 0 与*Ĵ* 0 6 = *Ĵ*和*˚F* 0 6 = *˚F*。为了创建相等数量的负对，我们从*T i* 0 *jf* 0中选择带有*i的面*0 6 = *i*和*j* 0和*f* 0随机选择，并将选定的脸与先前的脸*T ijf配对*。通过利用上述配对方法，我们能够从每种外观的每个身份中获得平衡的人脸对。
2. *VGG-Face＆SVM：*我们采用了预先训练的VGGface描述符模型，并将其用于人脸特征提取。我们首先将数据集中的人脸配对，然后使用VGG人脸描述符提取特征[51]。人脸图像从原始尺寸调整为224×224表示输入到网络。来自一对人脸的特征向量之间的欧几里得距离分配有正负二进制标签，以表示该对人脸是否来自同一个人。距离本身用于训练线性SVM模型以进行二进制预测。我们使用上述两个匹配协议进行了实验。我们选择了超参数值，并获得了线性SVM的最佳CV速率，这产生了0.695的测试AUC，如图6所示。该结果表明网络已成功地从监控视频中提取的劣质人脸识别出人脸特征。但是，由于VGG人脸描述符是在与人脸标志充分对齐的各种HR人脸上训练的，因此当用于监视既是LR又难以对齐的高质量人脸时，这个基线结果并不出色。此外，LR人脸图像在馈入深度预训练网络之前需要进行较大比例的缩放，这可能会引入伪像并还会增加计算复杂性。
3. *连体网络：*连体分类结构在[9]中首次在人脸验证中得到应用。暹罗体系结构在培训中不需要分类信息。它尝试通过共享权重的两个相同的网络层塔来学习特征表示。每个塔中都有一系列卷积层，激活层和最大池层。来自每个塔的两个特征表示被串联并馈入完全连接的层，这些层以对比损失的方式连接。由于两个塔的结构和重量相同，因此这种网络旨在通过端到端的训练将两个输入映射到相同的目标空间中。[73]提出了一个结构简单的小型基础网络，结论是，更深的网络架构和大量的过滤器通道可能会降低识别性能。我们的基本网络如表V所示。它具有三个卷积层，然后是最大池化层和一个完全连接的层。我们在小型网络中使用中等大小的过滤器和通道号。输入大小为32× 32被选择。我们从头开始训练网络（批处理大小为8）和SGD优化器。暹罗网络在5个纪元内收敛，站点1数据的AUC为0.861，站点2数据的AUC为0.871（单摄像机匹配），两个站点的数据为0.838（单摄像机和跨摄像机匹配）。
4. *MatchNet：* MatchNet [18]是另一种最先进的补丁匹配方法，它采用两塔结构，共享权重类似于暹罗网络。然而，与其小心地将两个塔产生的两个串联的特征向量直接馈送到决策层，不如将其输入决策层。

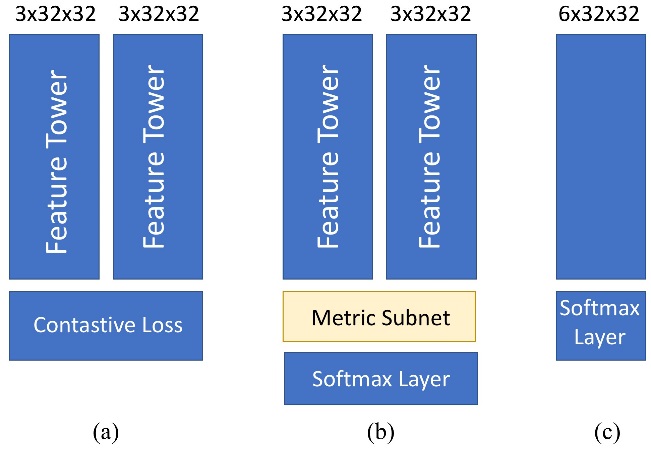
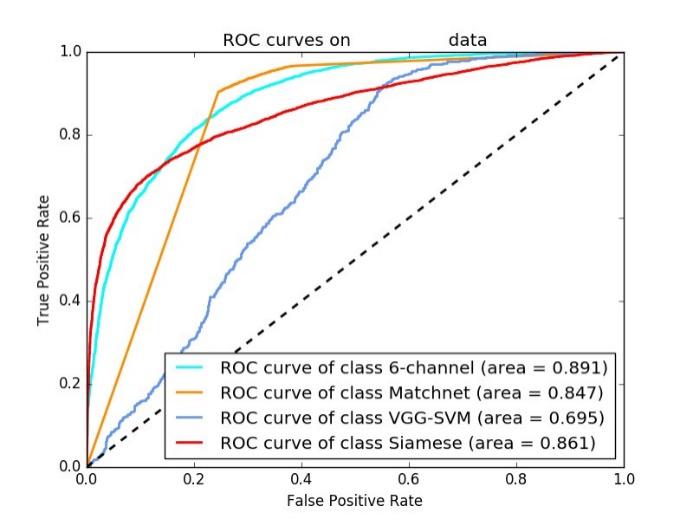


图5：三个深层架构的概述：（a）暹罗网（b）Matchnet（c）6通道网

设计了损失函数，它使用一系列完全连接的层作为子网，以学习使用交叉熵进行二进制分类的特征比较。与Siamese网络相比，MatchNet在图5所示的度量子网中具有更大的灵活性，它采用了配对的特征并将它们映射到统一的空间，从而使距离最小化。但是，由于完全连接的层具有更多的参数和更高的复杂度，因此收敛速度较慢。在训练期间采用softmax层和交叉熵损失。我们使用SGD优化器在基本网络上获得最佳结果。如图6所示，对于单台摄像机匹配，从站2获得的数据的AUC为0.847，从站1获得的数据为0.902，对于单摄像机和跨摄像机匹配，两个站的数据均获得0.827。

1. *六通道网络：*受[86]中提出的两通道模型的启发，我们通过合并三个颜色通道对其进行了改进。这种方法通过将两个人脸图像直接嵌入到六个通道中并馈入网络的第一层，并具有铰链损耗和一个bitbit二进制输出（如图5所示），从而放弃了两塔功能。 ，它具有更大的灵活性-它的参数是两塔结构的两倍，并且能够使用六个图像通道而不是三个图像通道来学习特征图。但是，它在三个函数中收敛最慢，需要使用L2正则化才能获得更好的性能。我们使用站点1和站点2单摄像机匹配的六通道网络实现的最佳AUC值是0.891和0.818。达到0.846的AUC，
2. *完全卷积结构和SPP池：*在本节中，我们尝试使用完全卷积结构和空间金字塔池（SPP）层来改进上述三种架构（暹罗网络，MatchNet和6通道网络[ 27]）。全卷积CNN（FCN）是所有可学习层都是卷积的CNN。卷积层比完全连接的层具有更少的参数，这将潜在地减少对小型数据集的过度拟合，但在要素中保留更多空间信息。我们在卷积层中用卷积层替换完全连接的层



s

t

a

t

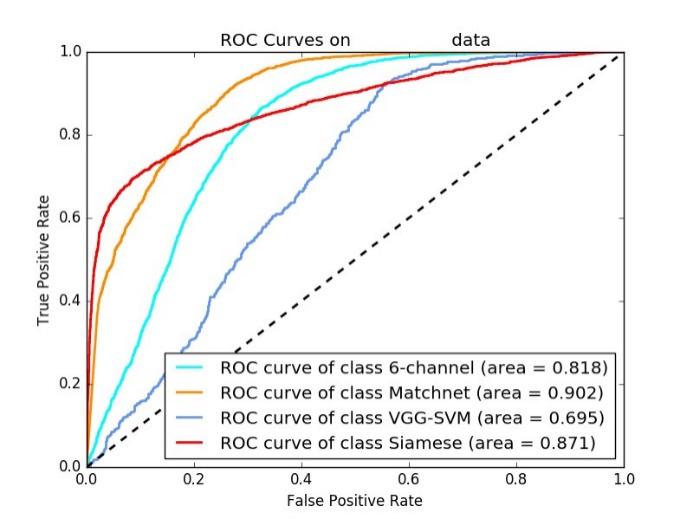
i

o

n

1

（一个）



s

t

a

t

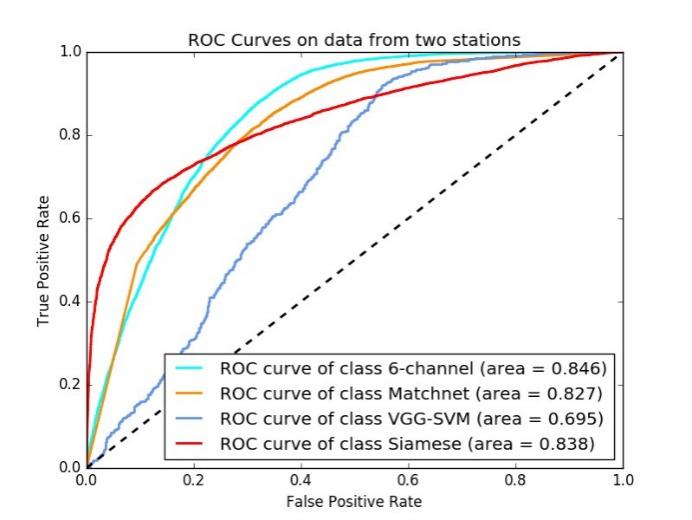
i

o

n

2

（b）



（C）

图6：针对三个数据子集的具有三个匹配基本网的人脸匹配（表V中的第一行）。（a）：车站1（b）：车站2（c）车站1和车站2。

表五：卷积核大小为的模型选择

3 × 3。显示不同网络布局时的测试准确性

|  |  |  |
| --- | --- | --- |
| 参量 | 方法 | 准确性 |
| c（32）-mc（32）-mc（64）-fc（64） | 6通道 | 77.1％ |
|  | 比赛网 | 77.3％ |
|  | 连体 | 78.3％ |
| c（8）-mc（16）-mc（32）-c（32）-米 | 6通道 | 76.7％ |
|  | 比赛网 | 78.1％ |
|  | 连体 | 77.8％ |
| c（8）-mc（16）-c（16）mc（32）-c（32）-米 | 6通道 | 77.5％ |
|  | 比赛网 | 79.3％ |
|  | 连体 | 79.8％ |
| c（8）-c（8）-mc（16）-c（16）mc（32）-c（32）-米 | 6通道 | 76.8％ |
|  | 比赛网 | 75.9％ |
|  | 连体 | 78.5％ |

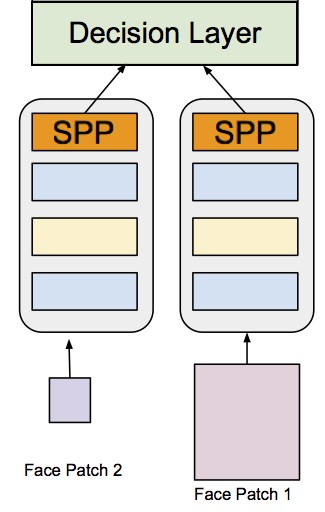


图7：具有SPP层的暹罗网络架构。

在前面的三个模型中，测试几种超参数设置以调整层数和滤波器数，并根据观察到的测试精度选择最佳设置。表V中汇总了性能比较。图8证明，FCN体系结构有效地提高了三种基本网络体系结构（暹罗语，Matchnet，6通道网）的性能（按AUC衡量），分别提高了1％，5％和9％。大约各占4％。此外，我们受到Zagoruyko等人的研究的启发。[86]将SPP层用于补丁匹配，并声称其性能得到了显着改善。与[86]仅使用尺寸相同的图像对通过SPP测试其架构的人相比，我们决定利用网络的完全连接架构，将各种尺寸的人脸图像馈入网络。通过在决策层之前用SPP层替换最后一个最大池化层，我们可以使用修改后的体系结构进一步测试VBOLO数据集中[86]中的假设。我们通过建立三个卷积暹罗网络来简化问题。每个都负责特定分辨率级别的匹配。在这种情况下，我们有

a）低到低，b）高到高，和c）低到高分辨率与三个独立但相同的网络匹配。这三个网络尝试通过脸对接近原始尺寸的方式来学习不同的指标和特征。我们将尺寸小于32 × 32的脸调整为16 × 16，并将其表示为LR脸。那些大于32 × 32的面将调整为64 × 64，并表示为HR面。我们训练和测试三个子网，图7所示的顶部是SPP层。4 ×在每个塔的末端应用4个SPP池。通过使用SPP层和完全卷积的体系结构，我们对AUC进行了轻微的改进（0.1％）。与以前的工作[36]相比，我们通过利用统一的深度特征和度量学习而不是分别优化特征和度量，显着改善了来自站点2的数据的性能。

*i）改进了对较大数据集的预训练方法：*

尽管我们在VBOLO数据集上获得了几套有希望的结果，但仍需要在较大规模的公共数据集上进行更全面的实验。此外，我们面临的挑战是，由于输入的过度拟合和固有维数有限，大多数深度架构都难以很好地利用LR图像。我们通过评估一些较大的普通LR无约束人脸数据集或监视质量的人脸数据集来执行一个项目，以改善训练过程。为了实现这个目标，我们采用[55]中介绍的DCGAN来获得预训练的鉴别器作为特征塔的初始化。使用此方法有两个优点：

* 通过优化LR训练集上的DCGAN，我们可以了解网络如何感知LR图像并通过查看生成器的中间输出来调整参数，例如激活函数和层数，如图9所示。来自过滤器和特征图的信息还不够直观，无法提供训练策略。
* 预先训练的鉴别器可以为一般的LR人脸提供初始权重，可以通过微调将其转移到其他LR人脸数据集，从而稳定和加速训练过程。

GAN鉴别器经过MegaFace挑战训练

2个LR子集，并使用目标数据集（VBOLO，SCface和UCCSface）进行微调。我们比较了从头开始训练的模型和使用DCGAN预训练的完全卷积MatchNet模型。

我们按照以下步骤进行训练和测试划分：对于UCCSface，我们选择90个身份进行训练，选择90个身份进行测试。对于MegaFace Challenge 2 LR子集，我们使用2999个身份进行训练，并使用6699个身份进行测试。对于SCface，我们选择65个身份进行培训，选择65个身份进行测试。我们遵循VBOLO配对和匹配协议进行培训和测试。验证集是训练集中随机选择的20％。所有实验进行了五次，平均错误率。当通过应用DCGAN鉴别器以使用大量LR面的预训练权重开始时，与从头开始训练网络相比，我们观察到验证率更稳定且往往更高。更好的性能

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据集 | 博宝 | UCCSface | MegaFace 2子 | 人脸 |
| DCGAN预训练 | 18.8 / 17.6 | 14.7 / 11.8 | 20.1 / 19.8 | 24.3 / 24.1 |
| 从头开始训练 | 20.7 / 19.5 | 18.8 / 18.6 | 22.5 / 21.8 | 27.4 / 28.5 |

表VI：Matchnet模型的平均错误率

在相同的数据集上实现。对于VBOLO，UCCSface，MegaFace挑战2 LR子集和SCface，与从头开始训练模型相比，我们的错误率降低了9.2％，21.8％，10.6％，11.3％。

我们还确定了DCGAN模型成功融合到LR人脸图像和HR图像上所需的一些体系结构更改。对于更高分辨率的建模，Radford等人。[55]建议采取以下步骤来稳定训练：

* 用跨步卷积（鉴别器）和分数跨步卷积（生成器）替换所有池化层。
* 用ReLU 代替 Tanh 激活 功能或

Leakyrelu功能。

* 在生成器和鉴别器中添加批标准化。

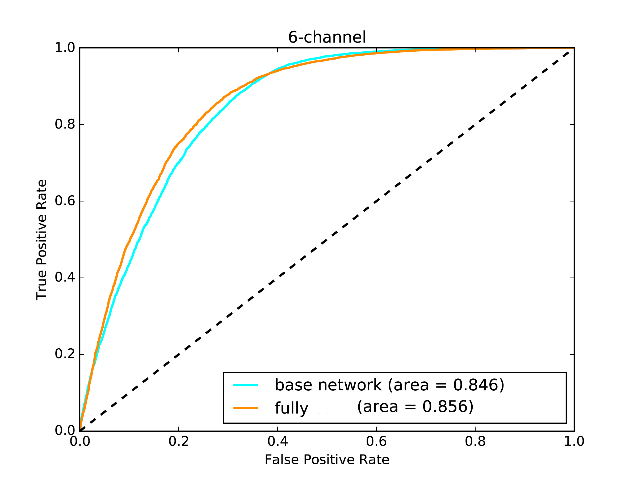
但是，除了鉴别器的最后一层（我们采用S型非线性）之外，我们仅在生成器和鉴别器中使用Tanh激活函数来实现稳定的对抗收敛。批量归一化对于DCGAN收敛到MegaFace LR子集上没有有用的作用，因此未在模型中应用。

IV。 ç 面上污染物和Ç OMMENTARY

在本文中，我们提供了一些新颖的贡献。首先，我们说明了使用最新的超分辨率算法时，LR无约束人脸与LR约束人脸识别之间的性能差距。其次，定义了两个基于LR人脸识别的重要应用场景：野外无约束的LR人脸识别和LR人脸重新识别。对于一般的LR人脸识别，我们利用一种新颖的方法来处理由于探针和图库中人脸图像的质量差异而引起的多维失配。与我们以前的工作相比，我们还设计了不同的深度网络来解决人员重新识别问题，以证明其性能更好[36]，[37]。我们提出一种使用DCGAN预训练的新策略，以获取网络的学习可视化效果并改善大规模数据集的结果。我们在选定的数据集上进行了广泛的实验，证明了结果，并发现尺寸失配是最具挑战性的一点，尤其是在从低到高分辨率的人脸识别任务中。结果表明，我们提出的方法针对不同的任务，有效地工作，并产生令人印象深刻的性能。

R 参考

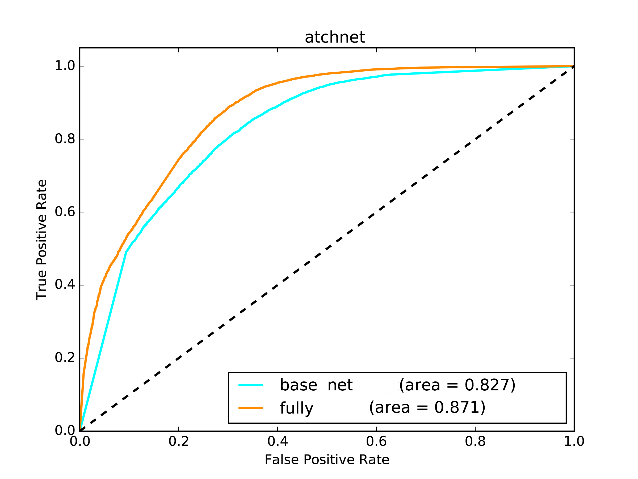
1. E. Ahmed，M。Jones和TK Marks。用于人员重新识别的改进的深度学习架构。在*IEEE计算机视觉和模式识别会议上*，第3908–3916页，2015年。



C

N

N



C

N

N

w

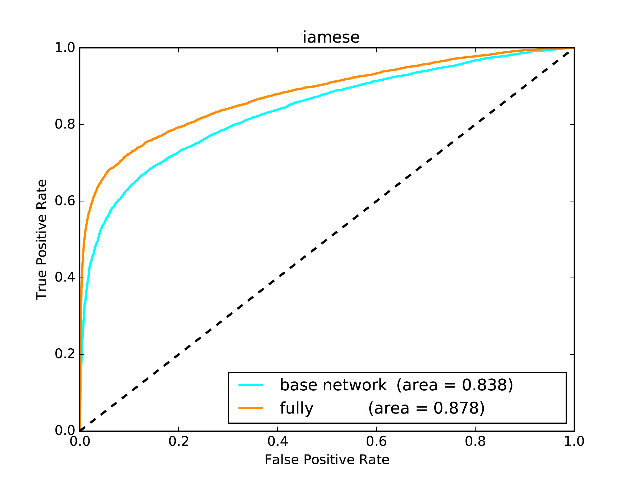
o

r

k

M

（a） （b）

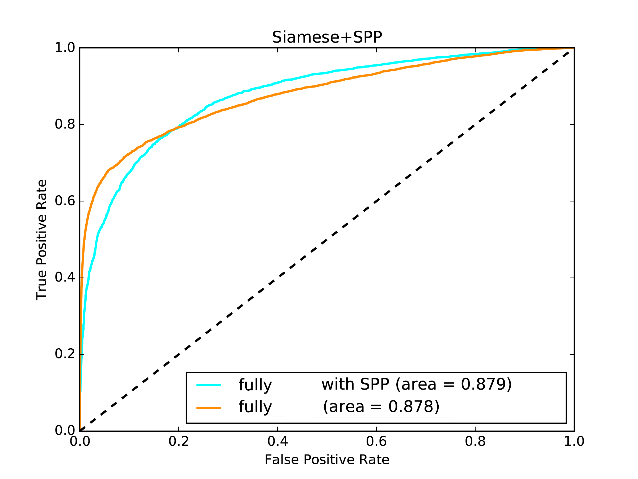


S

C

N

N



C

N

N

C

N

N

（c） （d）

图8：在具有多摄像头和多外观的两个站点的人脸上使用完全卷积网络进行人脸匹配。（a）：6通道。（b）：Matchnet。（c）：连体网。（d）：具有SPP层的全新暹罗语。



图9：DCGAN生成器生成的LR面

*原文*

提供更好的翻译建议