深度耦合ResNet用于低分辨率人脸识别

卢泽*，IEEE高级会员*姜旭东*，IEEE研究员* Alex Kot

***抽象*—监视摄像机捕获的面部图像通常具有低分辨率（LR），这会不利地影响其与高分辨率（HR）画廊图像的匹配性能。包括超分辨率，耦合映射（CM），多维缩放和卷积神经网络在内的现有方法只能产生适度的性能。在这封信中，我们提出了深度耦合ResNet（DCR）模型。它由一个干线网络和两个分支网络组成。主干网络由三种分辨率显着不同的面部图像训练而成，用于提取对分辨率变化具有鲁棒性的判别特征。由HR图像和目标LR图像训练的两个分支网络，作为特定于分辨率的CMs来将HR和相应的LR特征转换到一个其差异最小的空间。分支网络的模型参数使用我们提出的CM损失函数进行了优化，该函数不仅考虑了HR和LR特征的可分辨性，还考虑了它们之间的相似性。为了处理probe图像的各种可能分辨率， 我们在使用同一主干网络的同时训练多对小型分支网络。对LFW和SCface数据库进行的全面评估表明，所提出的DCR模型与现有技术相比，能够始终如一地实现更好的性能。**

***索引*词-卷积神经网络（CNN），耦合映射（CM），特征提取。**

# I. 简介

**F**

由于安全需求，商业应用和执法应用的增长，ACE识别（FR）一直是一个非常活跃的研究领域[1] – [3]。在具有挑战性的条件下，例如遮挡[4]，姿势变化和照明[5]，已经取得了可喜的结果。尽管已经开发出许多用于识别高分辨率（HR）人脸图像的FR方法[6] – [8]，但针对监视系统中的FR的研究很少，因为在这些系统中，HR摄像机不可用或者摄像机与摄像机之间的距离很远主题。在低分辨率（LR）图像的情况下，为HR图像开发的FR方法通常会下降[2]，[9]。

2018年1月15日收到手稿; 于2018年2月22日修订; 接受日期：2018年2月22日。发布日期：2018年2月27日；日期为当前版本2018年3月12日。这项工作是在新加坡南洋理工大学快速富对象搜索实验室中进行的，该实验室得到了新加坡信息通信媒体发展管理局的支持。这项工作得到了新加坡教育部学术研究基金1 RG 123/15的支持。塞巴斯蒂安·马塞尔（Sebastien Marcel）博士是负责协调对此稿的审稿并批准发表的副编辑。*（通讯作者：鲁泽）*

作者来自南洋理工大学电气与电子工程学院，新加坡639798（电子邮件：zlu008@ntu.edu.sg； exdjiang@ntu.edu.sg； eackot@ntu.edu.sg）。

该信函中一个或多个图形的彩色版本可从http://ieeexplore.ieee.org在线获得。

数字对象标识符10.1109 / LSP.2018.2810121

当只有LR探针图像可用时，识别面部仍然是一个挑战。

在这里，我们重点讨论将LR probe面部图像与HR gallery图像匹配的LR面部识别（LRFR）问题。为此任务建议的大多数方法通常可以分为两类。一种方法是通过超分辨率（SR）技术从LR重建HR探针图像并将其用于分类。尽管基于SR的方法（例如[10] – [13]）可以生成视觉上吸引人的HR图像，但是它们在计算上昂贵并且并未针对识别目的进行优化。因此，结果可以进一步提高[2]，[14]-[16]。

另一类是将LR探测器和相应的HR画廊图像同时转换到一个公共特征子空间，在这些子空间中它们之间的距离最小。Li *等。*[2]建议学习两种将不同分辨率的面部图像投影到统一特征空间的矩阵，其中LR图像与其HR对应物之间的差异最小。基于线性判别分析的思想，在[17] – [19]中提出了几种判别子空间方法。Ren *等人没有*使用线性方法*。*[20]通过最小化内核革兰氏矩阵捕获的不相似性，将LR和HR面部图像投影到一个无限的公共子空间中。在[14]中采用了多维缩放（MDS），同时以劣质探针图像和高质量图库图像之间的距离近似于图库图像之间的距离的方式来转换特征。同一作者在[21]中提出了一种基于参考的方法来降低计算成本。文献[16]中提出了两种区分性MDS方法，以充分利用身份信息，包括类间距离和类内距离。他们的新目标函数声称可以扩大类间距离，以确保可区分性。

通常，基于子空间的方法比基于SR的方法具有更好的识别性能。但是，基于子空间的方法通常从图像中提取像素值或尺度不变特征变换作为特征表示。通过使用对分辨率更改具有鲁棒性的功能表示，可以提高其性能。曾因卷积神经网络（CNN）的卓越性能而受到激励[22] *。*[15]训练一个深度卷积网络，分辨率不变CNN（RICNN），通过将真实的HR图像与上采样的LR图像混合，以监督的方式学习分辨率不变特征。尽管RICNN改善了LRFR的性能，但它对探针图像的分辨率变化很敏感，如[16]所示。

在这封信中，我们提出了一种基于CNN的方法，即深度耦合ResNet（DCR）模型，以解决上述问题

1070-9908©2018 IEEE。允许个人使用，但是重新发布/重新发行需要IEEE许可。有关更多信息，请参见http://www.ieee.org/publications standard / publications / rights / index.html。

LU *等。*：用于低分辨率面部识别的深层耦合网络 527

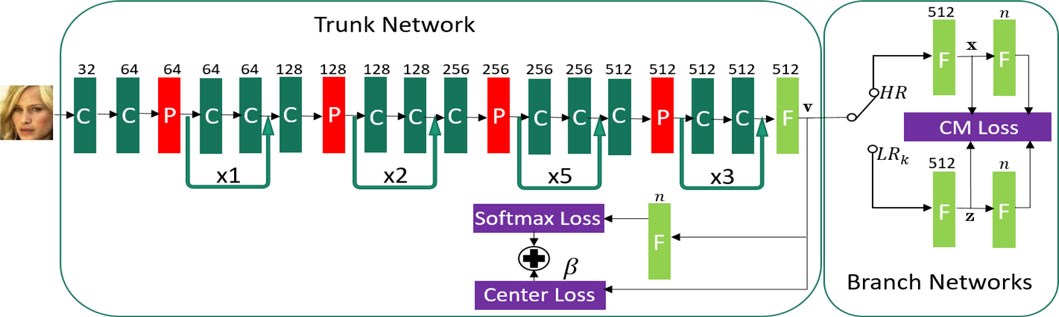


图1.提出的DCR模型的体系结构 中继网络学习由不同分辨率的图像共享的判别特征（用**v**表示），并将分支网络训练为CM（分别由**x**表示HR功能和**z**表示LR功能）。C，P和F分别表示卷积层，最大合并层和完全连接层。卷积层中输出特征图的数量和完全连接层中的输出数量由每层顶部的输出数量表示。“ × *h* ”表示重复*h*次的残差模块。*k*表示LR训练图像的分辨率，*β*是a

中心损耗的缩放参数。

LRFR中的问题。DCR模型的新颖性和贡献来自以下四个方面。

1. DCR模型由一个大主干网络和两个小分支网络组成。
2. 主干网络仅训练一次，以学习不同分辨率的面部图像共享的判别特征。它是基于最近提出的住宅模块构建的[23]。
3. 两个分支网络进行培训，以学习特定于分辨率的耦合映射（CM），以便将HR画廊图像和LR探测器图像投影到最小距离的空间。提出了CM损耗来优化分支网络的模型参数。
4. 实际上，可以有各种分辨率的probe图像与HR画廊图像匹配，提出的DCR模型通过训练多对小分支网络（每对2 MB）同时使用相同的大主干网络来解决此问题。 （105 MB）。分辨率指标方法[24]可用于确定探针图像的分辨率以及要使用的分支网络。

二。DCR 模型

## A. DCR的体系结构

LRFR的关键是提取对分辨率变化具有鲁棒性的特征，并测量HR画廊图像和LR探测器图像之间的相似性。近年来，CNN已被广泛用于FR中的特征学习，并且已经获得了出色的性能，如[23]和[25]。这促使我们采用CNN来学习由不同图像分辨率共享的判别特征，并推导CM以最小化HR画廊图像和特定分辨率的LR对应物之间的距离。提出的DCR模型的体系结构如图1所示。它由一个干线网络和两个分支网络组成。主干网络用于提取对人脸图像分辨率退化具有鲁棒性的判别人脸特征。两个分支网络被训练用以学习CMs，它使主干网络从HR图像中提取的特征向量与对应分辨率k的LR图像的特征向量之间的距离最小。

提出的DCR模型的干线网络是基于[8]中的CNN模型和[23]中提出的ResNet模型构建的。与以前的CNN架构（例如VGG）不同，[23]中的ResNet由残差模块组成，这些残差模块进行信号的可加性合并。He*等。*[23]认为，残余连接对于训练非常深的体系结构具有内在的重要性。ResNet已成为一项开创性的工作。如图1所示，干线网络将面部图像的原始像素作为输入。C，P和F分别表示卷积层，最大池化层和完全连接层。卷积核大小为3 × 3步长为1。最大池化层核大小为2 × 2，步长为2。每个卷积层后跟一个PReLU [26]非线性单元。卷积层中输出特征图的数量和完全连接层中的输出数量由每层顶部的输出数量表示。“ × *h* ”表示重复*h*次的残差模块。

主干网络中倒数第二个完全连接层的输出特征向量**v**形成分支网络的输入。HR特征向量及其对应的LR图像成对馈入分支网络。每个分支网络由两个完全连接的层和一个中间的PReLU单元组成。因此，CM被构成为两个非线性映射。第一个完全连接的层**的**输出特征向量（x和z）分别作为HR和LR图像的特征表示。主干网络输出对人脸图像分辨率变化具有鲁棒性的特征表示，其模型大小约为105 MB。分支网络缩短了HR特征与其特定分辨率的LR特征之间的距离。它的模型大小约为2 MB，占整个模型的1.9％。

## B.训练策略和CM损失

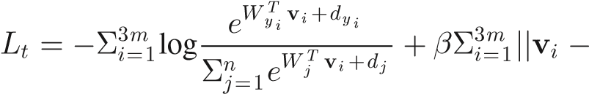
最近发布的CASIA-WebFace [27]数据库用于训练提出的DCR模型。包含9067个人物的434793张图像



图2。 来自CASIA-WebFace的面部图像示例。

主题集，每个主题至少包含14张图像，构成训练集。根据五个稀疏面部点（即双眼中心，鼻尖和两个嘴角）的坐标，通过仿射变换将面部图像标准化为112 × 96像素。我们采用现成的人脸对齐工具[28]进行人脸点检测。图2中显示了一些示例性面部图像。

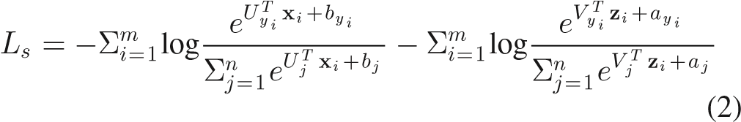
采用了两步训练策略来有效优化DCR模型的参数。在第一步骤中，主干网络是使用三种显著不同的图像分辨率的面部图像的训练，（112 × 96 *，* 40 × 40 *，* 6 ×6）。因此可以将其应用于分辨率从112 × 96到6 × 6的图像的特征提取。类似于[15]，首先将图像下采样为LR图像。之后，在将LR图像馈送到网络进行训练之前，使用双三次插值将其扩大为网络所需的输入大小。在我们的情况下，所需的输入图像尺寸为112 × 96。此过程有效地使LR图像变为模糊了的HR图像。softmax损失和中心损失的联合*L t*用于训练主干网络

**C**

其中*m是*相同分辨率的训练样本数量，*n*表示训练数据中的subject数量。**v** *i* 表示由主干网络从第*i*个训练图像中提取的特征向量。*W j*表示主干网络的最后完全连接层中权重*W的*第*j*列，而*d*是偏差项。*ÿ 我*是类标签为*我*个样本和**Ç** *v ÿ 我*表示*ÿ 我*深特征的第类中心**v**。*β*是比例因子。

第二步，主干网络的模型参数保持不变，并且使用分辨率与目标LR探针图像相似的HR图像（112 × 96）和LR图像训练两个分支网络。提出了CM损失函数，以监督分支网络的训练。

在分支网络的训练过程中，我们首先希望最大化每个人脸图像与来自于HR或LR功能不同类别的邻居之间的距离。下式定义的softmax损耗有助于类间分离：



其中 **x** *i*和**z** *i分别* 表示分支网络从第*i*个HR和LR图像中提取的特征向量。*U j，V j*表示分支网络的最后完全连接层中权重*U，V的*第*j*列，而*a，b*分别是HR功能及其LR对应项的偏置项。

为了保持类内的紧凑性，我们将最小化一幅图像的特征与其相邻的类内HR、LR图像的特征。在以下等式中定义的中心损耗可最大程度地减少组内差异：

- - （3）

其中**Ç** *X ÿ 我*和**Ç** *Ž ÿ 我*表示*ÿ 我* HR的第类中心设有**X**和LR设有**Ž**分别。

对于LRFR任务，重要的是要确保LR特征和相应的HR特征之间的一致性。具体而言，最终的LR和相应的HR特征向量应尽可能接近。这可以通过最小化的欧几里得损失来实现：

*大号Ë* =Σ *MI* = 1 || **x** *i* − **z** *i* || 22 *。* （4）

考虑到以上三个标准，用于优化分支网络参数的CM损耗可构造为：

*大号CM* = *大号小号* + λ *大号Ç* + *αL è* （5）

其中λ 和*α*是用于平衡三个损失函数的两个因子。以这种方式，不仅考虑了HR和LR特征的可辨别性，而且考虑了HR和相应的LR特征之间的关系。

## C.匹配人脸图像

训练完成后，可以将HR画廊图像和LR探针图像都输入DCR以获取其特征表示。可以通过计算面部相似度来执行面部验证或识别。大型干线网络包含整个DCR模型参数的98.1％，而小型分支网络包含所有DCR模型参数的1.9％。在实际应用中，监视摄像机捕获的探测图像可以具有许多不同的LR。因此，在DCR中训练了多对分支网络，以处理LR探针图像的不同分辨率。因此，多个分支网络极大地提高了DCR模型将HR图像匹配到探头面部图像的不同分辨率的有效性和效率。

# 三，Ë XPERIMENTS

在LFW [29]和SCface [30]数据库上进行了广泛的实验，以评估用于将LR探针图像与HR画廊人脸图像进行匹配的拟议DCR模型。LFW和SCface数据库都是无约束环境中FR的广泛使用的基准。在LFW上，DCR的面部验证性能分别与VGGFace [25]，LightCNN [31]，ResNet [8]以及它们的微调版本VGGFace-FT，LightCNN-FT和ResNet-FT进行了比较。我们使用与DCR模型的训练过程相同的纪元数，通过相同的LR训练数据对预训练模型进行微调。开源深度学习工具包Caffe [32]用于微调深度模型。在微调期间，批次大小设置为128。使用递减的学习比率对模型进行微调，并且当损失不再减少时，微调停止。此外，为了评估所提出的分支网络对其他网络的有效性，我们将中继网络替换为LightCNN和VGGFace，并将获得的模型分别命名为Coupled-LightCNN和CoupledVGGFace。此外，还报告了由三种不同分辨率的面部图像训练的干线网络的性能，以进行比较。使用不同大小的LR探针图像进行测试。请注意，使用不同大小的LR探针图像进行测试。请注意，使用不同大小的LR探针图像进行测试。请注意，

|  |
| --- |
| LU *等。*：用于低分辨率面部识别的深层耦合网络 529 |

# 表一

˚F ACE V ERIFICATION 甲作者CCURACY d IFFERENT 甲PPROACHES ù SING d IFFERENT P ROBE 小号IZES ON LFW

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 探头尺寸 | 8 × 8 | 12 × 12 | 16 × 16 | 20 × 20 | 112 × 96 |
| ResNet [8] | 72.7 | 84.1 | 92.3 | 95.4 | 98.7 |
| ResNet-FT | 88.9 | 93.8 | 95.9 | 96.8 | 98.8 |
| 中继网络 | 92.2 | 93.6 | 95.5 | 96.8 | 98.4 |
| DCR（联运） | 93.6 | 95.3 | 96.6 | 97.3 | 98.7 |

使用双三次插值法将脸部图像重新缩放为网络所需的输入大小，然后再将其输入到网络中进行特征提取。在SCface上，除了上述9种CNN模型之外，还尝试了四个最新的LRFR方法，即MDS [14]，[21]，DMDS，LDMDS [16]和RICNN [15]，以与DCR模型进行比较。在所有实验中，将（1）中的*β*，（5）中的λ 和*α*的值设置为0.008。

## A. LFW实验

LFW数据库包含5233个主题的13233张图像。该数据库中的图像显示出丰富的个人姿势，照明和表情变化。近年来，LFW已被广泛用于无约束FR的研究。我们遵循[29]中的“无限制的，标记的外部数据结果”协议，并通过对View 2数据的十倍交叉验证方案计算平均验证准确性。使用与CASIA-WebFace图像相同的方法对面部图像进行归一化和对齐。对于人脸验证范例中的两个图像，我们将第一个图像作为HR（112 × 96）画廊图像，然后将第二个图像降采样为8 × 8、12 × 12、16 × 16或20 ×20作为LR探针图像。相同大小的CASIA-WebFace图像用于训练相应的分支网络。相同的人脸图像用于LightCNN，VGGFace和ResNet模型的微调。PCA和余弦距离用于计算两个特征之间的相似度。在十倍交叉验证中，使用LFW数据的九个训练倍数计算PCA的总散布矩阵。VGGFace [25]，VGGFace-FT，Coupled-VGGFace，LightCNN [31]，LightCNN-FT，CoupledLightCNN，ResNet [8]，ResNet-FT，主干网络和在LFW上建议的DCR模型的人脸验证准确性如图所示。表I.在最后一列中，还提供了与画廊图像具有相同分辨率的HR探针图像的精度。

## B. SCface上的实验

SCface数据库包含使用五个质量各异的视频监控摄像机在不受控制的室内环境中拍摄的130名受试者的图像。对于每个被摄对象，通过监视摄像机和一个正面面部照片，分别在三个距离，4.20 m（*d* 1），2.60 m（*d* 2）和1.00 m（*d* 3）的三个距离处拍摄15张图像（每个距离五张图像）。数字拍摄的图像

# 表二

FRř ATES作者d IFFERENT 甲PPROACHES AT d IFFERENT d ISTANCES ON SC FACE

|  |  |  |  |
| --- | --- | --- | --- |
| 距离 | *d* 1 | *第* 2天 | *d* 3 |
| MDS [14]，[21] | 60.3 | 66.0 | 69.5 |
| DMDS [16] | 61.5 | 67.2 | 62.9 |
| LDMDS [16] | 62.7 | 70.7 | 65.5 |
| 里肯[15] | 23.0 | 66.0 | 74.0 |
| LightCNN [31] | 35.8 | 79.0 | 93.8 |
| 轻型CNN-FT | 49.0 | 83.8 | 93.5 |
| 耦合LightCNN | 50.5 | 85.0 | 94.0 |
| VGGFace [25] | 41.3 | 75.5 | 88.8 |
| VGGFace-FT | 46.3 | 78.5 | 91.5 |
| 耦合VGGFace | 62.3 | 91.0 | 94.8 |
| ResNet [8] | 36.3 | 81.8 | 94.3 |
| ResNet-FT | 54.8 | 86.3 | 95.8 |
| 中继网络 | 52.0 | 89.5 | 96.3 |
| DCR（耦合ResNet） | 73.3 | 93.5 | 98.0 |

相机。以下在实验设置[16]，正面面部照片的图像被用作由监控摄像机在拍摄距离图像画廊和图像*d I，I* = 1 *，* 2 *，* 3 被用作探针的图像。我们采取大小的CASIA-WebFace图像112 × 96 HR图像和那些的112 × 96，30 × 30和20 × 20作为在距离LR图像为分支网络的训练*d* 3， *d* 2，和*d*1，分别。与[16]相同，从SCface数据库中随机选择130个主题中的50个，用于分支网络的微调和PCA的训练。其余科目都在测试中。因此，训练和测试集之间没有身份重叠。来自CasiaWebface和SCface数据集的相同面部图像用于LightCNN，VGGFace和ResNet模型的微调。最近邻分类器用于对所有探针图像进行分类。我们报告了MDS [14]，[21]，DMDS，LDMDS [16]，RICNN [15]，VGGFace [25]，VGGFace-FT，Coupled-VGGFace，LightCNN [31]，LightCNN-FT，Couple -LightCNN，ResNet [8]，ResNet-FT，干线网络和表II中建议的DCR模型。

我们可以从表I和II中观察到：首先，分支网络大大提高了LightCNN，VGGFace和干线（ResNet）网络的性能；其次，对于LFW和SCface数据集上的探针图像的不同分辨率，所提出的DCR模型获得的FR准确度要比最新方法一致。对于非常LR的探头图像，性能提升非常重要。

# IV。ç ONCLUSION

在这封信中，我们提出了一种基于CNN的新颖方法，称为LRFR任务的DCRN模型。它首先通过类似ResNet的网络（主干网络）提取具有不同分辨率的面部图像所共享的区分特征。之后，通过分支网络学习CM，将特定分辨率的HR图像和相应的LR图像的特征投影到一个公共子空间中，在这些子空间中，它们的距离被最小化。在LFW和SCface数据集上进行的实验表明，所提出的DCR模型与现有技术相比，能够始终如一地且性能更好。

原文

提供更好的翻译建议