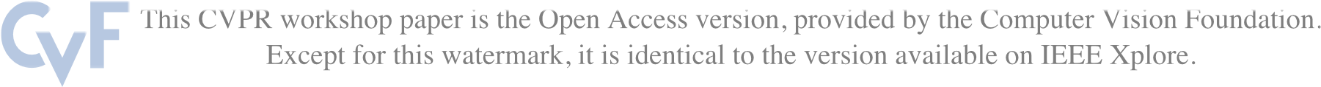
GenLR-Net：用于超低分辨率人脸和物体识别的深度框架

Sivaram Prasad Mudunuri [[1]](#footnote-0)[1] 1，Soubhik Sanyal \* 2 和Soma Biswas 1 1 电气工程系，印度科学研究院，印度班加罗尔。2 德国图宾根马克斯-普朗克智能系统研究所感知系统系。sivaramm@iisc.ac.in，soubhik.sanyal@tuebingen.mpg.de，somabiswas@iisc.ac.in[[2]](#footnote-1)

# 简介

从超低分辨率（LR）图像中识别人脸和物体在远程监视应用中很重要[33]。当从远处拍摄图像时，图像中的感兴趣区域通常非常小。因此通常在图像中缺乏区分信息，这有助于将一个物体与另一个物体区分开。但是，数据库中的图像通常具有高分辨率（HR），因此需要跨越分辨率的显着差异以及视点，照明等的变化来匹配图像。

同样由于对象类别的数量不断增加，假设在训练期间所有类别都可用[34]是不切实际的，这使任务更具挑战性。虽然LR面部图像的识别已经

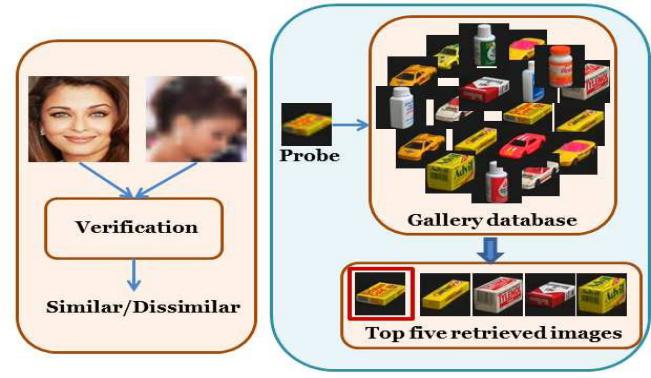


图1.低分辨率人脸验证[25]（左）和低分辨率对象识别[21]（右）的应用。

经过合理研究[39] [20]，但是识别普通物体的LR图像是相对未发展的。对于一般对象，正在对来自看不见类别的数据进行分类研究，这被称为零样本学习（ZSL）[34]。此处，提供了可见和不可见类别的属性，并且在测试期间，通过将不可见查询图像与可用属性进行匹配来对不可见查询图像进行分类。但是很多时候，我们可能希望检索相似的商品而不是对它们进行分类，例如，我们在街上给我们喜欢的衣服拍照，并希望从网上商店中检索相似的服装[10] [8] [ 18]。在我们的工作中，目标是当查询属于可见/不可见对象时检索相似的图像，因此我们不需要属性信息。因此本文中的这个问题与ZSL很不同。

在我们的工作中，使用相同的框架，我们解决了两个问题，即（1）LR面部识别 和（2）LR 对象识别。在第二个应用中，给定查询，检索相似项，并将与查询相似度最高的项作为正确匹配。

对于面部验证，训练和测试的subject是完全不相交的。对于一般对象，在测试过程中，我们首先评估当输入对象类为可见类别时评估所提出框架的性能，然后评估将其推广到看不见的类别的程度。为了进行训练，我们要求图像对属于相同和不同的类别，但是不需要类别标签。由于目标不是对象/人 分类任务，因此在训练过程中不需要对象类别或主题标识。

在这项工作中，我们提出了一种新颖的深度CNN架构，该架构可以处理分辨率上的巨大差异，并且可以推广到看不见的类别。我们在现有架构的基础上，通过合并不同阶段的不同损失以及适当的分析和评估，来构建自己的架构。我们使用VGG人脸网络[23]进行人脸验证实验，并使用VGG-object网络[27]进行对象检索任务，尽管也可以无缝使用任何其他相关架构。首先，我们分析当HR图像替换为LR图像时，如何影响标准体系结构的性能。为了补偿由于极低分辨率而导致的性能急剧下降，我们引入了三种不同的损失，即高级特征的对比损失，中间层特征的类间-类内分类损失和低层特征的超分辨率损失。我们分析了每种损失提升性能的原因和方式。在修改过的LFW [7]和用于人脸识别任务的CFPin wild数据库中进行了广泛的实验，以及用于对象检索任务的COIL-100 [21]和Toy Cars [22]数据库。对于看不见的物品类别的实验表明了所提出网络的泛化能力。提出工作的贡献如下：

* 分析极低分辨率对面部验证和物体识别应用的影响。
* 一种新颖的深度架构，可以处理分辨率上的巨大差异，并且可以泛化到看不见的类别。
* 在不同数据集上进行的大量实验表明，所提出的网络可以很好地匹配LR图像，并且具有泛化能力。

本文的其余部分安排如下。第2节给出了相关工作的指针。提议的方法和实验结果分别在第3节和第4节中介绍。最后，本文得出结论。

# 相关工作

在本节中，我们提供一些与低分辨率人脸和物体识别有关的作品的指针。低分辨率人脸识别：直到最近，研究人员才开始研究低分辨率问题。Wang *等。*[32]通过系统的基于深度学习的体系结构演示了在非常低分辨率的情况下的识别问题。Zeng *等。*[37]提出了一个分辨率不变的深度网络来学习整个领域的分辨率不变的特征。杨*等。*[35]通过添加额外的类间约束，提出**了一种判别性多维缩放方法，该约束扩大了学习空间中不同主题之间的距离。Farrugia *等。*[4]提出了一种使用耦合稀疏支持框架的线性模型的幻觉技术，该技术基于高分辨率流形的局部几何结构构造线性模型。朱*等。*[38]提出了一种深层的双网络体系结构，该体系结构解决了针对低分辨率人脸图像的幻觉和密集对应场估计问题。Mudunuri *等。*[19]提出了一种在基准点上基于MDS的自动低分辨率人脸识别方法。这些方法旨在从LR和HR图像中学习投影映射，从而使相同的对象在公共空间中显示相似的特征。还开发了基于合成的方法来从给定的LR图像重建HR图像。[26]中描述了一种基于综合的方法，该方法可学习类和解决方案的特定词典。Kolouri和Rohde [13]提出了一种通过学习HR图像上的非线性拉格朗日模型从给定的LR图像合成相应的HR脸部图像的方法。邹*等。*[39]提出了一种基于关系学习的超分辨率方法，它通过在HR和LR图像空间之间施加判别约束。作者演示了7 × 6 LR图像分析超低分辨率问题的性能。在[36]中设计了一种变换性和区分性自动编码器模型，以使未对齐且有噪声的LR面产生幻觉。在[17]中开发了一种更通用的相似性度量并将其统一到深度神经网络体系结构中以进行判别性特征表示学习。低分辨率对象识别的探索相对较少。Peng *等。*[24]提出了一种知识转移框架，用于区分LR图像中的细粒度对象。蔡*等。*[2]提出了一种端到端的分辨率感知CNN架构，通过对超分辨率和分类建模来对LR对象进行分类。在[31]中对光学模糊下不同应用的性能影响进行了详细研究。Su和Maji [28]通过解决包括图像降级在内的跨域变化，对交叉质量模型的适应进行了建模。大多数对象识别算法都解决了对LR图像进行分类的问题。相反，在这项工作中，我们解决了为给定的LR探针检索相似的HR图像的任务。**

# **提出的方法**

**在这里，我们以人脸验证为例，详细介绍了所提出的架构。我们逐步在网络的不同部分构建具有不同损失的体系结构，并对其进行分析和评估。**

## **问题陈述**

**在这项工作中，我们解决了两个问题，即在低分辨率条件下的面部验证和物体识别。**在后面的部分中，我们将**以人脸验证为例来说明所提出的网络体系结构**。在训练过程中，我们假设我们可以访问训练对象的LR和HR图像，训练数据以以下形式给出：



其中*N*是训练图像的总数。这里*l i*是二进制标签，如果图像属于同一主题（相似对），则为1 ；否则（异类对）为0 。**不使用subject编号，因为目标不是分类，而是验证**，并且测试样例与训练样例完全不同。尽管**x** *h i*和**x** *l i* 属于同一类，但它们在分辨率，视点，照明等方面有所不同（图2和3）。在测试过程中，给定一个图像对，其中一个是LR，另一个是HR，目的是验证它们是否属于同一subject。

## **动机**

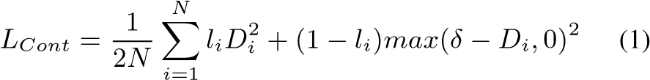
**首先，我们分析了LR输入图像对标准人脸识别网络的验证性能的影响。对于这项工作，我们采用了VGG人脸网络[23]作为基准架构，**尽管所提出的方法是通用的，并且也可以应用于其他基础网络。对于本实验，我们使用LFW数据库[7]和标准实验协议[7]。我们使用LFW-深漏斗图像在数据库的1fold上进行实验。由于VGG人脸网络是在HR图像上进行训练的，因此我们首先通过提供两种视图的HR图像来评估性能，并且我们获得93.83 ％的验证率。在我们所有的实验中，我们将相应的*fc* 7层（图2）的激活作为给定图像的特征。为了分析低分辨率的影响，我们保留了视图1的原始分辨率（224 × 224）的图像，并将视图2的图像缩减为20 × 20。然后使用双线性插值将这些图像的大小调整为224 × 224（图3）。以便可以计算和比较来自同一网络的特征。使用这些视图2图像，面部验证性能急剧下降至69.16 ％，这表明分辨率对于此应用非常重要。现在，我们描述在VGG框架之上构建的，被称为GenLR-Net的提议的深层体系结构，因为它适用于LR图像，并且可以泛化到看不见的类别。

## 提出的GenLR-Net

VGGFace结构[23]由图2中的阴影部分显示。该体系结构用于对面部图像进行分类，具有16个可训练层（每个Conv块里面有多个层），包括卷积层和全连接层。在训练期间采用softmax层之后的平均预测对数损失，以最大程度地减少分类误差。该网络接受了2.6M个身份的2.6M图像训练。

**在这项工作中，我们的目标是验证一对LR和HR图像是否属于同一subject。因此，我们希望属于同一主题的HR和LR图像彼此靠近，而来自不同主题的HR和LR图像彼此分开。**由于我们有两个分辨率不同的图像，因此提出的网络（图2）具有两个流，其中一个通道将输入作为HR图像，另一个通道将输入作为LR图像。HR通道保持参数固定（阴影部分表示在训练过程中权重已锁定且未更新），因为已经对网络进行了训练，可以从HR面部图像中计算出判别特征。我们的第一个修改是，我们将VGG网络的最终分类层替换为更高层次特征的对比层，即最终的完全连接层（*fc* 7 ）。

**较高级别特征的对比损失**：首先，对比损失[30] [6]应用于网络的两个通道的*fc* 7 层之间。这将有助于使HR和LR阳性样本更接近，并使阴性样本分开。任意两个特征之间的对比损失计算为



在此，*D i*是对应特征之间的*L* 2距离，而*δ*是用于分离不同对象的特征的余量（margin）。N是训练样本的总数，*l i*是每个训练对的相应二进制标签。如已经提到的，由于对应于HR图像的通道权重是固定的而不是更新的，因此loss将仅通过获取LR输入数据的通道传播。通过此修改，我们获得了84.00 ％的验证率。

在最近的一些文献[24] [16]中，已经观察到，加强中间层之间的损失可以提高性能。我们研究了在前面的层之间（如*fc* 6 ，pool 5 和pool 4 ）引入对比损失的影响，等等。我们观察到，当将这种损失加入*fc* 6 层时，结果将提高到86.00 ％ 。但是，当我们在前几层中施加相同的损失时，性能开始下降。

|  |
| --- |
| 图2.提议的GenLR-Net框架的简化框图。两通道CNN在一个通道中拍摄HR图像，在另一通道中拍摄LR图像。阴影块表示权重已初始化并固定为基准网络（在此示例中为VGG人脸网络）的权重。为了便于可视化，未显示所有层和非线性激活等。 |

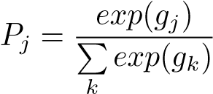
**这可能是因为该损失不适用于中级或低级特征。这就引出一个问题，*什么样的监督适合于初始层以提高性能？***

**中层特征的类间-类内分类损失**：基于先前的分析，我们引入了一个比对比损失*更“soft”*的约束，因此可能更适合中层特征。受[16] 启发，我们为中层特征提出了**类间-类内**分类损失。在此，计算两个图像（HR和LR）之间的差异，如果它们属于同一类别（对象），则分类为1；如果它们属于不同类别，则分类为0。因此，将N类问题转换为2类问题。这种损失还试图将来自同一类别的样本拉近，并将来自不同类别的样本推开。但是它比对比性损失少受约束，因为它并没有在相同类别和不同类别之间增加一个严格的margin，因此我们认为它适合。我们将两个通道的*池pool*5 的之间的activation差异考虑在内，并对这些特征应用分类损失，通过此修改，验证性能提高到87.24 ％。通过在pool 4 的向量上也加上该损失，我们观察到性能进一步提高到89 ％。与对比性损失一样，如果在pool 3 ，*pool*2 等初始低级特征中增加该损失，则性能开始下降。因此，在最终网络中，我们仅将此损失应用于pool 5 和pool 4 层。

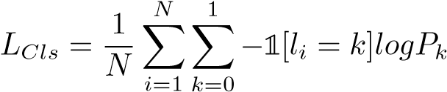
令*f* 1和*f* 2分别是LR和HR通道的*pool*5 层的激活值。在人脸验证实验中，*˚F* 1和*˚F* 2 ∈ *R* 7×7×512，因为在*CONV* 5 层有512个卷积核。特征差表示为（*f* 1 - *f* 2），结果转换为列向量（用*f*表示）。此向量连接到具有两个节点的softmax层，使得*f*如果输入图像对属于同一类别，则分类为1，否则分类为0。*θ ˚F*是*˚F*到SOFTMAX层的连接权重，激活值可以被计算为

 （2）

在此，*φ*是非线性函数*reLU*。softmax概率的计算如下

 （3）

交叉熵分类损失[29]用于softmax概率以对输入对进行分类

 （4）

如果输入对属于类别*k*，则1 [ *l i* = *k* ] = 1 ，否则为零。

**低级特征的超分辨率损失**：将LR图像与HR图像进行匹配的一种方法是，首先对LR图像应用超分辨率以使其成为HR，然后执行匹配。尽管超分辨率（SR）方法对于增强图像分辨率非常有用，但在识别应用中它们的设计效果不佳[1] [39]，因为很多SR网络不是专门为识别任务所设计的。我们还观察到，单独在LR图像上应用SR算法，并将SR图像用于识别不会导致识别性能的显着提高。因此，受[2]的启发，我们将超分辨率目标与验证任务结合。由于这可以使得网络的权重可以同时提高分辨率和验证性能，因此这应该会提高性能[2]。仅当与LR图像相对应的HR图像可用时，我们才可以使用此损失。这与可用于训练的HR图像对不同，因为HR图像对除了分辨率差异在姿势，照明，表情等方面也具有差异，并且可能使网络混乱。为此，我们将*conv1*层的输出，并将其作为*srconv1*层的输入。*srconv1*层预测残差图像，因此该输出与对应的LR图像相加，得到最终的超分辨率输出[12]。接下来，计算超分辨率输出和原始HR图像之间的重建损失。令**s** *l i* 为*srconv1*层的输出，则重建损失如下：

 （5）

在此，（**x** *l i*）*hr* 是相应**x** *l i*的HR图像。在这里，我们假设LR图像的HR版本可用于训练，这通常是超分辨率任务的情况。这样，我们观察到验证率提高到90.00 ％。

总的来说，我们通过联合减少所有损失来训练整个网络。具体而言，在*conv1* 层之后存在超分辨率损失，在*pool4*和*pool5*层分别存在两个分类损失，在*fc* 6 和*fc* 7 层处分别存在两个对比损失。最终损失可以表述为：

*大号* = *λ* 1 *大号SR* + *λ* 2 *大号池CLS* 4 + *λ* 3 *大号池CLS* 5 +

6 *FC* 7 （6）*λ* 4 *大号 fcCon* + *λ* 5 *大号精读*

表1给出了上述逐步应用的损失结果的摘要。

表1.说明了所提出方法的逐步动机。

|  |  |
| --- | --- |
| 方法 | 验证率（％） |
| *大号fcCon* 7 | 84.00 |
| *L fcCon* 6 + *L fcCon* 7 | 86.00 |
| *L poolCls* 5 + *L fcCont* 6 + *L fcCont* 7 | 87.24 |
| *L poolCls* 4 + *L poolCls* 5 + *L fcCon* 6 + *L fcCon* 7 | 89.00 |
| *L SR* + *L poolCls* 4 + *L poolCls* 5 + *L fcCon* 6 + *L fcCon* 7 | 90.00 |

在我们的实验中，我们设置*λ* 1 = 10 -3和*λ m*= 1 ，m = 2,3,4,5。我们使用Caffe [9]深度学习框架优化了建议的深度网络。对于具有五个损失的最终网络，我们使用预先训练的VGGFace（在面部验证实验的情况下）权重初始化网络，并以10 -8的学习率和0.0005的权重衰减对体系结构进行训练。动量固定为0.9。每20000次迭代后，我们将学习率逐步降低10倍，并且网络进行了80000次迭代。超分辨率层*srconv1*是kernel大小为3的3通道层（RGB）*。srconv1*的权重从*xavier*分布随机初始化，标准偏差为0.01。对于对象识别的实验，我们对VGG对象网络使用相同的参数，但是每20000次迭代后，学习率将降低10倍，网络学习60000次迭代。

# 实验结果

在这里，我们详细描述了为评估所提出的网络性能而进行的实验。具体来说，我们要解决以下问题：

* 提出的框架在各种分辨率的较大差异之间进行匹配/验证的效果如何。
* 提出的方法与最新的超分辨率方法相比如何？
* 提议的方法如何推广到看不见类别的图像？

在这项工作中，我们着重于两种应用：（1）交叉分辨率人脸验证，其中一个图像是LR，另一幅图像是HR，包括姿势，照明等方面的变化，（2）交叉分辨率对象识别，其中probe是LR，数据库中的图像是HR。

## 跨分辨率人脸验证

对于跨分辨率人脸验证的应用，我们在LFW数据库的修改版本[7]和CFP in wild上评估我们的方法。

（A）**在LFW数据库[7]上进行的实验**：LFW数据库[7]标记了在真正不受限制的环境下捕获的面部图像。面部图像具有广泛的姿势，表情，种族，服装，发型，灯光和其他参数，因此是用于面部验证应用的最广泛使用的数据库之一。因此，我们为我们的应用选择了对该数据库的修改以及标准实验协议。但是，与标准设置相比，我们的应用更具挑战性，因为在我们的情况下，其中一张图像的分辨率非常低。

我们在数据库的fold 1并使用LFW-深漏斗图像进行了实验。我们根据标准协议[7]在2700个相似对和2700个不相似对上训练了我们的网络，并在300个相似和300个不相似对上进行了测试，但具有修改的分辨率，如**3.2**节中所述。LR图像是通过对将原始图像下采样到20 × 20，然后使用双线性插值将其上采样到原始分辨率。根据我们的协议，部分示例图像显示在图3中。

表2. 修改后的LFW数据库上建议的方法的性能（％）[7]。还提供了超分辨率，度量学习和具有深度特征的领域自适应技术的比较。

|  |  |
| --- | --- |
| 方法 | 验证率（％） |

HR-HR (original VGG)

93.83

HR-LR (original VGG)

69.16

SSR [11] +

*fc*

7

features

72.10

SRCNN [3] +

*fc*

7

features

73.16

LapSRN [15] +

*fc*

7

features

76.16

*fc*

7

features + SA [5]

72.50

*fc*

7

features + LSML [14]

71.00

Proposed GenLR-Net

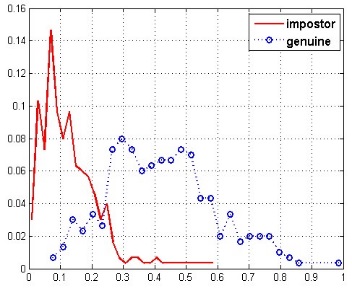
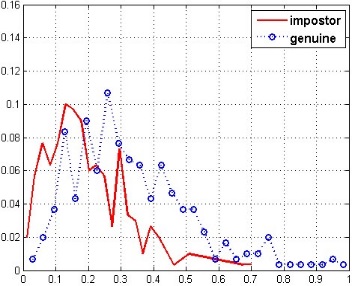
90.00



图3.我们实验中使用的修改后的LFW数据库[7]（左）和野生数据库[25]（右）中的CFP的面部图像样本。（a，c）和（b，d）的每一列显示了各自数据集的相似和不相似对的示例。根据我们的协议，视图1的图像是HR图像，视图2的图像是LR图像。

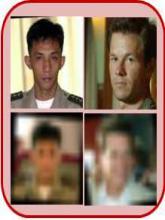
结果报告在表2中。正如动机部分中已经提到的那样，当两个HR图像都提供给原始VGG脸部网络时，我们的验证率为93.83 ％。根据我们的协议，当将视图2图像更改为LR时，性能下降到69.16 ％，这表明图像分辨率对人脸验证任务的重要性。该网络能够将验证性能提高到90.00 ％。

改善LR图像分辨率的标准技术之一是使用超分辨率（SR）技术。最近，该领域的研究取得了显着进展，并且提出了几种SR技术，这些技术可提供令人印象深刻的输出。在这里，我们评估了SR技术应用于LR图像增强其分辨率时的验证性能，然后使用原始的VGG网络提取特征。（先用SR处理图像，再用最原始的VGG验证）。我们评估了三种最新的SR技术[11] [15] [3]，以了解超分辨率对验证性能的影响。深度拉普拉斯金字塔超高分辨率网络（LapSRN）[15]是一种深度体系结构，可从HR图像系统地重建残留内容。SRCNN [3]是基于传统的基于稀疏编码的SR技术制定的轻型深度体系结构。



（a）VGG人脸网络 （b）建议的方法

图4.使用VGG网和建议的GenLR-Net的HR-LR图像的相似（称为真实）和不相似（称为冒名顶替）分数分布。使用GenLR-Net可以更好地分离分布。对于每个图，X和Y轴分别表示相似性得分和测试对分数。



的（a）类似的（X ） （b）中类似的（× ） （c）中异种 （d）异种

（X ） （× ）

图5.修改后的LFW数据库上建议的GenLR-Net的验证结果[7]。相似（不相似）表示属于相同（不同）主题的列。复选标记（X ）和（× ）表示网络分别正确和错误地分类了输入对。即，相似（X ）表示输入对（按列）相似，并且GenLR-Net对其进行正确分类。

学习LR和HR图像之间映射的编码方法。从表2中我们可以看到，SR方法可以成功地改善性能，但是提出的GenLR-Net的性能要好得多。与标准度量学习[14]和应用于*fc* 7特征的域自适应[5]技术相比，该方法的性能也明显更好。

我们使用相似和不相似对的距离分布进一步分析所建议网络的有效性。我们从图4观察到，使用建议的GenLR-Net，分布可以更好的相互分离，这对于获得良好性能至关重要。如表2所示，在具有挑战性的低分辨率设置下，与基本网络相比，这可以改善提议网络的面部验证性能。图5显示了很少的图像对，它们被提议的GenLR-Net正确和错误地分类。

**分辨率变化的影响**：在这里，我们研究在2个视图的不同分辨率下该方法的性能。对于这项研究，我们

表3. 使用建议的GenLRNet对LR图像的不同分辨率的验证性能（％）。

|  |  |  |  |
| --- | --- | --- | --- |
| 楷模 | 20 × 20 | 10 × 10 | 5 × 5 |
| 原始的VGG脸 | 69.16 | 56.17 | 54.50 |
| 拟议的GenLR-Net（微调20 × 20图像） | 90.00 | 67.70 | 62.30 |
| 拟议的GenLR-Net（用10 × 10图像微调） | -- | 72.17 | 65.00 |
| 拟议的GenLR-Net（微调5 × 5图像） | -- | -- | 65.80 |

实验使用大小为10 × 10和5 × 5的LR图像。将原始图像下采样到所需的大小，然后使用双线性插值上采样到相同的分辨率。对于这些甚至更低分辨率的图像，受[24]启发，以分阶段的方式进行训练。

例如，对于处理10 × 10 大小的LR图像，首先在20 × 20 LR图像上训练网络，然后针对10 × 10 LR图像进行微调，而不是直接针对较低的分辨率进行训练。这种方法有助于网络系统地学习变化[24]。结果记录在表3中。第二行表明，如果我们直接采用经过训练，以更低的分辨率对20 × 20的分辨率进行训练的网络，性能将开始下降。对于大小为10 × 10和5 × 5的图像，我们分别获得67.7 ％和62.3 ％的验证率。但是如果我们对网络进行微调以获取尺寸为10的图像× 10，性能提高到72.17 ％。从表中我们还可以看到，提出的GenLR-Net的性能随着分辨率的降低而逐渐降低，即使分辨率很差，其性能也相当不错。

（B）**在CFP in wild数据库中进行的实验**[25]：在这里，我们评估了在低分辨率设置下从正面到轮廓脸部验证的建议框架。图3（右）显示了少量根据我们的协议提供的数据库示例图像。与先前的情况相比，此方案更具挑战性，因为在此数据库中，前额（视图1）和轮廓（视图2）之间的姿势和表情存在很大差异。该数据集呈现了一个非常具有挑战性和现实性的场景，涉及跨分辨率和跨姿势匹配。与LFW相似，数据库有10个拆分，每个拆分有350个匹配对和350个不匹配对。同样在此实验中，probe人脸也被下采样为20 ×20并使用双线性插值上采样至原始分辨率。我们在one fold评估提议的方法，在该过程中我们在9个分组上训练我们的网络，并在其余分组上对其进行测试。我们遵循数据库从正面到轮廓的匹配协议（第1组），结果显示在表4中。我们观察到，与使用原始VGG *fc* 7 功能的基线特征71.71 ％相比，结果要好得多。

表4. 在野生数据库中对CFP提出的方法的验证性能（％）[25]。

|  |  |
| --- | --- |
| 方法 | 验证率（％） |
| HR-HR（原始VGG） | 88.57 |
| HR-LR（原始VGG） | 71.71 |
| 拟议的GenLR-Net | 77.28 |

## 跨分辨率物品识别

我们在COIL 100 [21]和ToyCar数据库[22]上进行了第二次应用实验，并评估了该方法的泛化能力。对于与对象识别相关的所有实验，我们使用VGG object网络权重[27]初始化了我们的网络。

（A）**在COIL-100数据库上进行的实验**[21]：解决对象分类任务的大多数方法都假设用于训练和测试的对象类别是一样的。即，在一组N个类别上训练模型，并且在测试期间，需要对属于这N个类别之一的查询图像进行分类。但是，由于不断发现新的对象类别，这是一个限制性假设。在许多现实情况下，查询图像可能来自模型在训练阶段未看到的一类。

例如，在在线购物中，用户想要搜索与他/她捕获的图片相似的连衣裙。首先，在网上商店中显示的图像通常是专业拍摄的，背景更干净，光线很好，而消费者拍摄的照片可以使用低分辨率相机拍摄。同样，新的服装项目会不断添加，因此不太可能对数据库中所有可能的服装项目进行模型训练。因此，最终目标是将不受控制的，看不见的查询与相对受控制的数据项进行匹配，并检索相似的数据项。

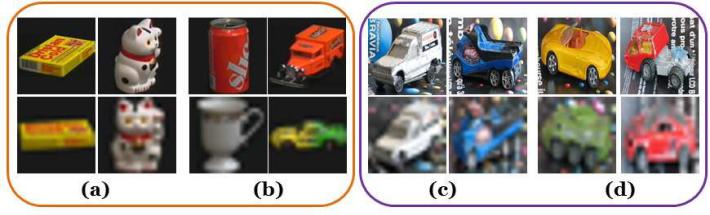


图6.实验中使用的COIL-100数据库[21]（左）和Toy Cars数据库[22]（右）的样本对象图像。（a，c）和（b，d）的每一列显示了用于训练的各个数据集的相似和不相似对的示例。

我们在COIL-100数据库上评估了所建议的GenLRNet的泛化性[21]（左图6）。该数据库有100个类别，每个类别包含72个姿势不同的图像，共7200个图像。每个图像的尺寸为128×128 。数据库中的对象具有



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表5. 在不同协议下，COIL-100数据库[21]的等级-1准确性（％）。   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 方法 | 看过看过 | 看不见的 | 全部看到 | 全部看不见 | | HR-HR（原始VGG对象） | 97.57 | 99.66 | 97.57 | 88.66 | | HR-LR（原始VGG对象） | 68.99 | 90.67 | 67.17 | 69.67 | | LR数据上的微调VGG对象 | 78.28 | 92.33 | 77.17 | 75.67 | | LapSRN [15] + *FC* 7 功能 | 88.18 | 94.00 | 87.47 | 76.67 | | 拟议的GenLR-Net | 93.13 | 98.00 | 91.21 | 81.00 | |

图7. GenLR-Net在COIL-100上的交叉分辨率对象检索结果[21]。每行显示与LR查询（第一列）相对应的前五个检索结果（第2-6列）。前两行在所有协议中都是可见的，后两行在所有协议中都是看不见的。正确的匹配由红色框表示。

各种各样复杂的几何和反射特性。我们将数据集分为两组，其中选择90个类别作为可见类别，其余10个视为不可见类别。我们从这90个对象中的每一个中随机选择60张（共72张）图像，以生成匹配和不匹配的对，以进行训练和验证。在每对图像中，一个图像具有高分辨率，第二个图像具有低分辨率。通过将原始图像下采样至20×20 ，然后将其上采样至224×224 （VGG对象网络输入所需的分辨率）来获得LR图像。来自90个类别的每类的其余12张图像用于测试。经过培训后，测试阶段分为四个不同的协议。

**Seen in Seen**：在每个90个训练类别的12个测试图像中，有一个图像被保留为gallery图像，其余11个图像被用作probe图像。在所有四种协议中，gallery是HR图像，probe是LR图像。

**Unseen in Unseen**：这里我们使用剩下的10个类别，这些类别没有用于训练。从10个类别的每一个中随机选择一张图像用作HR gallery。我们从这些类别的其余图像中随机选择30张图像，并将其用作LR probe。**Seen in All**：在这种情况下，我们将probe图像保留为“Seen in Seen”协议的设置。gallery分为可见的和不可见的两类，由“Seen in seen ”和“unseen in unseen”的gallery组成。

**Unseen in all**：我们将gallery与“seen in all”的协议保持一致，而probe 则与“seen in seen”相同。最后两个协议更为现实，因为我们通常不知道查询图像属于可见还是不可见类，因此通常不会先验。因此，必须将查询与可见和不可见类别进行比较。

表5列出了GenLR-Net在所有协议上的性能。我们观察到，看不见类别的结果比看过类别的结果要高。这是因为对于看不见的类别，样本数量要少得多。但是，我们看到，在所有情况下，与基本网络相比，与SR方法相比，提出的框架的性能都明显更好。图7中的每一行显示了使用所提出框架的给定LR探针图像（第一列）的前五个检索结果（第2-6列）。红色框表示正确的匹配项。

（B）**在Toy Cars数据库上进行的实验[**22]：在这里，我们像在面部验证应用程序中一样，探索了在不受控制的设置中进行对象验证的问题。该数据库包含14种不同的玩具卡车和小汽车的图像。训练集包含7个对象实例，并具有1185个相似和7330个不相似的图像对。其余7个对象实例用于测试。数据集中的图像具有多种姿势和光照变化（右图6）。任务是验证图像对是否属于同一对象。表6中报告了所提出方法的性能。我们观察到，即使在具有挑战性的LR环境中，GenLRNet仍能成功地改善基础网络的性能。

表6. 玩具汽车数据库中建议的方法的验证性能（％）[22]。

|  |  |
| --- | --- |
| 方法 | 验证率（％） |
| HR-HR（原始VGG） | 94.54 |
| HR-LR（原始VGG） | 86.15 |
| 拟议的GenLR-Net | 88.09 |

# 结论

在这项工作中，我们提出了一种新颖的深度学习框架，以解决将低分辨率探针图像（面部/对象）与数据库中高分辨率图像进行匹配的难题。我们还解决了看不见的物体识别这一极具挑战性和实用性的问题，这是一个相对未开发的领域。我们相信，这项研究将为研究人员解决这一具有挑战性的问题提供动力，同时也将作为基准，以解决更为传统的可见物体分类任务。

1. 原文

提供更好的翻译建议

1. 平等贡献。2这项工作是在作者还是班加罗尔IISc的研究生时完成的。 [↑](#footnote-ref-0)
2. [↑](#footnote-ref-1)