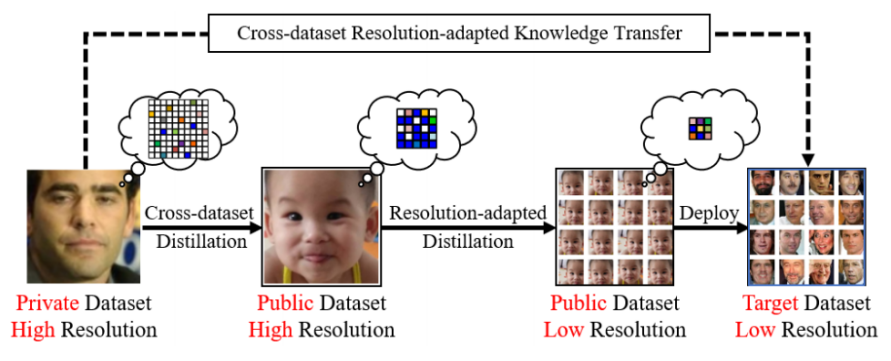
通过桥蒸馏进行的有效的低分辨率人脸识别

**Abstract**

野外人脸识别现在正朝着轻量级模型，快速推理速度和分辨率适应能力发展。在本文中，我们提出了一种桥蒸馏方法，将预先在高分辨率人脸上训练的复杂人脸模型转换为用于低分辨率人脸识别的轻量级模型。在我们的方法中，通过**两步蒸馏**解决了这种跨数据集分辨率的知识转移问题。第一步，我们进行跨数据集蒸馏，以将先验知识从私人高分辨率人脸转移到公共高分辨率人脸，并生成紧凑而有判别性的特征。在第二步中，进行分辨率适应的蒸馏，以通过多任务学习将先验知识进一步转移到合成的低分辨率人脸。通过学习低分辨率人脸表示并模仿适应的高分辨率知识，可以在识别低分辨率人脸时以高效率和有希望的准确性构建轻量级学生模型。实验结果表明，该学生模型在仅使用0.21M参数和0.057MB内存的低分辨率人脸识别中表现出色。同时，在GPU，CPU和移动电话上，其速度分别达到每秒14705、934和763个人脸。

**Introduction**

总之，将知识从高分辨率模型转移到低分辨率模型是有帮助的，并且可以避免计算量大的重构。因此，帮助识别低分辨率人脸的学习框架应该能够以有原则方式有效地传递信息丰富的高分辨率知识。也就是说，**它实际上需要解决两个子问题：应该从高分辨率模型中传递什么知识以及如何进行这种传递。**这样，通过单个框架即可同时解决低分辨率人脸识别和知识蒸馏方面的挑战。



（桥蒸馏的动机：

从私人高分辨率人脸到目标低分辨率人脸的直接知识转移可能很困难。因此，我们使用公共的高分辨率和低分辨率面孔作为桥梁，通过交叉数据集蒸馏和适应分辨率的蒸馏逐步地蒸馏和压缩知识。请注意，公共低分辨率人脸是从公共高分辨率人脸生成的，用于模拟目标低分辨率人脸的可能分布。

）

受此启发，如图1所示，我们提出了一种新颖的桥蒸馏方法，该方法可以将在其私有数据集上预先训练的现有高分辨率模型转换为用于目标数据集上的低分辨率人脸识别的简单模型。在我们的方法中，公开的高分辨率人脸及其分辨率降低的版本**被用作桥梁**通过两步蒸馏将复杂的高分辨率教师模型压缩为更简单的低分辨率学生模型。第一次跨数据集蒸馏使从私人到公共高分辨率人脸的预训练知识得以适应。它学习了一种特征映射，**该映射保留了对公共高分辨率人脸的判别能力以及原始私人知识中编码的详细人脸模式。**然后，第二步分辨率适应的蒸馏以多任务方式学习学生模型，以共同模仿适应的高分辨率知识并识别公共的低分辨率人脸，这些人脸被合成以模拟目标低分辨率的可能分布面孔。这样，学生模型只需要知道在低分辨率面孔上仍然可以区分的高分辨率细节，而与其他面孔无关，从而可以进行紧凑的知识转移。

所做的贡献概括如下：

1）我们提出了一种桥梁蒸馏框架，该框架能够将高分辨率的人脸模型转换为更简单的低分辨率人脸模型，同时大大降低了计算和存储成本，并且性能下降最小。 2）我们提出了跨数据集蒸馏技术，该技术将预先学习的知识从私人人脸数据集那里适应为公开的高分辨率人脸，从而保留了紧凑而有判别性的高分辨率细节； 3）进行了全面的实验，结果表明，学生模型与最新的高分辨率人脸模型具有可比的准确性，但是存储成本极低，推理速度也很快。

**THE PROPOSED APPROACH**

该框架基本上遵循师生框架，但具有两步蒸馏。首先对复杂的老师模型进行高分辨率人脸识别任务的训练，然后针对低分辨率场景将其蒸馏为更简单的学生模型。该框架引用以下描述符。

1. **域**：框架涉及三个域：1）私有域是指用于培训教师模型的外部高分辨率私有人脸数据集IP，通常较大且对于框架而言是不可见的； 2）公共域是公开的人脸数据集IS，并通过适应和模仿老师的知识，作为学习简单学生模型的桥梁，3）目标域是指用于识别看不见的低分辨率人脸IT（例如监视人脸）的部署场景。 在不失一般性的前提下，我们假设目标域中的人脸分布是可观察的，但标签不可用。
2. **教师模型**：在提出的框架中，教师模型被假定为现成的，这意味着它已经在IP上进行了预训练。通常，预先训练的教师模型会编码高分辨率人脸识别的丰富和泛化的知识。**因此，当将其应用于新的数据集（例如，IT）时，希望跨数据集迁移预先学习的高分辨率知识，而不是从头开始训练模型。**请注意，它与许多师生框架不同，后者的训练仅限于私有数据集或相同分辨率或可用的目标数据集。在这项工作中，我们假设教师模型Mt的形式为Mt =（Ft，St），由特征提取后端Ft和softmax分类层St组成。这是一种广泛应用的架构，大多数人脸分类模型都符合。因此，给定输入的面部图像I，首先使用ft（I）= Ft（I, wft）来计算高级特征；通过st（I）= St（ft（I），wst）处理，以获得分类分数。**Ft的设计和IT可能对所提出的方法是看不见的。**

**3）学生模型**：学生模型的设计比老师模型的设计简单得多。通过在**公共数据集IS上训练**学生模型来模仿老师的行为，该模型的复杂性大大降低了。在实际环境中，我们的目标是获得一个在低分辨率目标数据集IT上运行良好的优化学生模型Ms。通常，**IS，IT和IP不共享身份**。私有数据集和目标数据集之间的分辨率差异也存在。因此，问题在于如何正确地传递从IP中学到的高分辨率知识，以通过使用IS促进在IT上的低分辨率识别。 这种转移需要同时解决分布和分辨率差异。

**4）公式化**：在上述设置的基础上，可以将低识别度的人脸识别公式化为跨分辨率背景下的开放集域适应问题，其目的是实现从判别性高分辨率教师模型Mt（I; wt）到一个较简单的低分辨率学生模型Ms （I‘; ws）的有效知识转移，**私有数据集Ip和目标数据集It之间存在巨大的域偏移**。在此，两个数据集通常不共享身份，从而导致**特征分布的差异**。此外，额外的分辨率差异进一步增加了域偏移。因此，**知识转移需要同时解决分布和分辨率差异。**为此，我们将公共源人脸集Is作为桥接域，并建议通过两步蒸馏进行桥蒸馏以执行知识转移：

1）跨数据集蒸馏将从**私有域高分辨率人脸中预先学习的知识适应到公共域，**并将其蒸馏学习紧凑特征，

2）分辨率适应的蒸馏：**通过将适应的特征作为监督信号来训练学生模型，从而将知识从高分辨率公共面孔转移到其低分辨率版本**。在这种情况下，可以通过许多公共基准轻松实现Is，同时可以使用许多现成的面部识别模型并将其用作教师模型。 因此，我们的方法提供了一种经济的方法来学习用于促进低分辨率人脸识别的有效模型。

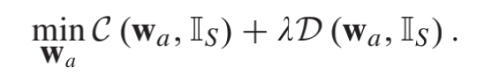
**Cross-Dataset Distillation**

跨数据集蒸馏解决了一个子问题，即应该从高分辨率模型中传递什么知识。它用于通过**适应过程**来压缩和适应从Ip到Is学到的教师知识，并应考虑两个方面。首先，**它应该保留从Ip学习到的高分辨率知识，同时有选择地增强它们在Is上的判别性**，以便在老师模型的帮助下正确的高分辨率知识可以从Is中提取出来。**其次，适应后的特征应该紧凑**，以便学生可以用极其有限的资源模仿它们。

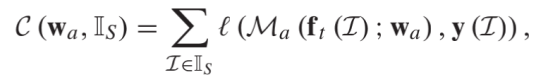
先前的工作表明，直接在低分辨率的人脸图像上进行训练通常会导致识别精度无法令人满意。 这种退化的基本原理是，当图像分辨率降低时，重要的面部细节会逐渐丢失。 为了缓解此问题，应将高分辨率的面部知识编码到训练好的模型中，以便可以**在识别过程中隐式执行细节重建。**

在我们的设置中，在丰富的高分辨率数据集IP上学习的教师模型Mt（I; wt）通常强大到可以从其自己的数据集中识别高分辨率的面孔。在这里，I代表输入的高分辨率人脸图像，而wt是模型参数。我们**希望老师的知识可以转移给学生模型Ms（I‘；ws）**，其输入是来自独立数据集IS或IT的低分辨率面部图像I’。但是，我们假设IS还包含高分辨率的人脸，以便教师网络可以学习高分辨率的细节。为了提高清晰度，我们使用符号I和I‘分别表示高分辨率和低分辨率的人脸图像。鉴于IP和IS可以有不同的分布，我们建议**首先将从IP中学习到的老师的知识适应到IS**。适应过程应考虑两个方面。**首先，它应该保留从IP中学到的高分辨率知识优势，同时有选择地增强它们在公共数据集IS上的判别性。其次，适应后的特征应紧凑，以便学生模型可以用极其有限的计算资源模仿它们**。

为了实现这一思想，我们**将自适应函数实例化为一个小型子网Fa（ft（I）； wfa）**将高维特征ft（I）映射到一个缩小的特征空间。通过**将此适应模块插入教师模型**，可以以数据驱动的方式学习最佳特征映射。给定适应的特征fa（I）= Fa（ft（I）；w fa），我们假设一个简单的softmax分类器Sa（fa（I）; wsa）可以很好地识别公共数据集Is中的面孔。将**Ma =（Fa，Sa）表示为完全自适应模块**，其中wa = [wfa; wsa]是要学习的参数。我们建议以以下目标**训练适应模块**，可以将其视为知识蒸馏的一种变体，

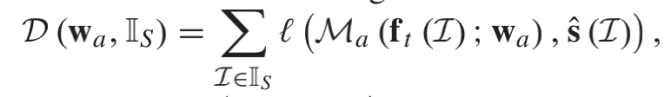


提出的目标是**分类损失C和蒸馏损失D**的平衡总和，其中λ为平衡权重。为了对高分辨率公共人脸集进行分类，我们有



其中我们采用了广泛使用的**交叉熵分类损失**l（·，·），而y（I）是输入图像I的identity。

**蒸馏损失D增强了适应特征在人脸分类上模仿原始特征的行为**。为此，我们首先微调教师在IS上的softmax层，以获得Sˆt （ft（I）；ˆwst），其中wˆst表示再训练的层权重。 可以将Sˆt视为一种特征选择器，它保留了最初学习的知识的判别性成分，用于识别IS中的人脸。那么蒸馏损失为：



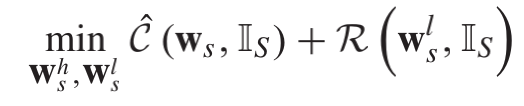
其中sˆ（I）= Sˆt （ft（I）; ˆwst ）/ T，T是用于软化softmax输出的温度。上面公式使用交叉熵来**近似软化后的教师的softmax特征**，从而确保适应过程引入了从IP中学习到的其他教师的知识。除了方程2给出的关于IS的识别能力外，还保留了教师网络的知识。 结果，**适应的特征模仿了原始高分辨率特征的判别选择的影响，但是尺寸大大减小**。训练自适应模块Ma之后，我们丢弃softmax层，仅保留**自适应和缩小的特征fa（I）**= Fa（ ft（I）；wst）**作为训练学生模型的监督。**

1. 与其他蒸馏方法的区别：与经典蒸馏方法不同，提议的桥蒸馏方法经历了两个蒸馏步骤：第一步是**将预训练的复杂模型适应到公共数据集上**，第二步**学习用更简单的模型模仿它**。 第一步，给定私有和公共数据集蒸馏模型本身，因此我们将所提出的算法称为跨数据集蒸馏。 因此，**教师无需接受有关公共数据集的全面重新训练**。在公共数据集上直接对教师进行再训练不仅耗费大量时间，而且可能会过度你和数据集并丢失以前学习的知识。 相比之下，提出的适应方法可以保留此类信息，并且训练起来要快得多。
2. 与其他学习方法的区别：最近的继续学习或增量学习方法[44]，[45]专注于使用新任务数据来训练网络，同时保留从原始任务数据中学到的知识。相比之下，我们的跨数据集蒸馏设置**主要执行知识适应**，而不是知识保存，**因此教师的能力可以转移到公共域**，这可以促进高分辨率和低分辨率实例之间的知识对齐。这样，可以保留高分辨率识别的能力，而将适应性知识转化为公共高分辨率数据集则可以避免被污染，从而导致有效的知识转移。 我们还注意到，我们的**跨数据集蒸馏将知识适应和知识转移结合在一起**，这是通过**微调和迁移学习**来实现的。

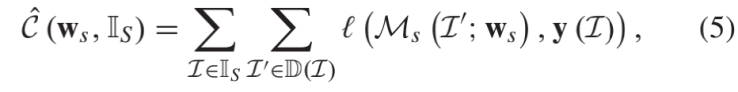
*Resolution-Adapted Distillation*

分辨率适应的蒸馏解决了子问题，即如何执行从**高分辨率人脸到低分辨率人脸的知识转移**。 给定适用于公共数据集的高分辨率知识，我们要求学生在推理过程中对它们进行近似。 由于学生的能力较弱，因此模仿适应知识的特征层应足够深。 随便说说，我们发现模仿层最好在身份层之前插入，以进行softmax分类，如图2所示。

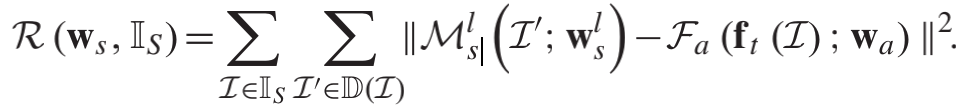
基于此设计，我们将Ms分为**Mls和Mhs（即下部和较高部分）**，并将其参数ws分为wls和whs。下部Mls对应于主要特征分支，直到模拟层为止，而Mhs由其余层组成。使用以下目标训练学生模型：



将人脸分类和特征回归结合在一个统一的多任务学习任务中。对于**分类损失Cˆ，我们仍然采用交叉熵**，但是这次对来自公共数据集**IS的下采样的低分辨率人脸版本**图像进行训练：



其中D（I）表示从I降采样的图像集。回归损失R定义为：



训练后，教师模型和适应模块将被丢弃。在推理过程中，学生模型根据任务将低分辨率的面部图像作为输入，并输出其分类的身份特征或标签。

*Implementation Details*

提出的框架设计得很灵活，因此原则上**教师模型可以采用任何模型**或其集合的形式，只要它们以用于面部分类的softmax层结束即可。请注意，许多最新模型[2]，[5]，[6]，[47]都满足此假设。在这项工作中，我们采用了两种最新的体系结构VGGFace2 [6]，CosFace [5]，它们具有112×112的输入分辨率和1024的嵌入特征尺寸，以及它们在教师模型中的集成。 VGGFace2在来自VGGFace2数据集的大规模高分辨率人脸数据集上进行了预训练[6]，并以224×224的分辨率工作，而CosFace在CASIA WebFace数据集[48]和大型私有数据集上进行了预训练。 在评估期间，不会对这些数据集进行任何重新训练或微调。 另外，我们假设其数据集是私有的，即，所提出的方法无法访问它们。

对于学生模型，我们设计了一种**轻量级的架构**，类似于[49]，[50]中提出的架构。如图2所示，将低分辨率的面部图像作为输入。 我们使用p×p的各种分辨率训练学生模型，其中p = {96，64，32，16}。 该架构具有十个卷积层，三个最大池化层和三个完全连接的层，这些层和ReLU非线性层交错。 在3×3层之间相交了几个1×1卷积层，以节省存储空间并提高推理速度。 建立两个跳过连接以增强信息流。 全局平均池用于进行最终预测，以便体系结构可以处理任意分辨率。在这种架构下，**参数的大小仅为0.21M，仅为教师规模（VGGFace2为26M或CosFace为37M）的0.81％或0.57％。**

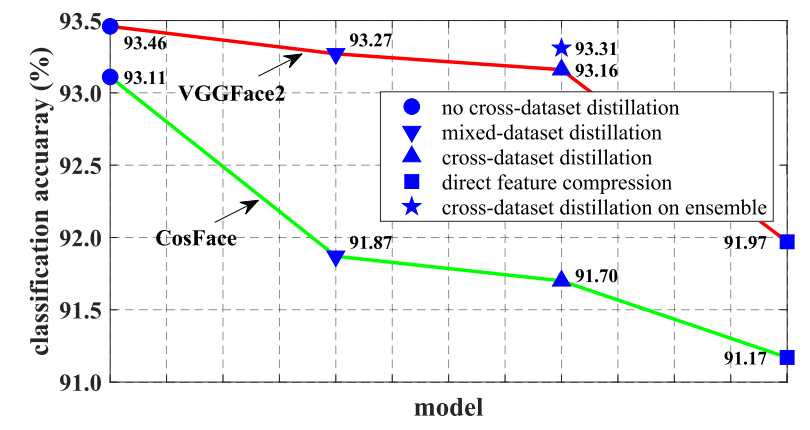
我们的**适应模块将教师模型的softmax层之前的特征作为输入**。**它具有两个全连接层**，分别具有512和128个单元。可以使用更复杂的体系结构，但要花费额外的训练时间。学生模型和适应模块中的所有权重均通过Xavier的方法进行初始化。

在训练过程中，首先在IS的低分辨率面孔上对学生模型进行预训练，然后在**根据教师模型提取的高分辨率知识进行调整的监督下进行微调。** 来自IS的训练图像按不同尺度进行下采样，以模拟各个级别的退化。 执行批量标准化以加快收敛速度。 我们使用随机梯度下降来训练学生模型。 在所有实验中，批次大小和学习率分别设置为256和0.001。

**EXPERIMENTS**

**数据集：UMD Faces，LFW和UCCS数据集**

Evaluating Adaptation on UMDFaces

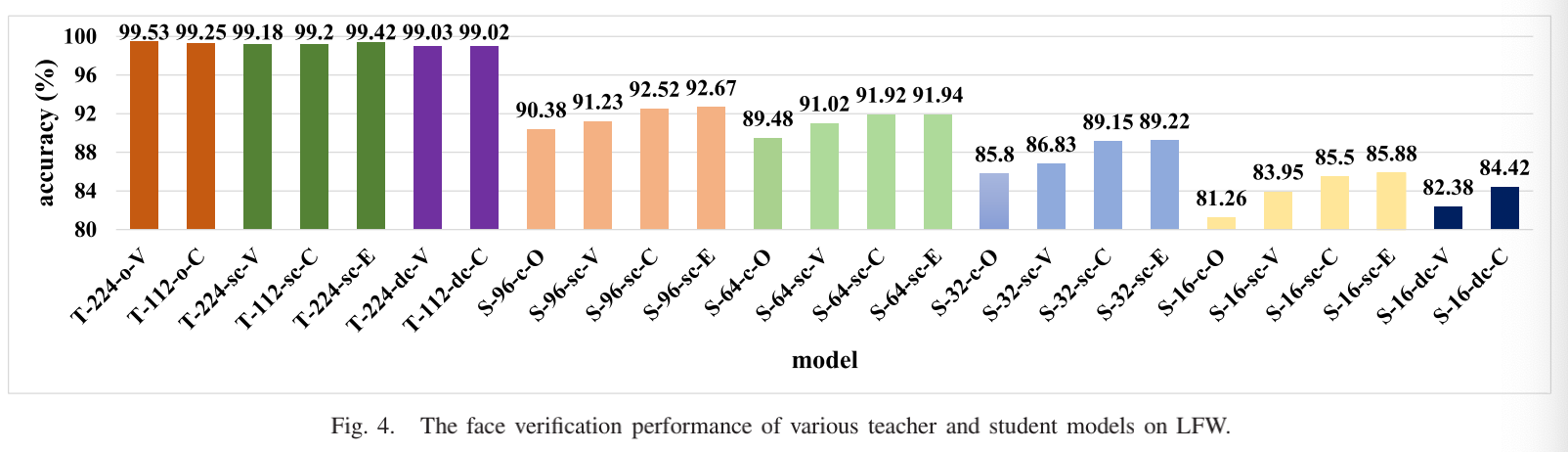


在第一个实验中，我们旨在研究通过跨数据集蒸馏提出的自适应模块带来的改进。为此，我们评估了四种情况：

1. 没有跨数据集蒸馏。在这种情况下，由VGGFace2或CosFace教师模型提取的预训练的2,048或1,024维特征将直接用于监督学生模型。
2. 直接特征压缩，无需蒸馏。此设置仅将预学习的老师的知识适应于具有分类损失的低分辨率设置，但不包括蒸馏损失。
3. 提出的交叉数据集蒸馏，它等于结合了两种损失的公式
4. 混合数据集蒸馏。在这种情况下，我们假定可以访问私有高分辨率数据集并将其用于训练。在我们的实验中，我们从私人高分辨率数据集（用于VGGFace2模型的VGGFace2数据集或用于CosFace模型的CASIA-WebFace数据集）中随机选择1,000个受试者的训练实例，并将其与UMDFaces训练集组合为混合训练数据集，然后进行训练较大的9,419路分类器，以将教师的知识适应到公共高分辨率数据集中。

在后三种情况下，通过自适应将特征减少到128维。图3显示了这四种情况在UMDFaces上的识别准确度，表明通过跨数据集蒸馏的特征压缩比直接压缩具有更好的性能。此外，从案例1到案例3，性能仅略有下降，特别是在没有交叉数据集蒸馏的情况下，由两名教师进行的跨数据集提纯相对于VGGFace2监督的学生模型下降了0.15％，同时特征大大缩小了128个维度，因此大大压缩了冗余并节省了资源消耗。此外，使用VGGFace2和CosFace模型，经过训练的混合数据集蒸馏分类器分别达到91.87％和93.27％的精度。它表明我们的跨数据集精馏仅在没有合奏的情况下精度下降很小，而在合奏的情况下实现了精度提高。这意味着我们的方法即使不访问私有数据集也可以执行有效的知识适应。

**Evaluating Face Verifification on LFW**

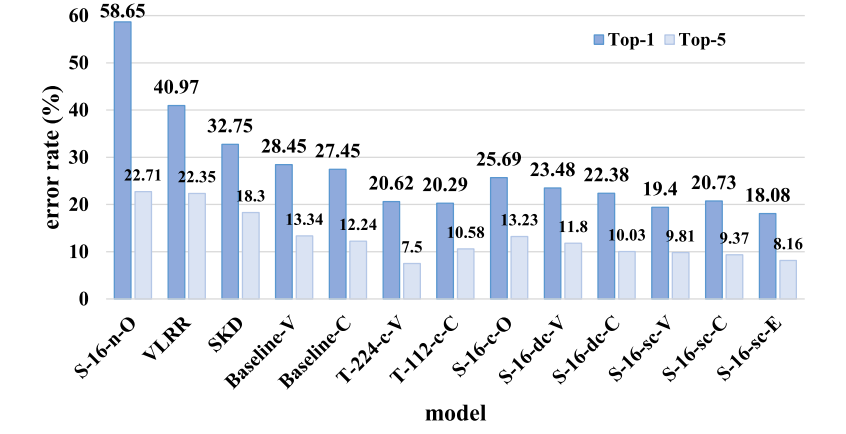


在第二个实验中，我们在各种**输入分辨率p** = {96，64，32，16}，监督信号x = {c，s，dc，sc}和蒸馏教师T={O,V,C,E}的情况下，广泛评估了该方法在面部验证任务上的性能。此处，监督信号的缩写仅表示类级别监督（c），丢弃类监督时的跨数据集蒸馏（s），除类监督的直接蒸馏（dc）和类级别监督下的跨数据集蒸馏（sc）。在dc的情况下，自适应模块将被丢弃，学生模型将直接从老师那里学习模仿2,048或1,024维的预训练特征。蒸馏教师的缩写分别表示没有教师**（O），VGGFace2（V），CosFace（C）和他们的集合（E）**。在这种情况下，特定设置的学生模型表示为其分辨率，监督信号和蒸馏教师的组合，例如S-p-x-t，其中p∈p，x∈x和t∈T。为简洁起见，我们将教师模型表示为相同的规则，例如T-112-sc-C代表以112×112的分辨率并与监督信号sc相适应的交叉数据集蒸馏CosFace教师模型。我们还使用T-224-o-V和T-112-o-C表示原始的VGGFace2和CosFace教师模型，而没有进行适应。

从图4可以总结出几个重要的观察结果。首先，**精度随着分辨率的降低而逐渐下降**，这是所期望的。尽管S-96-sc- {V，C，E}的表现仍然比其老师差，但它们达到了相当不错的准确度（例如S-96-sc-E为92.67％），其模型大小为0.21M ，表明桥梁蒸馏提供了一种在资源有限的设备上重新部署现有重量级模型的方法。其次，只有类监督的学生模型S-p-c-O的表现始终比S-p-sc- {V，C，E}差，其中p∈p。这意味着由于对低分辨率面部进行直接训练，因此没有交叉数据集蒸馏的学习模型可能缺乏捕获判别性高分辨率细节的能力。相反，学生模型可以通过模仿高分辨率老师的知识来明确学习重建缺失的细节。第三，从集合中蒸馏出的学生模型优于其从单一老师那里提炼出的相应模型，这表明转移更多信息的知识优势将带来更好的绩效提升。最后，为了显示交叉数据集蒸馏的影响，我们将每个S-16-sc-t与它的直接蒸馏模型S-16-dc-t进行比较，其中t∈{V，C}。结果表明，通过交叉数据集蒸馏将自适应结合起来可以提高推理性能。

我们进一步注意到，这种改进在各种分辨率和不同蒸馏教师的指导下都是一致的。 因此，可以公平地得出结论，所提出的桥蒸馏方法的适应性对于学习判别特性是必要的。我们怀疑引入蒸馏损失来补充分类损失可以在很大程度上防止针对目标域的训练偏向，从而导致过拟合的发生率降低。此外，详细信息显示在LFW的接收器工作特性曲线（ROC）中（请参见图5）。 与其他五种型号相比，我们的S-16-sc-E型号拥有最佳的性能。 特别是当假阳性率为10.0％时，S-16-sc-E的真实阳性率为80.4％，比S-16-cO的阳性率高9.6％，表明我们提出的桥式蒸馏的有效性 假阳性率低的方法。

Evaluating Face Identifification on UCCS



在最后的实验中，我们使用具有挑战性的UCCS数据集，与低分辨率人脸识别任务的最新模型进行了比较。在此数据集上，我们直接与超低分辨率人脸识别模型（VLRR）[7]和选择性知识蒸馏（SKD）[1]作比较，这两种模型都是在以下情况下工作的：分辨率16×16。我们遵循类似的实验设置，将180个对象的子集随机分为训练集（4,500张图像）和测试集（935张图像），其中面部图像归一化为16×16，而两组中的身份是互斥的，在训练集上微调各种学生模型S-16-xT的softmax层，并对测试集进行评估。我们还以类似的方式微调了教师模型T-224-c-V和T-112-c-C，而无需在UCCS上进行蒸馏。然后，我们将它们直接分解为两个基准（Baseline-V和Baseline-C）。此外，我们从零开始在UCCS上训练了新模型S-16-n-O进行比较。

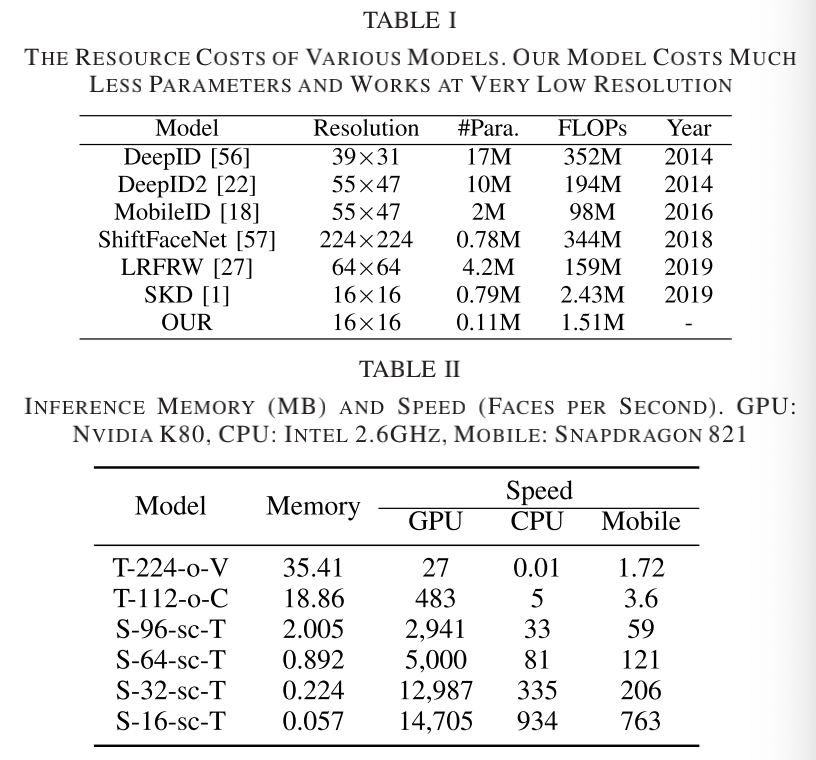
图6显示了结果，其中报告了top-1和top-5错误率。我们可以发现，经过微调的教师T-224-cV和T-112-c-C以及经过蒸馏的基线均明显优于VLRR和SKD，但S-16-n-O则更差，这表明模型经过了预训练，外部数据集上的数据可以提供有关低分辨率识别问题的有价值的先验知识。 因此，与finetune教师相比，所有finetune学生的错误率均比VLRR和SKD低得多，而参数却大大减少。 此外，经过微调的带有交叉数据集蒸馏的学生模型可以达到最佳效果。 我们怀疑，即使输入分辨率非常低，也可以利用来自经验丰富的老师的丰富预训练知识，使学生能够成功地重建内容丰富的高分辨率细节。

Effificiency Analysis

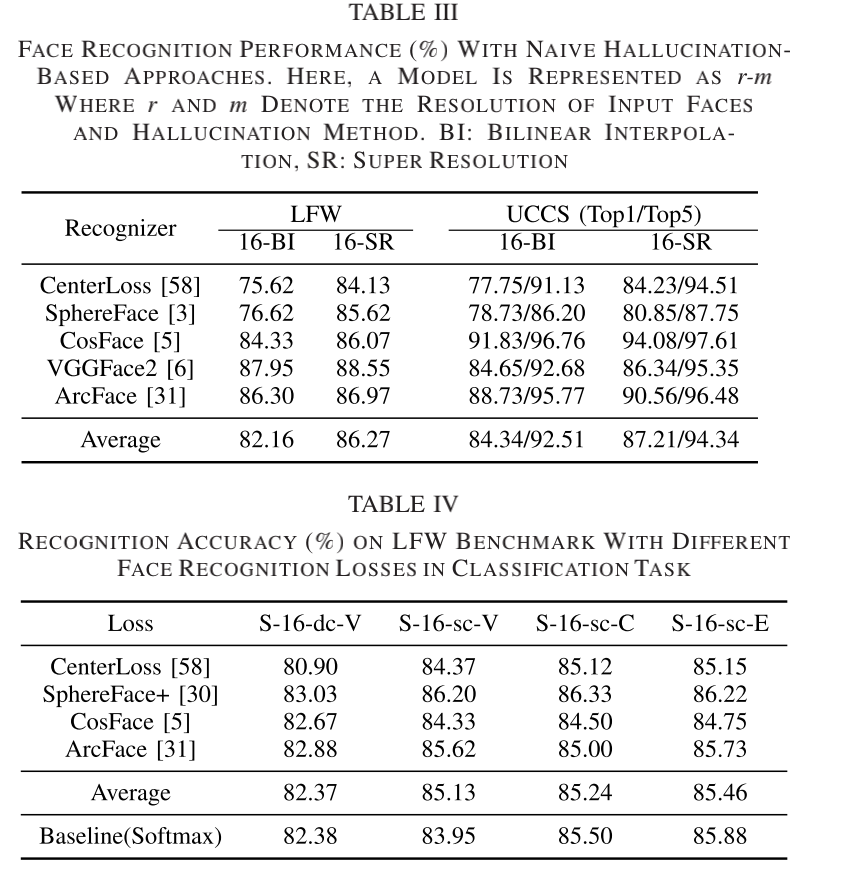
学生可以模仿老师的表现，同时大大降低存储，内存和计算的成本。

如表I所示，我们与最先进的轻型人脸模型进行了比较。 DeepID [56]和DeepID2 [22]模型采用低分辨率输入和数十个网络，导致计算复杂。 MobileID [18]使用较少的2M参数压缩DeepID2模型，该参数仍需要98M FLOP。 最近，ShiftFaceNet [57]应用具有0.79M参数的轻型网络来识别224×224人脸图像，因此仍然要付出很大的换行负担。 LRFRW [27]接受64×64和4.2M参数的低分辨率输入，但FLOP达到159M。 最新的SKD [1]可以识别16×16的人脸图像，而参数成本为979万。 与它们相比，我们的学生模型仅具有非常少的参数，而参数为0.11M，计算成本更低，而分辨率却非常低。

从表II中我们可以发现，与分辨率分别为96×96、64×64、32×32和16×16的学生相比，内存降低比率因子分别为18×，40×，158×和621×。使用VGGFace2老师时，或与CosFace老师相比为9×，21×，84×和331×。特别是，对于16×16的学生，内存仅为0.057MB。除了显着节省内存之外，两者的计算成本仍然大大降低。如表II所示，在NVIDIA K80 GPU上，教师模型VGGFace2或CosFace的推理速度仅为每秒27或483个面，而Sp-sc-t的推理速度则高达2,941、5,000、12,987和14,705架构，其中t = {V，C，E}，分辨率大小p分别为96、64、32和16。即使在CPU中，推理速度也非常快。特别是，将模型部署在移动设备上时，模型S-16-sc-T的推理时间仅为1.31ms。结果，建议的小型学生模型S-16-sc-T能够在GPU，CPU或移动电话上每秒处理14,705、934或763张脸。这些结果表明，我们的桥式蒸馏方法为在低端设备上重新部署现有的重型预训练模型提供了一种实用的解决方案。



*Comparison Analysis*



为了进一步研究该方法的有效性，我们进行了两个实验比较，包括与基于幼稚幻觉的方法的比较以及具有不同面部识别损失的模型的比较。首先，我们在LFW和UCCS基准测试上对基于幼稚幻觉的方法进行了实验比较。在我们的实验中，我们检查了两种幻觉方法，BI（双线性插值）和SR（最近的FSRNet方法[59]），以及五个高分辨率人脸识别模型（CenterLoss [58]，SphereFace [3]）。 ]，CosFace [5]，VGGFace2 [6]和ArcFace [31]）。请注意，这些模型由其作者提供，其参数在特定数据集中进行了预训练。输入的脸部分辨率为16×16。标签。 III显示结果。从结果中我们可以发现，基于幻觉的方法通常可以实现良好的识别精度。例如，即使使用简单的双线性插值对人脸图像进行幻觉，在LFW上的平均识别精度也可以达到82.16％。还应注意，复杂的幻觉方法通常会导致更大的性能改进，同时在重建过程中会花费更多的计算负担。此外，由于SphereFace在CASIA-WebFace数据集上进行了预训练，因此在UCCS上的性能比CosFace和ArcFace差[48]，其中人脸图像的分布与UCCS中的监视人脸有很大不同。我们的方法可实现可比的性能，同时节省更少的计算和内存资源。例如，我们的S-16-sc-E学生模型在LFW上的准确度为85.88％，在UCCS上的准确度为91.84%@Top5。接下来，我们检查分类任务的不同损失函数对识别性能的影响。除了基线softmax损失以外，我们还对四种学生模型进行了四种人脸识别损失的实验（CenterLoss [58]，SphereFace + [30]，CosFace [5]和ArcFace [31]）。标签。 IV显示了结果，其中我们提出的方法在具有更好的蒸馏损失的情况下，在各种面部识别损失下具有一致的性能提升。这意味着我们提出的方法是不可知的。

**conclusion**

本文提出了一种新颖的桥梁蒸馏方法，以解决资源有限的低分辨率人脸识别任务。 这种方法的核心是有效的师生框架，该框架依赖于新颖的跨数据集分布和适应分辨率的蒸馏算法，该算法首先调整教师模型以保留具有区别性的高分辨率细节，然后使用它们来监督 训练学生模型。 大量的实验结果表明，该方法能够将信息性的高分辨率知识从教师传递给学生，从而大大减少了模型，减少了参数，并且推理速度极快。 在未来的工作中，我们将探索在广泛的视觉任务上进行多桥知识蒸馏的可能性。