双定向胶囊网络可实现非常低分辨率的图像识别

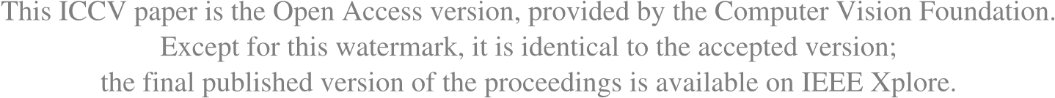
Maneet Singh，Shruti Nagpal，Richa Singh和Mayank Vatsa IIIT-德里，印度

{ 鬃毛，兽皮，rsingh，玛雅克} @ iiitd.ac.in

# 抽象

*低分辨率（VLR）图像识别对应于对分辨率为* 16×16的*图像进行分类或更少。尽管当以很大的对距距离（例如，监视场景）或从广角移动摄像机捕获对象时，它具有广泛的适用性，但受到的关注有限。这项研究提出了一种新颖的双重定向胶囊网络模型，称为DirectCapsNet，用于解决VLR数字和面部识别问题。所提出的体系结构利用胶囊和卷积层的组合来学习有效的VLR识别模型。该体系结构还合并了两个新颖的损失函数：（i）建议的HR锚损失和（ii）建议的目标重建损失，以克服VLR图像中信息量有限的挑战。拟议的损失在训练过程中使用高分辨率图像作为辅助数据，以“直接”进行区别性特征学习。针对VLR数字分类和VLR人脸识别进行了多次实验，并与最新算法进行了比较。拟议的DirectCapsNet始终展示最先进的结果；例如，在UCCS人脸数据库中，当*16×16 *图像与* 80×80 *图像匹配。*

# 简介

在典型的监视场景中，通常会在较大的距离内捕获图像，从而使目标区域具有非常低的分辨率（VLR），通常小于16×16 [ 33 ]。图1 （a）显示了VLR识别的实际示例应用，其中感兴趣的区域可以是人脸，可疑物体或移动车辆的车牌号。这些样本证明了问题的艰巨性，其中VLR识别的一些关键挑战是信息内容有限和模糊不清。VLR识别在图像标记中也具有适用性，在该图像标记中，在框架中捕获了多个对象/人物，并且这些实体中的每个实体的分辨率都很小。



(

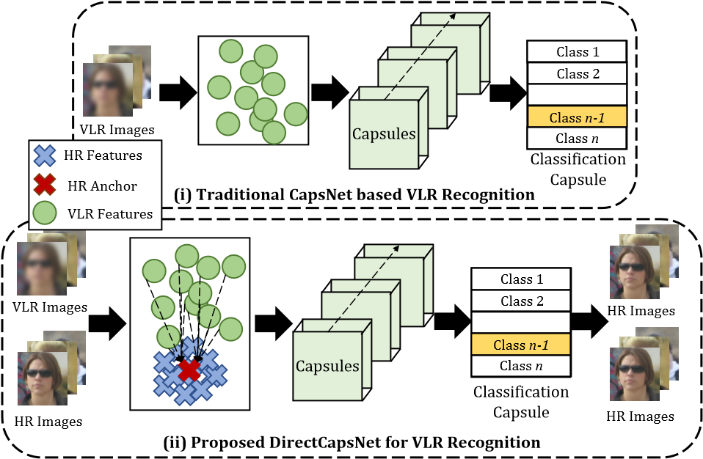
i) Digit Classification

(

ii) Face Recognition



1. VLR识别的实际应用。图片来源：（i）互联网，（ii）UCCS数据集[ 24 ]



1. 建议的双重定向胶囊网络（DirectCapsNet）

图1：拟议的DirectCapsNet利用HR样本，通过拟议的HRanchor损失和目标重建损失，*指导* 对VLR图像识别更有意义和更具区别性的特征的学习。

Netzer *等。*[ 17 ]证明了人类在识别在真实环境中捕获的VLR数字时表现不佳。对于街景房门号码（SVHN）数据集，作者观察到人类对101-125 像素高度的样本的准确性为百分之几。另一方面，性能下降到82 *。*当对分辨率非常低的样本（即高度不超过25像素的图像）进行分类时为0％±2％，从而恢复了问题的挑战性。直接上采样经由内插可以被看作是用于识别VLR一个可能的解决方案，但是，多个研究已经证实由于需要大的放大系数[性能差14 ，25]和

1 340

噪声的可能引入，也可以在图1 （a）（i）中观察到。此外，在文献中，研究人员还证明了在高分辨率（HR）图像（包含高信息含量）上训练的模型无法在（V）LR图像上表现良好[ 25 ]。当前稀缺的解决方案以及VLR识别的广泛应用使其成为一个重要问题，需要引起人们的关注。

这项研究提出了一种新颖的基于胶囊网络的VLR图像识别模型。Hinton *等。*[ 7 ]提出了学习“胶囊”，它表示实例化参数的向量，以便更有效地编码输入。实例化参数可以构成图像的属性，例如*相对于该实体的隐式定义的规范版本的视觉实体*的*姿势，光照和变形* [ 7]。我们相信，这些参数对于图像的分辨率将是不变的，因此展现了对VLR识别有用的潜力。由于VLR图像中的信息内容有限，因此VLR识别模型也可以从信息丰富的HR样本中受益。为此，我们提出了*双重定向胶囊网络*（称为*DirectCapsNet*）（图1 （b）），以学习由HR样本指导（或指导）的VLR识别的有意义的功能。这项研究的贡献如下：

* 一种新颖的双定向胶囊网络（*DirectCapsNet*）模型，提出了VLR识别，其*指示* 从包含朝向HR图像的更有意义的和判别特征有限信息的VLR图像学的特征。
* 为指导VLR识别模型提出了两种损失：（i）HR锚固损失和（ii）目标重建损失。为特征学习模块建议了HR锚丢失，它将特定类别的VLR特征推向该类的代表性HR特征（锚）。在分类模块处利用目标重建损失，其中从VLR图像的胶囊输出中重建HR图像，从而迫使相同类别的VLR和HR图像的胶囊相似。
* 实验结果和分析证明了建议的DirectCapsNet模型在VLR数字分类和VLR人脸识别方面的优势。

在SVHN [ 17 ]，CMU Multi-PIE [ 6 ]和UCCS [ 24 ]数据库上进行了实验，并使用最新算法进行了比较。所提出的模型在具有挑战性的UCCS人脸数据库上的准确率超过95％。在SVHN数据库上，使用8×8 VLR图像可获得约84％的分类精度，与现有结果相比，几乎提高了27％。

# 相关工作

目前已在低分辨率识别领域进行了若干改进[ 11 ，14 ，18 ，31 ]; 但是，相对低分辨率（VLR）识别的领域仍然相对较少探索。如前所述，超低分辨率（VLR）识别是指以16×16 或更小的分辨率识别感兴趣的区域。由于在给定的VLR图像的有限信息内容，一个潜在的解决方案是超解析或合成其较高分辨率的图像[ 20 ，28 ]，然后将其用于识别。而存在于超分辨率或合成的算法[大量文献13 ，23 ，29 ]，其中大多数主要集中在生成图像的视觉质量上，而不是识别任务上。Zou和Yuen [ 33 ]提出了一种最初的超分辨率技术，特别着重于VLR人脸识别。所提出的算法利用了基于视觉质量的约束来实现高质量的HR合成，并使用了区分性约束来学习对识别有用的特征。辛格*等。*[ 25 ]提出了一种身份感知的面部合成技术，用于从给定的LR输入中生成HR图像。合成的图像被提供给现成的商业（COTS）系统进行识别。

除了基于超分辨率的技术外，在文献中，研究人员还提出了用于通过使用从HR图像中提取的信息来*增强*或*改善* VLR图像所学特征的算法。例如，Bhatt *等。*[ 2 ]提出了一种基于集成的共转移学习算法用于人脸识别。共转移算法通过利用HR图像信息增强VLR分类，在共训练和转移学习的交叉点进行操作。Wang *等。*[ 30提出了用于VLR识别的稳健的部分耦合网络。HR图像在训练过程中用作“辅助”数据，以学习更多区分性信息，这在VLR图像中可能不可用。如通过多次实验所证明的，在训练时使用HR图像可增强学习的VLR功能，从而提高识别性能。Mudunuri和Biswas [ 16 ]提出了一种基于参考的方法以及多维缩放，用于学习HR和VLR图像的公共空间。最近，李*等人。*[ 14 ]通过学习HR和LR样本的公共特征空间，分析了用于LR和VLR人脸识别的不同度量学习技术。Ge *等。*[ 4提出了一种用于（V）LR人脸识别的选择性知识蒸馏技术。在HR面部图像上训练的基础网络用于为（V）LR CNN模型选择最有信息的面部特征，以增强（V）LR特征和分类性能。

在文献中，通过学习HR和VLR样本之间的共享表示，VLR识别算法已从HR样本中受益[ 30 ]或通过



（一种）



（b） （c）

图2：来自（a）SVHN数据集，（b）CMU Multi-PIE数据集和（c）UCCS数据集的样本HR和VLR图像。HR图像（第一行）包含较高的信息内容，这在VLR样本（第二行）中通常会丢失。

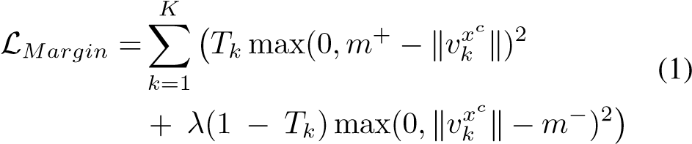
将HR数据学习到的模型信息传递到VLR识别模型[ 4 ]。与从VLR图像中独立学习的算法相比，通过在训练时利用HR图像中的其他信息，此类算法能够学习更多的判别和有意义的特征。本研究提出培训过程中利用辅助HR样品*直接* 在VLR向更丰富的人力资源的功能特点，通过新颖的DirectCapsNet模型。

# 建议的双定向胶囊网络

如图2 所示，非常低的分辨率（VLR）识别问题受到输入图像中信息内容有限的挑战，这常常导致缺乏可用于识别/分类的区分特征。为了克服这一挑战，我们提出了一种新颖的双重定向CapsNet，称为*DirectCapsNet*。DirectCapsNet通过两种方式指导VLR表示，从而增强了VLR表示：通过提议的（i）HR锚定损失和（ii）有针对性的重建损失，两者均使用HR图像提供了额外的监督。HR信息用于指导/引导框架以甚至从具有有限信息内容的VLR图像中提取判别表示。这是通过使用HR锚丢失来实现的，它使VLR图像的表示更接近其相应HR样本的表示。在分类阶段还通过有针对性的重建损失来强制执行此操作，这将促进同一类别的HR和VLR样本的相似功能。由于所提出模型的基本架构是胶囊网络，因此我们首先简要说明其功能，

## 初步：胶囊网络

Hinton *等。*[ 7 ]提出了*胶囊*的概念作为学习表征的有效方法。它由Sabour *等人*进一步开发*。*[ 22 ]，其中介绍了胶囊网络（CapsNet）进行分类。胶囊是“一组神经元，其活动矢量代表特定类型的实体（例如对象或对象部分）的实例化参数”。换句话说，不是单个标量输出，而是每个封装输出一个向量，其值称为活动向量。每个胶囊矢量（长度ķ *。* ķ 2）中的范围是有界的[0-1] 。Sabour *等。*[ 22 ]提出了胶囊之间的动态路由的概念，其中多层胶囊被堆叠以用于对象分类。最后一层包含尺寸为*k* × *m*的分类胶囊，其中*k*为类别数，*m*为胶囊尺寸。对于给定的输入，预测类别是与具有最大活动矢量（长度）的胶囊对应的类别。为了学习有效的分类模型，使用余量损失来学习网络。给定一个*ķ*类的问题，与作为输出 *ķ 个*类胶囊，用于将输入*X Ç*（属于类*Ç*），和*Ť ķ*为对应于该标签*ķ 个*类，CapsNet的边缘损失被定义为：

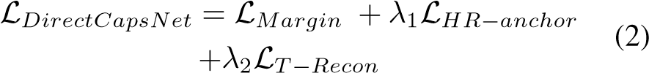


其中，*Ť ķ* ∈{0 *，* 1} ，即，输入采样是否属于类*ķ*（*Ť ķ* = 1 ）或否（*Ť ķ* = 0 ）。*米*+和*米*-对应于用于增加类内的相似性和减少级间的相似性，分别为正的和负的裕量，并且*λ*是用于控制每个术语的重量恒定。上述损失（等式1 ）促进了更长的胶囊长度（k *v k* k）以获取正确的类别，并为与其他类别相对应的胶囊缩短长度。在文献中相对较少地探索胶囊网络，其架构或损失功能的修改有限或没有修改。它们已用于脑肿瘤检测[ 1 ]，海草检测[ 9 ]，生成合成数据[ 10 ]和图像分类[ 32 ]。胶囊网络对给定输入的实例化参数进行编码，因此具有成为VLR图像识别合适网络的潜力。

## 拟议的DirectCapsNet

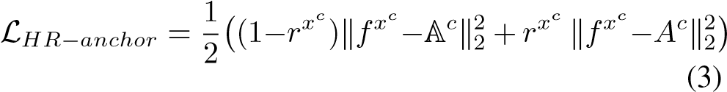
|  |
| --- |
| 图3：针对SVHN数据集提出的双重定向胶囊网络（DirectCapsNet）的体系结构[ 17 ]。给出了给定类别的HR锚损失的示意图。通过指导模型学习判别性和信息丰富的特征，HR图像可用于补充VLR识别模型所学习的特征。 |

如图3 所示，建议的DirectCapsNet网络可以分为三个部分：（i）输入，（ii）特征提取和（iii）分类。在训练时，输入包括HR和VLR样本。特征提取模块由卷积层和建议的HR锚损失组成，分类模块由胶囊网络和建议的目标重建损失组成。通过建议的（i）HR锚损失和（ii）有针对性的重建来实现双向，拟议的DirectCapsNet致力于在辅助HR样本的帮助下为VLR输入学习有意义的，功能丰富的表示。建议的DirectCapsNet的损失函数公式为：



其中，*λ* 1和*λ* 2用于平衡HR锚的权重和针对保证金损失的目标重建损失。边距损失在类别之间引入了可区分性，而HR锚点损失和定向重建损失则在功能和分类级别上强制执行信息丰富的表示。在测试时，对于给定的VLR输入，将选择具有最大长度的类囊作为给定输入的类。必须注意的是，DirectCapsNet在模拟实际情况时仅在训练时才使用HR样本，并在测试过程中使用给定的VLR图像进行操作。正如本节其余部分将要说明的那样，建议模型的每个组件都有助于学习VLR识别的判别特征。

建议的HR锚丢失：图3 中的输入样本是来自SVHN数据集的HR（32×32×3 ）和VLR（8×8×3 分辨率放大为HR分辨率）图像[ 17 ]。VLR图像中信息内容的局限性使得难以提取区分性信息，从而常常导致识别无效，这也是在人类中也观察到的现象[ 27 ]。拟议的HR锚丢失通过将VLR功能推到更靠近其HR计数器部件来解决此挑战。这样即使在信息有限的情况下，也可以确保学习用于VLR识别的区分空间。对于属于类*c的*，具有特征*f x*的输入*x cc*从卷积层获悉，HR锚损失公式为：



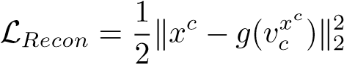
其中，*r x c*是表示样本分辨率的二进制变量，即，对于HR样本，*r x c* = 1 ，对于VLR样本，*r x c* = 0 。由于HR样本仅在训练期间使用，因此此信息随时可用。*f x c表示*从特征模块中的卷积层提取的特征，*A c*和A *c* 均表示类*c*的HR锚，用于增强VLR表示。具体来说，A *c*“ HR锚”是指处于恒定状态的HR锚，而“ *A c”*表示参数形式的HR锚，需要对其进行优化。特定类别的HR锚对应于属于该类别的所有HR样本的平均特征向量。给定一个VLR样本（*r x c* = 0 ），等式的第一部分处于活动状态，其中*c*类的HR锚辅助VLR特征*f x c*靠近锚点，从而有助于学习对分类。

对于HR样本（*r x c* = 1 ），等式3的后半部分

）变为活动状态，同时更新了HR锚和功能。

拟议的HR锚丢失是学习HR锚和学习更接近HR特征空间的VLR特征的组合，以便学习判别性VLR特征。第一项试图将VLR功能指向HR锚点，第二项从HR功能中学习代表性的HR锚点。重要的是要注意，VLR功能在锚生成中没有贡献，因为HR锚在第一项中是恒定的。这确保了VLR功能是针对更高质量的HR功能，而不是相反的方向。因此，公式3 借助HR样本促进了信息性VLR功能的学习。

拟议的目标重建损失：方向的第二种形式是通过分类模块（胶囊网络）的目标重建损失（图3 ）来确定的。有针对性的重建损失促进了对HR和VLR样本的相似分类胶囊的学习。如前所述，胶囊是对输入样本[ 22 ] 的实例化参数进行编码的向量。对于给定的输入，胶囊的激活被称为活动向量。为了重建，仅选择目标类别的活动向量并将其用于重建输入样本。对于属于类*c*的输入图像*x c*，重建损失在数学上公式为：

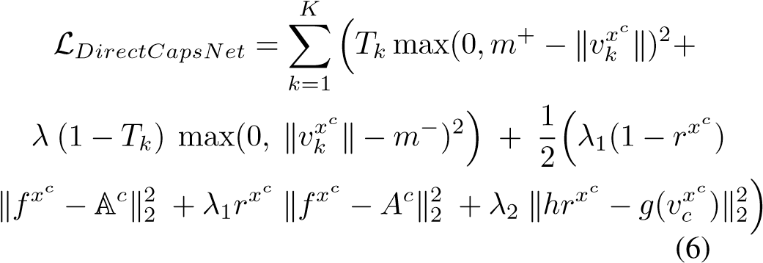
 （4）

其中，是的分类胶囊的活性矢量*Ç 个*类的输入*X Ç*，和*克*（*。*）指的是重建网络。重建损失尝试对能够解释输入图像的实例化参数进行编码，从而能够重建输入。直观上，我们认为HR样本及其对应的VLR样本的实例化参数应该相似。因此，为了纳入第二层次的指导，在建议的DirectCapsNet中引入了目标重建损失。

目标重建损失会在重建网络的输出处强制执行VLR图像的HR对应部分。无论是HR还是VLR输入，都将重构的样本强制为HR图像。对于输入*x c*，目标重建损失可以写成：

 （5）

其中，*HR X Ç*是对应于输入HR / VLR样品和HR图像是活性矢量*Ç 个*类。对于HR输入图像，公式5确保在重建网络的输出处重建HR输入。对于VLR图像，将其HR对应部分作为重建网络的目标。由于重建网络在最终分类胶囊上运行，因此目标重建损失将HR和VLR样本推向具有由HR样本驱动的相似的胶囊活性矢量。因此，通过指示模型从提取的VLR特征中重建HR样本，重建损失促进了在分类阶段直接为HR和VLR样本学习相似的胶囊特征。

结合公式3 和5 来更新公式1 ，对于输入*x c*（属于*c*类），建议的DirectCapsNet的损失函数写为：

## 实施细节

DirectCapsNet已使用NVIDIA Tesla P-100 GPU上的PyTorch框架以Python实现。Adam优化器[ 12 ]已用于学习模型。的HR-锚损耗（重量*λ* 1方程6 ）被设置为1个*Ë* -3 ，和目标重建损失（重量*λ* 2方程6 ）被设置为1个*Ë* - 5 。对于边缘损失（正和负边缘*米*+和*米*-方程1 ）分别设定为0.9和0.1。如图3 所示，对于所有实验，DirectCapsNet模型都包含*n个*卷积层，然后是两个胶囊层。HR锚损失应用于DirectCapsNet的最后卷积层。最终的胶囊层连接到三个完全连接的层的重建网络。对于其中HR样本大于例96×96 ，三与卷积层[16 *，* 32 *，* 128] 过滤器具有32个样本的批量大小使用。如果HR样本较小，则使用具有128个过滤器的卷积层，批量大小为100个样本。在卷积层之间使用*ReLU*激活功能以及批处理归一化[ 8]。所有模型均已从头进行了培训，并且未使用任何预培训的网络。

# 实验与实验方案

已对DirectCapsNet提出的三个非常低分辨率（VLR）识别问题进行了评估：（i）VLR数字识别，（ii）VLR人脸识别和（iii）无约束VLR人脸识别。有关每个案例研究的数据集和协议的详细信息如下：案例研究1-VLR数字识别：街景房门号码（SVHN）数据集[ 17 ]已用于VLR数字识别。数据集包含[0-9] 范围内的数字的真实图像。预定义的基准协议已用于给定的10类问题，其中73,257位数字用于训练，而26,032位数字用于测试。对于VLR识别，与现有协议[ 30 ] 一致，使用了32×32 HR图像，并且使用8×8 VLR图像。结果按前1名和前5名的准确性进行报告。

案例研究2-VLR人脸识别：VLR人脸识别在图像标记场景或单个图像中捕获多个人的情况下具有直接的适用性。对于此特定案例研究，已在CMU Multi-PIE数据集[ 6 ] 上进行了实验，该数据集模拟了受约束的设置。根据现有协议[ 25 ]，使用了237个主题。每个对象一个图像添加到由HR图像组成的训练集/图库中，每个对象一个图像添加到测试集/探针（VLR）中。HR图像的分辨率为96×96 ，VLR图像的分辨率为8×8 和16×16 。报告结果使用rank-1识别准确性进行报告。

案例研究3-无约束的VLR人脸识别：无约束的VLR人脸识别在监视场景中具有广泛的适用性，其中VLR的人脸图像通常包含姿势，照明和遮挡等其他变化。已经在两个数据集上进行了实验：（a）无约束的监视设置的不受约束的大学生（UCCS）数据集[ 24 ]，以及（b）半约束的设置具有姿势和光照变化的CMU Multi-PIE数据集[ 6 ]。

UCCS数据集包含大学生的图像，这些图像是使用相距100至150米的远程高分辨率监控摄像头拍摄的。图像显示了学生在上课时在校园里走来走去。较大的隔离距离和不受约束的数据性质可模拟现实世界的监视设置。数据集包含1732个身份的标记子集。与现有的协议相一致[ 4 ，30 ]，含有该顶部180的身份（在图像的数量而言）的子集被用于评价。根据协议，每个受试者的图像被分为对应于训练：测试的4：1 比例。VLR图像的分辨率为16×16 ，而HR图像的分辨率为80×80 像素。

如上所述，CMU Multi-PIE数据集[ 6 ]包含具有姿势，表情和照明变化的图像。根据现有协议[ 16]，在这种情况下，将针对VLR图像的姿势和照明变化执行人脸识别研究。涉及50个主题的图像用于训练，其余主题的图像构成测试集。在我们的实验中，我们不使用训练集，而仅使用测试集的图库图像来训练建议的DirectCapsNet模型。画廊包括正面图像（用于训练建议的模型），探针（测试集）是具有不同姿势（数据集的“ 05 0”）和照明的图像。在五对不同的照明条件和平均等级1识别上进行了实验。表1：SVHN数据集[ 17 ]的Top-1和top-5精度（％），用于VLR数字识别（8×8 ）。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 算法 | 准确性 （％） | |
| 前1名 | 前5 |
|  | CNN（VLR）（2016）[ 30 ] | 45.29 | 66.78 |
| RPC网（2016）[ 30 ] | 56.98 | 70.82 |
| Proposed | CapsNet（HR） | 77.82 | 87.86 |
| CapsNet（VLR） | 79.19 | 88.89 |
| DirectCapsNet-（人力资源损失） | 82.42 | 90.15 |
| DirectCapsNet-（目标侦察。） | 81.95 | 90.35 |
| 拟议的DirectCapsNet | 84.51 | 91.20 |

已经报道了准确性。与一致[ 16 ]，则HR图像的36×30 分辨率，而VLR图像具有的分辨率18×15 ，15×12 ，12×10 ，和10×9 。

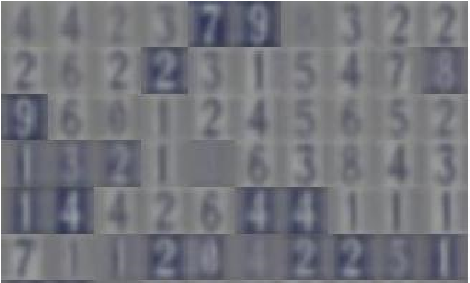
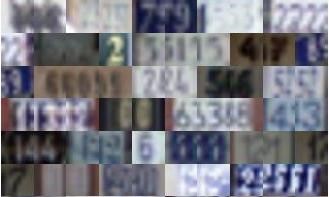
图2 展示了三个案例研究中使用的数据集中的一些HR和VLR图像。双三次插值用于从HR转换为VLR，反之亦然。在训练时，HR和VLR对用于目标重建损失。通过引入亮度变化，沿y轴翻转以及随机裁剪来应用数据增强。在测试时，仅提供VLR图像进行分类。

# 结果与分析

表1 - 3 和图4 - 6 本的结果为三个案研究：（ⅰ）VLR数字识别，（ⅱ）VLR面部识别，以及（iii）无约束的VLR人脸识别。为了证明每个组件的有效性，还对建议的DirectCapsNet进行了分析。由于现有协议已用于分析，因此直接从相应出版物中报告了结果。

DirectCapsNet的消融研究和分析：已经对SVHN数据集进行了实验，以分析所建议的DirectCapsNet的每个组件，并激励它们包含在最终模型中。从表1可以看出，在VLR图像（CapsNet（VLR））上进行训练时，原始CapsNet模型（具有边际损失）达到了top.-1分类准确度，为79.19％，与原始CNN架构（45.29％）相比有很大改进）[ 30 ]。改进的性能促进了将胶囊网络用于VLR识别的任务。与文学一致[ 22]，我们相信，由于CapsNet尝试对数据的实例化参数进行编码，因此会导致学习特征始终保持微小变化不变，这是鲁棒的VLR识别模块的理想特性。

此外，为了重申VLR识别模型的必要性，仅在HR图像上训练具有相同架构的CapsNet。在这种情况下，模型



(

a) VLR Input

(

b) Reconstructions

图4：从VLR输入在SVHN数据集上获得的样本重构。DirectCapsNet能够在可用信息内容有限的情况下重建数字（例如，绿色框），但是它也无法正确地重建一些具有挑战性的情况（例如红色框）。

在训练时看不到任何VLR图像，并且会在VLR测试图像上进行评估。可以观察到，CapsNet（HR）的分类精度达到77.82％，因此重申了在培训时需要开发专用的VLR识别网络或利用任务特定信息的需求。我们还进行了McNemar检验[ 15 ]，并在建议的DirectCapsNet和CapsNet之间的置信区间（CI）为99％（*p*值*<* 0.01）的情况下实现了统计差异。表1还可以进行分析，以了解所提出的DirectCapsNet模型的每个组件的效果。从DirectCapsNet模型中删除HR锚损失后，可以达到82.42％的top-1准确性，而去除目标重建损失后可以达到81.95％的top-1准确性。与建议的DirectCapsNet模型相比，这两种模型均显示出较差的性能，因此支持在DirectCapsNet模型中包括HR锚固损失，定向重建损失和胶囊。

案例研究1-VLR数字分类：表1 列出了建议的DirectCapsNet的SVHN数据集的top-1和top-5分类准确性，并与其他技术进行了比较。所提出的DirectCapsNet模型实现了top-1精度为84.51％和top-5精度为91.20％。DirectCapsNet展示了相对于健壮的部分耦合网络（RPC Nets）的最新结果[ 30 ]，在top-1方面的改进超过27％，这是一个基于CNN的框架，用于学习VLR的部分共享权重和HR样本，以及两者的部分独立权重。所提出的DirectCapsNet模型的出色性能激发了其在VLR识别中的用途。图4 展示了从DirectCapsNet获得的用于8×8 VLR样本。令人鼓舞的是，DirectCapsNet模型能够为输入样本重建数字，从而激励了目标重建损失的包含。对于相同类别的样本，获得了相似的重建结果，这说明了HR锚丢失对于增加特征之间的类别内相似度的有效性。案例研究2-VLR人脸识别：表2 列出了



1. VLR输入 （c）HR输入



1. 重建VLR输入 （d）重建HR输入

图5：从CMU Multi-PIE数据集上的拟议DirectCapsNet模型获得的样本重构。对于同一个类别，DirectCapsNet能够将VLR和HR样本投影到相似的目标上，从而提出了可靠的分辨率不变特征表示。

|  |  |  |
| --- | --- | --- |
| 算法 | 准确性 （％） | |
| 8×8 | 16×16 |

|  |  |  |
| --- | --- | --- |
| 原创+ COTS（2018）[ 25 ] | 0.0 | 0.0 |
| 双三次插值。+ COTS（2018）[ 25 ] | 0.1 | 1.1 |
| SHSR（Synthesis + COTS）（2018）[ 25 ] | 82.6 | 91.1 |

|  |  |  |
| --- | --- | --- |
| 拟议的DirectCapsNet | 94.5 | 97.4 |

表2：CMU Multi-PIE数据集上的VLR识别的等级1准确性（％）[ 6 ]。HR图像的分辨率为96×96 。

表3：用于VLR人脸识别（16×16 ）的UCCS数据集[ 24 ]的等级1准确性（％）。HR图像的分辨率为80×80 。

|  |  |
| --- | --- |
| 算法 | 累积 （％） |
| 稳健的部分耦合网（2016）[ 30 ] | 59.03 |
| 选择性知识蒸馏（2019）[ 4 ] | 67.25 |
| LMSoftmax for VLR（2019）[ 14 ] | 64.90 |
| 适用于VLR的L2Softmax（2019）[ 14 ] | 85.00 |
| VLR的中心损失（2019）[ 14 ] | 93.40 |
| 拟议的DirectCapsNet | 95.81 |

两种VLR人脸识别协议的等级1识别（或顶级1识别）准确性。所提出的DirectCapsNet模型对于8×8 和16×16 VLR图像实现了94.5％和97.4％的精度，而在受约束的CMU Multi-PIE数据集上的HR辅助图像为96×96 （表2 ）。对于8×8 分辨率的图像，DirectCapsNet与最新技术（通过分层稀疏表示法（SHSR）合成）[ 25 ] 相比，显示出近12％的改进。图5展示了样本VLR和HR人脸图像，以及从DirectCapsNet获得的重建图像。所提出的模型能够将属于同一主题的面部重构到相似的目标上，这表明类内相似度很高。VLR和HR样本都被重建为相似的图像，从而恢复了目标重建和HR锚定丢失的好处。

4

90.3

91.55

92.15

95.17

40

50

60

70

80

90

100

15

10

x

9

12

x

10

x

12

18

x

15

Accuracy (%)

Resolution

HR-LR (MDS)

Mudunuri and Biswas

Proposed DirectCapsNet

图6：建议的DirectCapsNet在具有姿态和照明变化的VLR人脸识别分辨率变化时的性能。HR分辨率固定为36×30 像素。与HR-LR（MDS）[ 3 ]和Mudunuri和Biswas [ 16 ]进行了比较。

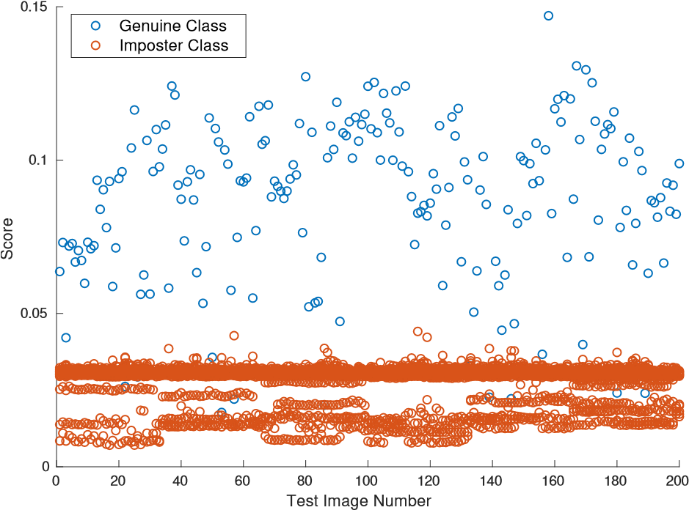


图7：提议的DirectCapsNet在UCCS数据集的某些样本上获得的用于VLR识别的分数。每个测试图像都有一个真实分数（正确等级）和179个冒名顶替者分数（错误等级）。

案例研究3-无约束的VLR人脸识别：表3 和图6给出了无约束的VLR人脸识别的等级1识别（或前1名识别）准确性。如表3 所示，在UCCS数据集上，DirectCapsNet模型实现了95.81％的1级准确度，表明与最新网络相比提高了近2.5％，与当前第二好的网络相比提高了近10％[ 14 ]。还与最近提出的大幅度softmax（LMSoftmax），*l* 2约束的softmax（L2Softmax）和基于中心损耗的VLR识别系统进行了比较[ 14]。与公制学习技术相比，拟议的DirectCapsNet的改进性能证明了合并辅助HR信息以在使用拟议的双向定向损失函数进行训练的同时提供方向的好处。图7 展示了DirectCaspNet模型在UCCS数据集样本上获得的分数。分数对应于用于分类的胶囊的活性矢量的长度。图7 表明，该模型能够为正确的类别生成高分，而为其他类别生成小分，这促进了可分离性，从而实现了较高的识别性能。

在具有姿势和照明变化的CMU MultiPIE数据集（图6 ）上获得了相似的性能，其中所提出的DirectCapsNet实现了95.17％的平均识别性能，与当前的最新算法相比，提高了约1.64％[ 16]。]。图6 证明了所提出的DirectCapsNet不会像其他降低分辨率的技术那样，在准确性方面遭受重大损失。该模型对15×12 和10×9 的识别准确率分别为92.15％和90.34％，而性能第二好的模型[ 16]表示两种分辨率之间的降幅接近9％。跨多个极低分辨率的识别性能的提高激发了所提出的DirectCapsNet模型在现实世界中的适用性。

# 结论

现有研究主要集中在高分辨率和低分辨率图像识别上。但是，VLR识别问题受到的关注有限。VLR识别是一个艰巨的问题，在现实世界中具有广泛的应用性，但它遭受的主要挑战是信息含量低。这项研究提出了一种用于VLR识别的新型双向定向胶囊网络（DirectCapsNet）。DirectCapsNet将分类的边际损失与建议的HR锚点损失和目标重建损失相结合，以增强VLR功能。HR图像在训练过程中用作“辅助”数据，以补充VLR功能学习。在VLR数字识别（SVHN数据库）和受约束/不受约束的VLR人脸识别（CMU Multi-PIE和UCCS数据库）上的实验结果证明了该模型的有效性，并提高了其在不同VLR任务中的可用性。将来，我们计划扩展提出的算法以解决多个协变量；例如，在人脸识别应用中，伪装下的VLR识别[26 ，老化[ 21 ]，频谱变化[ 19 ]和对抗攻击[ 5 ]。

# 致谢

这项研究得到了印度IIIT-德里的Infosys人工智能中心的部分支持。印度政府还通过Swarnajayanti奖学金来支持M. Vatsa。S. Nagpal通过TCS博士奖学金获得支持。

# 参考文献

1. Parnian Afshar，Arash Mohammadi和Konstantinos N Plataniotis。通过胶囊网络对脑肿瘤类型进行分类。在*IEEE国际图像处理会议上*，第3129–3133页，2018年3
2. Himanshu S. Bhatt，Richia Singh，Mayank Vatsa和Nalini K. Ratha。使用基于整体的共转移学习来改善跨分辨率的人脸匹配。*IEEE Transactions on Image Processing*，23（12）：5654–5669，2014. 2
3. Soma Biswas，Gaurav Aggarwal，Patrick J Flynn和Kevin W Bowyer。低分辨率人脸图像的姿态稳健识别。*模式分析与机器智能的IEEE交易*，35（12）：3037-3049，2013年8
4. 葛世明，赵胜伟，李晨雨和李嘉。通过选择性知识蒸馏技术在野外进行低分辨率人脸识别。*IEEE TRANSACTIONS ON图像处理*，28（4）：2051年至2062年，2019年2 ，3 ，6 ，7
5. Gaurav Goswami，Akshay Agarwal，Nalini Ratha，Richia Singh和Mayank Vatsa。检测并减轻对抗性干扰，以增强人脸识别能力。*国际计算机视觉杂志*，127（6）：719–742，2019. 8
6. 拉尔夫·格罗斯，伊恩·马修斯，杰弗里·科恩，武雄加纳德和西蒙·贝克。多PIE。*图像与视觉计算*，28（5）：807-813，2010年2 ，6 ， 7
7. Geoffrey E Hinton，Alex Krizhevsky和Sida D Wang。转换自动编码器。在*国际会议上人工神经网络*，44-51，2011页2 ， 3
8. 谢尔盖·艾菲和克里斯蒂安·塞格迪。批量归一化：通过减少内部协变量偏移来加速深度网络训练。*的arXiv预印本的arXiv：1502.03167*，2015年5
9. 卡齐·阿米努尔·伊斯兰教，丹尼尔·佩雷斯，维多利亚·希尔，布莱克·沙伊夫，理查德·齐默曼和姜莉。通过深囊网络在沿海水中检测海草。在*中国模式识别与计算机视觉会议上*，第320–331页，2018. 3
10. Ayush Jaiswal，Wael AbdAlmageed，Yue Wu和Premkumar Natarajan。CapsuleGAN：生成对抗性胶囊网络。在*欧洲计算机视觉会议上*，第526-535页，2018年3
11. 监卫健和林健民。基于奇异值分解的低分辨率人脸同时幻觉和识别。*IEEE视频技术电路和系统交易*，25（11）：1761-1772，2015年。2
12. Diederik P Kingma和Jimmy Ba。亚当：一种随机优化方法。*的arXiv预印本的arXiv：1412.6980*，2014年5
13. Christian Ledig，Lucas Theis，Ferenc Huszar，Jose Caballero，Andrew Cunningham，Alejandro Acosta，Andrew Aitken，

Alykhan Tejani，Johannes Totz，Zehan Wang和Wenzhe Shi。使用生成对抗网络实现逼真的单图像超分辨率。在*IEEE计算机视觉和模式识别会议上*，第105–114页，2017年。2

1. Pei Li，Loreto Prieto，Domingo Mery和Patrick J Flynn。在野外进行低分辨率人脸识别：比较和新技术。*IEEE Transactions informa-*

*灰取证和安全*，14（8）：2000-2012，2019年1 ，2 ，7 ，8

1. 奎因·麦克内玛。注意相关比例或百分比之间的差异的抽样误差。*Psychometrika*，12（2）：153-157，1947. 7
2. Sivaram Prasad Mudunuri和Soma Biswas。姿势和照明变化时的低分辨率人脸识别。*IEEE TRANSACTIONS ON模式分析和机器智能*，38（5）：一〇三四年至1040年，2016年2 ，6 ， 8
3. Yuval Netzer，Tao Wang，Adam Coates，Alessandro Bissacco，Bo Wu和Andrew Y. Ng。通过无监督功能学习在自然图像中读取数字。在*NIPS研讨会深度学习和无监督特征点学习*，2011年1 ， 2 ，4 ， 5 ，6
4. 欧阳书信，蒂莫西·霍斯佩达莱斯，宋义哲，李学明，陈易洛伊和王小刚。异类人脸识别的调查：素描，红外，3D和低分辨率。*图像与视觉计算*，56:28 – 48，2016. 2
5. 欧阳淑欣，蒂莫西·霍斯佩达莱斯，宋义哲，李学明，陈易·罗伊和王小刚。异类人脸识别的调查：素描，红外，3D和低分辨率。*图像与视觉计算*，56：28–48，2016. 8
6. 宋哲公园，闵圭公园和文基刚。超分辨率图像重建：技术概述。*IEEE信号处理杂志*，20（3）：21–36，2003. 2
7. Narayanan Ramanathan和Rama Chellappa。跨年龄进度的人脸验证。*IEEE Transactions on Image Processing*，15（11）：3349–3361，2006年。8
8. Sara Sabour，Nicholas Frosst和Geoffrey E Hinton。胶囊之间的动态路由。在*在神经信息处理系统的进步*，3856-3866，2017年页3 ，5 ，6
9. Mehdi SM Sajjadi，Bernhard Scholkopf和Michael Hirsch。Enhancenet：通过自动纹理合成实现单幅图像超分辨率。在*IEEE国际计算机视觉会议上*，第4501-4510页，2017年.2
10. Archana Sapkota和Terrance E Boult。大型无约束开放式面孔数据库。在*IEEE国际会议生物识别：理论，应用和系统*，1-8页，2013年1 ，2 ，6 ，7
11. Maneet Singh，Shruti Nagpal，Mayank Vatsa，Richia Singh和Angshul Majumdar。用于交叉分辨率面部识别的身份识别合成。在*IEEE会议计算机视觉与模式识别研讨会*，592- 59209页，2018年1 ，2 ，6 ，7
12. Maneet Singh，Richia Singh，Mayank Vatsa，Nalini K. Ratha和Rama Chellappa。识别野外伪装的面孔。*IEEE Transactions on Biometrics，Behavior，and Identity Science*，1（2）：97–108，2019. 8。
13. Pawan Sinha，Benjamin Balas，Yuri Ostrovsky和Richard Russell。人脸识别：所有计算机视觉研究人员都应该知道的19个结果。*IEEE论文集*，94（11）：1948–1962，2006. 4
14. 王楠楠，陶大成，高新波，李雪龙和李杰。面对幻觉的全面调查。*国际计算机视觉杂志*，106（1）：9-30，2014年。2
15. 王新涛，柯宇，赵东和陈易·罗伊。通过深度空间特征变换以图像超分辨率恢复逼真的纹理。在*IEEE计算机视觉和模式识别会议上*，第606-615页，2018年.2
16. 王章阳，张世玉，杨应珍，刘鼎和黄Thomas铭。使用深度网络研究非常低的分辨率识别。在*IEEE会议计算机视觉与模式识别*，4792-4800页，2016年2 ，6 ， 7
17. 王志飞，苗镇江，QM吴宗盛，万艳丽和唐振。低分辨率人脸识别：评论。*视觉计算机*，30（4）：359–386，2014. 2
18. 灿群，张璐，唐怡，邹文斌和陈旭。Ms-capsnet：一种新颖的多尺度胶囊网络。*IEEE信号处理快报*，25（12）：1850至1854年，2018年3
19. Wilman WW Zou和Pong C Yuen。低分辨率人脸识别问题。*IEEE TRANSACTIONS ON图像处理* 327-340，2012：21（1）1 ，2
20. 原文

提供更好的翻译建议