通过选择性知识蒸馏在野外进行低分辨率人脸识别

葛世明，IEEE高级会员，赵胜伟，李晨宇，李嘉，IEEE高级会员

***摘*要—通常，野外使用人脸识别模型需要以极低的计算成本来识别低分辨率人脸。为了解决这个问题，一种可行的解决方案是以最小的性能下降为代价，压缩复杂的face模型以实现更高的速度和更低的内存。受此启发，本文提出了一种通过选择性知识提炼识别低分辨率人脸的学习方法。在这种方法中，首先对two-stream卷积神经网络（CNN）进行初始化，以分别通过teacher流和student流识别高分辨率面孔和分辨率退化后的面孔。教师流由复杂CNN表示，用于用于高精度识别，学生流由更加简单的CNN表示，用于用于低复杂度识别。为了避免在学生流中出现显着的性能下降，我们然后通过解决稀疏图优化问题，有选择地从教师流中提取最有用的面部特征，然后将其用于正则化学生流的微调过程。通过这种方式，实际上通过同时处理两个任务且使用有限的计算资源来训练学生流：通过特征回归来近似提供最多信息的面部线索，以及通过低分辨率的面部分类来恢复丢失的面部线索。实验结果表明，学生流在识别低分辨率人脸方面表现出色，仅花费0.15 MB的内存，在CPU上每秒以418张面孔运行，在GPU上以每秒9433张面孔运行。**

***索*引词-野外人脸识别，两-流体系结构，知识提炼，CNN。**

1. 介绍

ACE是将一个主题与另一个主题区分开的基本属性，需要多次识别

**F**

于2018年4月3日收到手稿;于2018年10月10日修订;公认

发布日期：2018年11月22日。发布日期：2018年11月30日。日期为当前版本的2018年12月12日。这项工作部分受2016YFC0801005资助的国家重点研究和发展计划的支持，部分受中国自然科学基金会资助的61772513和61672072的支持，部分受北京新星计划的支持。该计划由Z181100006218063资助，部分由中国科学院信息工程研究所的国际合作项目资助，由Y7Z0511101资助。S. Ge的工作得到了中国科学院青年创新促进会的支持。Aydin Alatan教授是负责协调对该稿的审稿并批准发表的副编辑。（通讯作者：贾莉）

Ge。S.加入中国科学院信息工程研究所，北京100095（电子邮件：geshiming@iie.ac.cn）。

Zhao S.和C. Li在中国科学院信息工程研究所，北京100095，以及中国科学院大学网络安全学院，北京100049。

李，北京航空航天大学计算机科学与工程学院，虚拟现实技术与系统国家重点实验室，北京100191，北京航空航天大学北京大数据与脑计算高级创新中心100191，中国（电子邮件：jiali@buaa.edu.cn）。

数字对象标识符10.1109 / TIP.2018.2883743



图1.对于熟悉相应高分辨率面部的对象，可以识别出低分辨率面部。尽管低分辨率人脸可能质量低劣，纹理模糊，照明不佳以及遮挡物多样化，但来自高分辨率人脸的知识可以帮助提取有区分性的特征以实现有效的人脸识别。

现代计算机视觉和多媒体应用中的每一天。在这些应用中，需要在移动电话[5]甚至智能相机[6]上重新部署许多著名的面部识别模型[1] – [4]，以满足在具有极低的计算成本和内存占用量的情况下识别低分辨率面部的现实要求（即，野外人脸识别[7]）。为此，有必要探索一种可行的解决方案，以解决人脸识别中的关键挑战：如何将现有的复杂人脸模型转换为更有效的模型，该模型仍然适用于低分辨率人脸而不会显着降低识别精度？

与高分辨率面部相比，低分辨率面部具有其独特的视觉属性。如图1所示，低分辨率的面部缺少许多细节。然而，对于熟悉相应的高分辨率面部的受试者仍然可以识别它们，这意味着人类的神经系统可能具有恢复熟悉的面部缺少的细节的能力。受这一事实的启发，已经提出了许多现有的低分辨率人脸模型，这些模型可以大致分为两类：幻觉（hallucination）类和嵌入（embedding）类。

幻觉类别中的模型建议在识别之前重建高分辨率面部[8] – [10]。例如，Kolouri和Rohde [8]描述了基于运输的

单帧超分辨率方法可自动构建高分辨率的面部表情的非线性拉格朗日模型。之后，通过找到最适合给定低分辨率数据的模型参数来增强低分辨率面部图像。Jian和Lam [9]观察到在不同分辨率下的人脸图像的奇异值具有近似线性关系。基于这种观察，他们首先将奇异值分解用于面部表示，以学习低分辨率和高分辨率面部对之间的映射功能，然后同时执行幻觉和低分辨率面部的识别。Yang等人提出的类似方法[10]使用稀疏表示来执行联合的幻觉和识别，这可以合成具有识别保证的特定于每个人的低分辨率面部图像。通常，这些方法在识别重构的高分辨率面部时表现出令人印象深刻的性能，而超分辨率操作通常会带来额外的计算成本，并导致识别速度较低。

与基于幻觉的模型不同，嵌入类别中的模型通过使用各种外部人脸上下文（external face contexts）直接从低分辨率人脸中提取区分特征。例如，Biswas[11]提出将低分辨率的面部图像嵌入到欧几里得空间中，使得它们之间的距离在变换后的空间中可以很好地近似于高分辨率面部的最佳距离。任等人[12]提出了耦合核嵌入，以将具有不同分辨率的面部图像映射到无限子空间上。然后，通过最小化低分辨率和高分辨率空间中它们的kernel gram矩阵所捕获的差异，在新空间中进行识别过程。一般而言，基于嵌入的方法中最重要的过程是将知识从高分辨率的面孔转移到低分辨率的面孔。但是，在此过程中需要仔细解决的关键问题是正确地仅转移所需的知识，而不是将所有知识从高分辨率域转移到低分辨率域。这种选择性的知识转移是将现有的面部模型转换为更有效的模型（在低分辨率的面部模型上也能正常工作）的最重要挑战之一。

受这一事实的启发，我们提出了一种选择性的知识蒸馏方法，用于野外低分辨率人脸识别。如图2所示，首先对两流CNN进行训练，以通过使用两个流同时识别高分辨率人脸及其分辨率降低的版本。 这两个流包括在高分辨率的面部上操作的教师流和在低分辨率的面部识别上更为简单的学生流。为了确保学生流具有与教师流相当的识别性能，我们接下来通过解决稀疏图优化问题，选择性地从教师流中提取包含最多有用信息的面部特征，然后将其用于调整学生流的微调过程。通过这种方式，实际上通过同时处理两个任务且使用有限的计算资源来训练学生流：通过特征回归来逼近最多有用信息的面部线索，以及通过低分辨率的面部分类来恢复丢失的面部线索。请注意，教师流可以

采用现有的任何深度人脸模型的体系结构，这意味着所提出的方法可以将任何现有的面部模型转换为更简单的模型，具有更快的速度和更低的内存，并且性能下降最小。在四个公共数据集上的实验结果表明，学生流在以极低的分辨率识别人脸方面表现出色。特别是，它仅使用0.15MB内存，在单个CPU线程上以每秒418个face的速度运行，在GPU上以9 433个face的速度运行。

本文的主要贡献概述如下。1）我们提出了一种通过选择性知识蒸馏的人脸模型压缩方法，该方法可以在不显着降低性能的情况下大大减小模型的大小和复杂性；2）我们提出了一种基于图的优化算法，该算法可以从现有的面部模型中提取出最具判别力的面部特征，从而可以用于监督低分辨率面部模型的训练过程；3）我们进行了全面的实验，表明压缩模型可以以与最新的高分辨率人脸模型相当的精度实现极高的识别速度。

本文的其余部分安排如下：第二节回顾了相关工作，第三节介绍了选择性知识提炼方法。在第四节中进行了广泛的实验，以评估所提出的方法，在第五节中对本文进行了总结。

1. 相关作品

我们在本文中提出的方法旨在从复杂的人脸模型中提取知识以进行低分辨率人脸识别。因此，我们从通用的人脸识别模型，低分辨率人脸识别和知识提取三个方面简要回顾了相关工作。

1. *一般人脸识别模型*

最近，一般的人脸识别技术已经从经典的浅层框架[13]，[14]演变为深层的[1]，[3]，[4]，[15] – [18]，其性能得到了显着改善。对于深层方法，区分它们的关键因素是它们采用的损失函数。例如，DeepFace [1]是通过构建具有识别损失的3D人脸来整合卷积神经网络（CNN）的早期尝试。之后，已经提出了各种损失函数来训练人脸识别CNN，例如三重损失[3]，[4]，中心损失[15]和范围损失[17]。在[18]中，识别面孔及其属性的任务被同时考虑以提高识别性能。对于DeepID系列，首先在[2]中分别训练了几个使用不同面部补丁的小型CNN，其后续工作结合了面部验证信号[19]并更改了基础网络[20]以提高准确性。

一般来说，这些深度模型在识别一般面孔方面取得了令人印象深刻的性能。但是，如表I所示，许多此类通用模型具有大量参数，高维特征表示和复杂的分类函数以进行推理。这些模型的复杂性阻止了

表我

一般人脸识别的代表性深度模型。

#train：训练图像数量，分辨率：输入面部分辨率，暗淡：输出特征尺寸，＃lyr：网络层数量，＃par：

模型参数数量



它们可以直接部署在计算资源有限的野外。尽管DeepID系列将低分辨率人脸作为输入，但未探索低分辨率人脸的独特属性。为了进一步提高低分辨率人脸识别的性能，有必要在分辨率降低的时候探索缺失的特征。

1. *低分辨率人脸识别*

通常，低分辨率人脸识别有两种方法。幻觉类别（hallucination category）旨在在识别之前重建高分辨率面部，而嵌入类别建议通过嵌入模式直接从低分辨率面部提取特征。在幻觉类别中，Kolouri和Rohde [8]构建了高分辨率的面部表情的非线性拉格朗日模型，然后找到最适合低分辨率面孔的模型参数。Jian和Lam [9]提出了一个基于奇异值分解的框架，并同时进行了幻觉和识别。在[10]中，提出了一种基于稀疏表示的联合人脸幻觉和识别框架。该框架可以合成特定于人的低分辨率人脸以进行识别。在[22]中，提出了一种系统，该系统通过使用稀疏表示来识别人脸，使用了具有涉及许多自然和面部图像的特定词典。而且，像[23]和[24]这样的深度模型可以从低分辨率的面孔生成极其逼真的高分辨率图像。但是，由于复杂的高分辨率人脸重建过程，这种基于幻觉或超分辨率方法的速度可能会有些慢，这阻碍了它们在计算资源有限的现实情况下的直接部署。

替代重建高分辨率面部过程，一种更直接的方法是将低分辨率面部嵌入各种外部上下文（external contexts）中，以恢复分辨率降低丢失的特征。受此启发，一些方法提出将高分辨率和低分辨率的人脸都转换为 [25]-[31]统一特征空间中用于匹配，而在[32]和[33]中采用多尺度（多分辨率）同时分析人脸以提取更好的特征。在[34]中，

通过多维缩放，学习通用的变换矩阵，以同时变换低分辨率和高分辨率训练图像的面部特征。Shekhar等[35]提出了一种联合的稀疏编码技术，用于低分辨率的鲁棒识别，而Wang等人[36]尝试使用深度学习方法解决非常低分辨率的识别问题。在[37]中，CNN被采用了基于流形的轨迹比较策略，用于视频中的低分辨率人脸识别。

从这些方法中，我们发现基于嵌入的方法的核心思想是转移（或利用）来自高分辨率面部的知识。结果，低分辨率人脸识别的性能主要受两个关键因素影响：传输什么知识以及如何利用它。换句话说，应该从高分辨率数据（或模型）中选择性提取期望的知识，并以正确的方式指导低分辨率人脸识别过程。这也是本文的核心思想。

1. *知识提取*

除了从高分辨率的面孔中挖掘知识之外，获取低分辨率的人脸模型（即学生网络）的另一种方法是直接从预先训练的复杂面孔模型（即教师网络）中提取此类知识。随着更深，更宽的网络的发展，许多工作[38]-[49]都采用了这种蒸馏技术，将复杂的模型（或整体模型）压缩为更易于部署的简单模型。其中，Luo等人[42]利用大型教师网络或某些网络集合的所学知识作为监督，来训练一个紧凑的学生网络进行人脸识别。在他们的方法中，在较高的隐藏层中选择了最相关的用于人脸识别神经元进行知识转移。Lopez-Paz等[50]提出了将蒸馏和学习与特权信息相结合的通用蒸馏框架。Su和Maji [51]提出了交叉质量蒸馏，以通过分别使用标记的高分辨率图像，标记的局部对象和彩色图像来学习用于识别低分辨率图像，非局部对象和线条画的模型。Radosavovic等。[52]提出了数据提炼，以整合来自未标记数据的多次转换的预测，以自动生成新的训练注释。

综上所述，知识蒸馏的核心部分是速度和性能之间的权衡，而这种技术提供了将许多复杂模型转换为可以在野外部署的简单模型的机会。请注意，在这项研究中，我们不仅尝试将复杂的人脸模型提纯为简单的人脸模型，而且还探索了使用分辨率降低的人脸作为输入以进一步加快识别速度并同时保持识别准确性的可行性。通过这种方式，低分辨率人脸识别和知识提炼可通过单个框架同时解决。

1. 该方法

我们的两流知识提炼框架由教师流和学生流组成（见图2）。

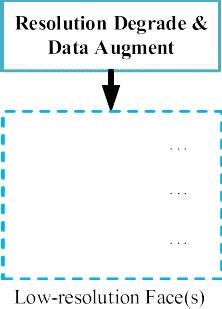
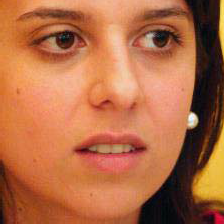
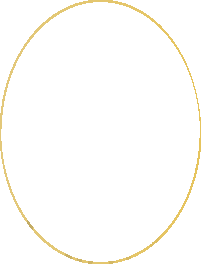
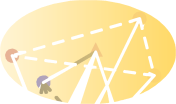
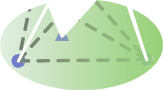
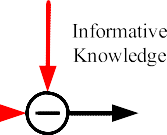
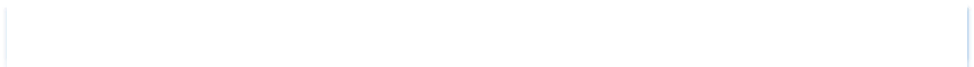
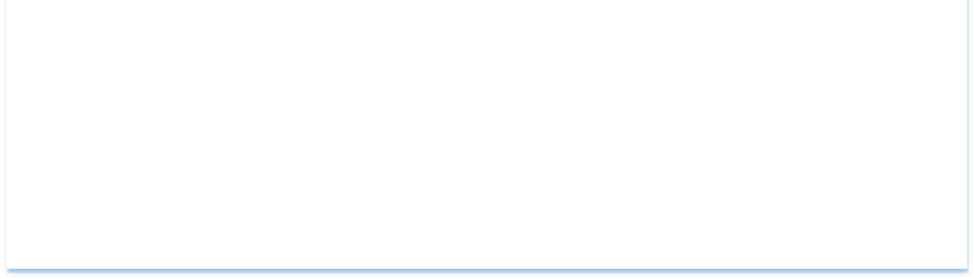
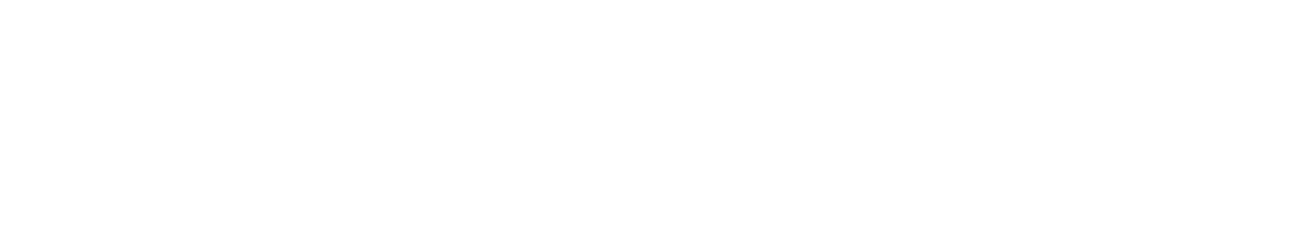


图2.我们的方法框架。蒸馏过程包括三个阶段。在第一阶段，我们初始化教师和学生流以分别识别高分辨率的面孔和其低分辨率的版本。在第二阶段，我们通过解决稀疏图优化问题，选择性地从教师流中仅提取最有用的面部特征。在第三阶段，被选择的特征用于指导学生流的微调过程。以此方式，实际上通过在有限的计算资源下同时处理两个任务来训练用于识别野外低分辨率人脸的学生流：选择性特征逼近和低分辨率人脸识别。

教师流可以采用以前已经训练过的任一复杂的面部识别神经网络（并且训练数据可能不再可用）。蒸馏过程旨在学习简单而紧凑的学生流

例如，使用来自私有数据集中的其他人脸图像对在互联网上发布的CNN模型进行了预训练）。

**学生流：**

它模仿教师流在野外真实场景中的实际部署

学生流C:\Users\pc\AppData\Roaming\Tencent\Users\904759374\QQ\WinTemp\RichOle\0(1HU7LON7PL110T9$HFTHU.png是一个更加简单的CNN模型，用来对LR人脸图像进行识别。

从StudentsFaceSet Ds进行学习，|Ds|是高分辨率人脸图像数目，对于每一个高分辨率的脸Fi，StudentsFaceSet包含n个对应的resolution-degraded版本。将第j个版本记为Fij。LR-HR图像对应相同的标签Li。这里假设identityset Ls里面有C类，第C类的HRfaces数量记为Kc，Ds等于Kc的和

*j* =1

*i*=1

学习过程包括三个阶段：1）

使用一个复杂的CNN或在高分辨率的人脸图像上预先训练的几个CNN的集合作为教室流的初始化，然后学生流对带有身份标签的低分辨率的人脸图像进行分类。

*j* =1

2）选择阶段从教师流中提取最有价值的知识

3）微调阶段：通过联合进行特征回归和人脸身份分类，将教师流提取到的知识和LR人脸图像同时监督学生流微调进程

对这三个阶段的详细描述如下：

。



1. *定义*

为了简单起见，我们将两流CNN的关键组件定义如下：

**教师流**：教师流φt（F; Wt）是一个复杂的CNN（或几个复杂的CNN的集合），具有

预训练好的以识别高分辨率面部F的参数Wt的集合。在这里，我们假设Wt从教师脸部集合Dt吸收了大量高分辨率脸部图像中编码的丰富知识，Dt中每个脸部都由身份集合Lt中的整数标记。通常，训练集面部图像的数量| Dt |非常大，可能对学生流是不可见的

1. *两流CNN的初始化*

如图2所示，我们的两流CNN通过教师流φt（F; Wt）和学生流同时进行高分辨率和低分辨率人脸识别

。教师的参数集Wt

可以通过最新的人脸识别模型初始化

，例如具有VGG16架构的VGGFace [4] [53]，具有GoogLeNet架构的FaceNet [3] [54]和具有ResNet50架构的VGGFace2 [21] [55]。作为一个有代表性的例子，我们在教师流中使用VGGFace [4]的体系结构，并使用作者发布的模型初始化Wt。请注意，VGGFace已在庞大的人脸图像数据集Dt上进行了预训练，Dt我们认为在知识蒸馏过程中不再可用。

22471;

图3.提出的学生网络的结构。它包含十个卷积层，三个最大池化层和三个全连接层。它仅需要79万个参数（不包括最后一个softmax层），比现有的高分辨率人脸识别模型（例如VGGFace）小得多。



学生流φs（F˜; Ws）旨在识别低

*分*别为φs（F˜; Ws）。φˆt（F; Wˆt）是特征提取

ˆ φt(F; ˆ Wt)，是用来提取高分辨率图像身份特征的在softmax层之前的特征提取尾端。ˆ φs( ˜ F; ˆ Ws)是在模仿层之前的主特征分支，用来提取近似特征来和教师流进行匹配。学生流的微调可以描述为通过ˆ φs模仿特征表示ˆ φt来提高φs最终的识别能力。问题是：如何进行跨网络监督？

分辨率人脸F˜

在Ds上训练的紧凑网络上进行。

用于提取身份的softmax层之前的后端

因此，我们采用了一种轻量级的网络架构，

高分辨率人脸图像的功能，而

*φ*ˆ*s (F*˜ ; Wˆ *s)*

类似于[56]。如图3所示，学生流将低分辨率（例如32×32）的人脸作为输入，模型大部分使用3x3卷积核，并在每个池化步骤之后增加通道数量。此外，全局平均池化用于进行预测以及在3×3卷积之间使用使用1×1过滤器来降维。请注意，此处采用1×1×D模拟层来接收将来来自教师流的知识，其中D为

教师流中学习的高分辨率人脸表示的维度。因此，来自模仿层的特征可用于与教师网络进行特征近似。此外，由于学生模型的能力较弱，因此模仿所传递知识的特征层应足够深，因此我们在模拟层和softmax层之间经验性地插入了身份层。身份层还扮演特征压缩的角色。最后，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。结果，以Ws的参数数量仅达到0.79M，

这仅是教师参数集Ws的0.57％（即，

138M).这些参数首先使用xavier初始化，然后通过最小化Ds上的分类损失进行优化：

DSN

W∗

对应于主要特征分支，直到模拟层

并用于提取近似特征以匹配教师。学生流的微调过程可以是

描述为模仿特征表示φˆt（F; Wˆt）与φˆs（F˜; Wˆs）以提高φs（F˜的最终识别能力）; Ws）。问题是：如何进行跨网络

监督？

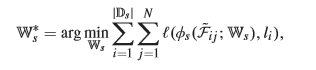
通常，教师流具有非常强大的使用教师面部集合Dt中可用的标识来识别高分辨率人脸的功能。然而，由于来自Dt和Ds identity的不同，它不能直接从学生面部集合Ds识别不熟悉的面孔。在这种情况下，来自教师流的知识可能包含一些错误，这将误导学生流的微调过程。从而，

我们选择性地提取最多可利用信息的知识，拒绝从φˆt中提取错误的知识，以提高子网络φˆs特征提取的能力和学生流φs的人脸识别能力。

为此，一种可行的解决方案是通过使用教师流提供的特征从Ds中找出最多可利用信息的面孔，而这种有信息的面孔可以定义为类间相似度小而类内部信息量大的面孔。

为此，我们将选择性知识提取过程公式化为在图上的推理问题，图的节点表示面部，节点之间的边表示他们之间的相关性。

*s*＝ARG min � �T（ωs（f˜IJ；WS），Li）；(1)

瓦数*i=1 j =1*

其中l（ ）是用于衡量分类损失的softmax函数。最小化问题（1）可以通过采用标准反向传播的随机梯度下降法来解决[57]。

C教师流中的选择性知识提炼

如图4所示，一个紧密连接的图将包含来自C个面部类的所有节点之间的大量边，从而减慢了推理过程。为了有效地进行基于图的推理，我们为每个面部类别添加了一个质心节点，然后构造了一个稀疏的

连通图G = {V，E}。在图G中，节点集V

包含两种类型的节点：面节点{Fi}ddt和质心

*i*=1

*c*=1

在初始化阶段之后，由于分辨率降低会使图片丢失许多身份识别线索，因此学生流通常在低分辨率的面孔上遭受较低的识别精度。结果，需要在教师流的监督下再次对其参数进行微调，以学习如何提取最具区别性的特征，甚至

节点{Uc}*c*。第i个面节点Fi和第c个质心节点Uc分别由D维列特征向量fi和uc表示。老师模型从高分辨率的人脸图像中提取这两种类型的特征向量

可以计算为：

当脸部分辨率非常低时。

φˆt和φˆs是由教师流网络φt和学生流网络φs的前几层分别组成。

Y·DS*δ*（Li＝δ）

*（c）FI*

*φ*ˆs（F˜; Wˆs）

**F***I＝*ˆ*T（FI；W*ˆ*T），UC=*

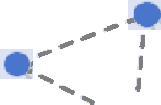
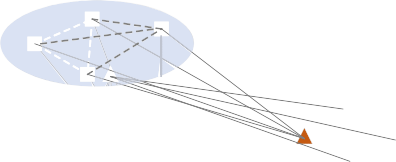
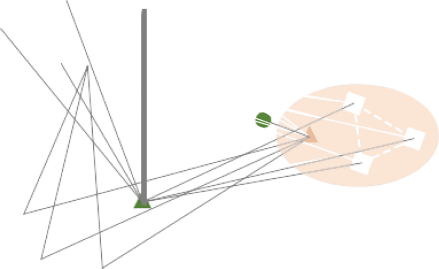
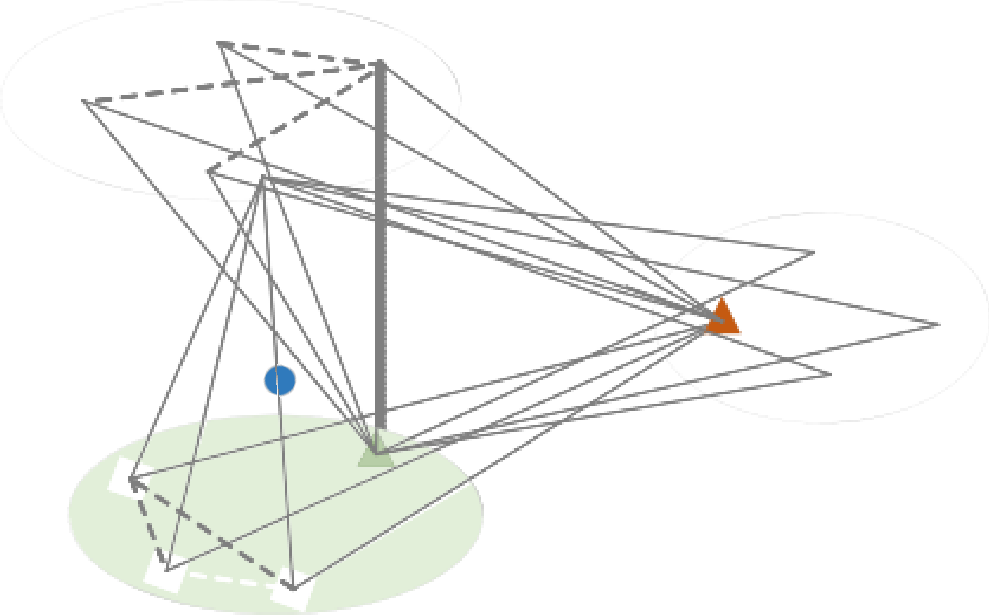
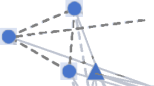
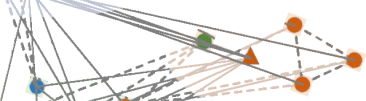
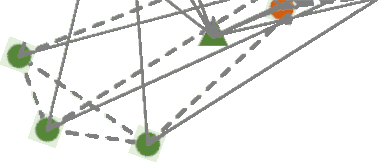
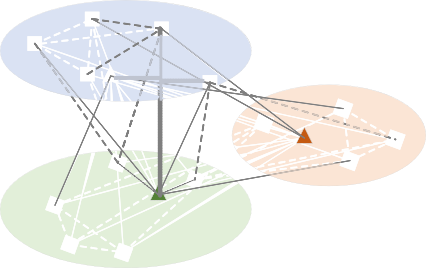
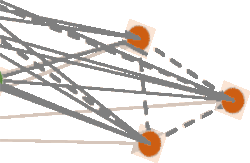
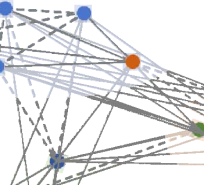
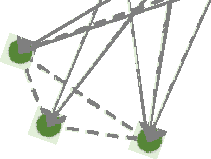
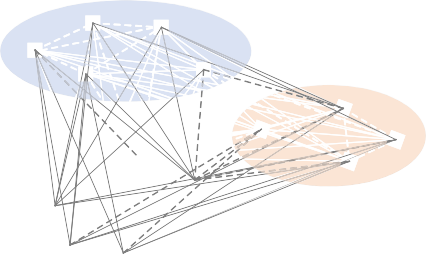
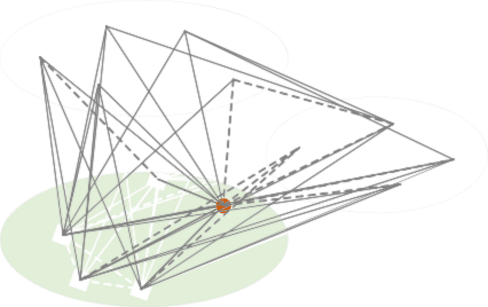
*i*=1

Y·DS*=*（li = c）



*i*=1

*,* (2)



27923; 27925; 27927;

27959; 27961; 27992; 28023;



图4.图优化的选择性知识提取。密集连接图（左）在特征节点之间的边过多，因此很难解决优化问题。为了解决这个问题，通过为每个节点类添加虚拟质心节点来构建稀疏连接图（中间）。然后，将该问题表述为可以通过能量最小化有效解决的图形标注问题（右）。

其中δ（li = c）是一个指标函数，如果li = c等于1，否则等于0。我们可以看到，每个面部节点是由特定面部的外观特征化形成的，并且

质心节点由面部平均外观表示。

借助质心节点，我们可以构造一个稀疏图，其边集E由紧密连接的类内边{（Fi，Fj）|∀i，j，li = lj}组成，这些边链接同一对象中的所有面节点，以及稀疏连接的类间边缘{（Fi，Uc）|∀i，c，li / = c}，它们将一个类中的面节点与该类之外的其余质心节点链接在一起。这样，稀疏连通的图G包含（| Ds | + C）个节点，总边数：

知识可以被准确地提取，大大减轻通过教师网络引入的噪音知识的影响，从而很好地完善了对学生网络训练的特征监督。可以通过λ控制丢弃的异常值的数量（将在实验中讨论λ的影响）。

1. *老师流指导的学生流微调*

通过教师流提取所选的信息丰富的面孔及其特征，对学生进行微调

c

*c* 1

=

[Kc（Kc-1）/ 2 + Kc（C-1）]。

将共同解决两个问题：

对于给定稀疏图，我们可以通过解决二进制标签问题来选择有用信息最多的面孔：

DSi *C* \

1、通过特征回归逼近教师流给出的含有丰富信息的脸的特征以及2）从低分辨率的面孔中恢复丢失的面部线索。因此，我们可以通过解决最小化问题来微调学生流

：

分钟 �

***α***

*i*=1

*αI* �*δ（LI/C＝C）·D（FI，UC）*

*c*=1

分

*Lcls（WS；DS）+LReG（W*ˆ*s；DS；α）；* (4)

DS

+*λ* �

DS

�

*αiαj·δ（Li＝Lj）·d（Fi，Fj）；*

宽，宽ˆs

其中分类损失和规则的影响-

*i*=1 *j* =1*,i< j*

*STαi { 0, 1 }，i＝1，…，* (3)

其中α=（α1，α2，...，αDS）是具有| Ds |的二元向量。

α=(α1α2,…,αDs)是一个二进制向量，如果第i个人脸其第i个成分αi = 1。注意，我们使用余弦距离d(·)来度量两个特征向量之间的相似性。我们可以看到，第一项倾向于选择可用信息较少的面部节点，这些节点与其他类中的“平均”面部的相似性较低。λ是负的权重，用来平衡两项,第二项prefer有很高的相似性与其他face有同样的身份标签的face node。特别是,与非负距离测量d(·)和负权重λ,第一项往往选择更少的面孔和第二项往往选择更多。在实践中，我们可以使用图割算法[58]来解决(3)问题。

分量，且如果面Fi为，则其第i个分量αi等于1。

失误同等重要地组合在一起，形成一个多任务学习问题。第一项Lcls（Ws; Ds）是学生流在所有低

分辨率的面孔。与（1）类似，它定义为

信息丰富的面孔，否则为0。注意我们使用余弦

DSN

距离d（·）以测量两个特征向量之间的相似度。我们可以看到第一个词更喜欢选择

*Lcls (Ws ; Ds） |*

� �

*i*=1 *j* =1

*（S（F*˜*IJ；WS），Li）；* (5)

与其他类别的“平均”面部相似度较低的信息量较少的面部节点。λ是一个负权重，平衡了两个项，因此第二项更喜欢

与其他人脸具有高度相似性的人脸节点

2

（4）中的术语Lreg（Wˆs; Ds;α）是由特征提取后端形成的子网φˆs（F˜; Wˆs）的特征回归损失学生流。可以定义为

相同的身份标签。特别是在非负距离度量d（·）和负权重λ的情况下，第一项倾向于选择较少的面孔，第二项倾向于选择

*LReg（Ws；DS；α）=*

DS

� *i*

*i*=1

*N*

�

*j* =1

I*φ*ˆ*s（F*˜*ij; W*ˆ*s）-fi I。* (6)

更多。在实践中，我们可以使用图割算法[58]解决问题（3）。

解出（3）后，我们可以选择数量有限的，具有较高的类内相似度和较低的类间相似度的信息面孔。在此过程中，从特征聚类的角度出发，排除了可能是教师流所产生的错误的异常值。这样，很多有用的

通过将这两项纳入（4），我们可以通过带有标准反向传播的随机梯度下降算法来解决分类和回归任务[57]。这样，可以在教师流的监督下以特征回归的形式微调学生流，从而以有限的计算成本提高了低分辨率人脸识别能力。

1. 实验

在本节中，我们首先介绍实验设置，然后进行四个实验以验证所提出的方法。进行第一个实验以分析和讨论选择性知识提炼的影响，第二个实验比较教师和学生网络在面部验证任务中的表现。在第三和第四实验中，我们分别在面部识别任务和面部检索任务中将学生流与最新的低分辨率面部模型进行比较。最后，我们对学习的学生模型进行效率分析。

1. *实验设定*

我们在四个著名的人脸数据集上进行了实验：UMDFaces [59]，LFW（野外带有标签的人脸）[60]，UCCS（不受约束的大学生）[61]和SCface（监控摄像机的人脸）[62]。分别从选择性知识提取，面部验证，面部识别和面部检索的角度来验证所提出的方法。以下列出了这三个数据集（和实验设置）的详细信息。我们使用TensorFlow [63]在NVIDIA GPU K80和单核Intel CPU 2.6G上实现所有模型。

UMDFaces数据集[59]包含367、888个图像，这些图像带有来自8,419个主题的注释，这些图像是通过在Internet上爬网公共图像而获得的。在实验中，我们使用该数据集训练所有学生模型并验证选择性蒸馏操作。对于每张训练脸，我们首先使用[64]中的算法来进行脸部对弈，以定位脸部landmarks。然后修剪脸部将其，

标准化为224×224高分辨率图像，用作教师网络的输入。同样，形成用于训练每个学生网络的低分辨率面孔，我们对定位后的人脸地标进行16次随机扰动，将人脸区域裁剪并归一化为大小为p×p的人脸图像，其中分辨率p∈{16,32,64,96}。

在UMDFaces数据集上，所有224×224人脸图像都是

首先输入教师流模型VGGFace中进行提取4096D特征向量

通过解决方程（3）中的图形优化问题，选择有用丰富信息的特征并用指示向量α表示。之后，通过选择的信息特征，面部识别标签和学生网络输入的面部图像，使用标准BP算法训练学生网络。在训练中，我们将批次大小设置为256。引入了批次归一化层，以加快网络训练并防止过度拟合。

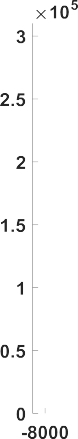
在LFW数据集[60]上，我们在人脸验证任务中评估所有学生模型。在实验中，评估使用了6000对脸部图像，包括3,000对正向图像和3,000对负向图像。性能reported as Area under ROC curve（AUC）。在实验中，首先从一对面部图像中提取隐藏层（模拟层和identity层）中的特征向量并将其标准化。使用简单的阈值计算它们之间的相似度以进行验证。与[19]不同

35797;

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

36004;

图5.非正参数λ对选择的信息脸的数量的影响。



经过训练的联合贝叶斯[65]用于面部验证，在整个实验中都使用了相似性，以直接显示出通过更好地指导学生进行训练所带来的好处。

在UCCS数据集上[61]，我们将学生模型与人脸识别任务中的最新技术进行了比较。采用1,732个带标签的身份主题的人脸，其中模糊，遮挡和照明不良的图像通常很常见。请注意，培训和测试中的identity是互相隔离的。该数据集适用于在监视条件下对更具挑战性的无约束人脸识别进行基准测试。

在SCface数据集[62]上，我们将学生模型与人脸检索任务中的最新技术进行了比较。数据集包含

使用不同质量的监视摄像机从三个距离（分别为4.2m，2.6m和1.0m）捕获了130个对象，每个对象都有一个高分辨率的正面人脸图像和多个低分辨率的图像。在实验中，随机选择50名受试者进行训练，其余80名受试者进行测试。在测试图像中，对于每个主题，使用一个高分辨率的面部图像来构建检索数据集

15张低分辨率人脸图像用于检索。结果，总共80×15 = 1的每个低分辨率面部图像都与80个高分辨率面部图像匹配。的

分别报告了具有不同距离和总集合的三个子集上的rank-1识别精度。

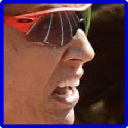
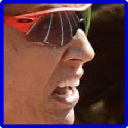
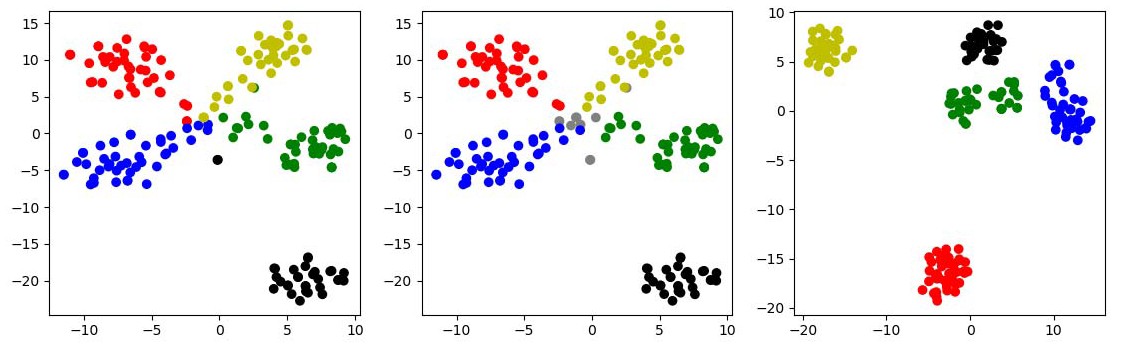
1. *选择性知识蒸馏*

为了研究选择性知识蒸馏，首先我们要探讨参数λ的不同设置对解析图优化算法的影响。因此，我们以2的整数次幂将λ从-8192逐渐增加到0，然后研究selected informative faces数量的减少趋势。在图5中，我们显示了影响

参数λ对selected informative faces的数量的影响。

当负常数λ非常小时，信息面的数量会非常缓慢地减少。λ之后

增加到-1024附近时，信息面的数量开始急剧减少。继续增加，数量继续减小，并且在λ大于-32之后保持为0。



38518;

图6.使用t-SNE可视化选择性知识精炼在丢弃嘈杂的面孔上的示例[66]。远离他们自己的类别或更接近其他类别的嘈杂的面部节点（b）被丢弃，这在再次执行t-SNE（c）时会导致更紧凑的可视化。相应的废弃面部图像（d）通常具有侧面姿势，重度遮挡，照明不一致和外观模糊，这使得它们很难被教师网络识别。

我们进一步研究了选择性知识蒸馏过程中丢弃面部的过程。图6给出了一个示例，该示例显示了当λ增大时在五个identity类中丢弃面部的过程，其中我们采用t分布随机邻居嵌入（t-SNE）[66]来可视化高维面部节点。通过解决方程4中的稀疏图优化问题。在图4中，一些noisy人脸节点（请参见图6（a）中的原始人脸节点和图6（b）中的灰色噪音人脸节点）通常远离其类质心或更接近其他类被丢弃，导致更紧凑的可视化效果（见图6（c））。在图6（d）中，我们显示了被丢弃的noizy人脸图像，在其中我们可以看到被丢弃的人脸通常以侧面姿势，重度遮挡，照明不一致或外观模糊为特征。这些具有挑战性的图像（可能超出了教师流的识别能力）被选择并丢弃。这意味着选择性知识蒸馏确实选择了信息量更大的面孔，而丢弃了信息量较少的面孔。理想情况下，教师模型应具有强大的能力来处理各种人脸变化并正确地聚类人脸，这意味着教师模型通过提取特征可以实现较高的类内相似度和较低的类间相似度。但是，在某些挑战性情况下（例如，较大的姿势变化），教师模型可能会失败，从而导致类内部相似度较低（如图6所示）。在这种情况下，提取的教师知识被认为是“错误的”，因此在特征回归任务中将被我们的方法丢弃。

1. *LFW上的低分辨率人脸验证*

通过选择的面部图像，我们以不同的输入分辨率以及各种监督信号压缩教师模型训练了许多学生模型。

。我们研究的监督信号缩写为：

* 1. *c*：仅有脸部类的监督（无蒸馏）。
  2. *s*：无面类监督的选择性蒸馏
  3. *s*c：选择性蒸馏，具有面类监督
  4. *d*c：直接蒸馏，具有面类监督

为了简单起见，将学生模型表示为S-p-s，其中监督信号s∈{c，s，sc，dc}。例如，模型S-32-sc表示学生模型使用32×32的输入分辨率，并经过选择性蒸馏和类监督进行训练。请注意，S-pc是在脸部类的监督下直接训练的基线学生模型。同样，教师模型表示为具有224×224作为输入的T-224。所有学生模型都采用图3所示的相同架构进行训练。

各种学生模型的性能如图7所示，从中我们可以看出，识别精度随着面部分辨率的降低而降低。学生模型S-96-sc通过使用选择性知识提炼和人脸班级监督达到了95.03％的准确度，仅比没有metric学习的教师模型T-224低2.12％。请注意，S-96-sc中的模型参数比教师模型VGGFace小得多（0.79M对138M），并且面部特征向量的维数

压缩率高达32倍。根据这些结果，我们可以放心地说，这种性能非常有竞争力-

特别适合在资源受限的设备上进行实际部署。

从图7中，我们还可以发现，在没有老师指导的情况下，基线学生模型S-32-c的准确率非常低，仅为70.23％，这意味着，在低分辨率人脸上直接训练时，基线模型本身可能缺乏提取出具有判别性特征的能力。



42907; 42967;

图7. LFW上各种师生模型的识别准确性。结果表明，选择性知识提炼和人脸分类监督可以帮助识别低分辨率人脸。



在接受来自教师流和面部标签的联合监督信号训练后，S-32-sc模型的识别准确度大幅提高了19.49％（即，从70.23％提高到89.72％）。在S-16-sc和S-16-c以及S-16-dc和S-16-c之间可以发现类似的精度提高，这意味着选择性或直接的知识提炼可以有效地将教师的知识转移到学生中网络从而显着提高识别性能。

为了进一步验证知识选择的重要性，我们比较了S-16-sc和S-16-dc之间的性能。通过仔细选择有用的知识，S-16-sc相对于S-16-dc可获得2.04％的准确度增益，而S-16-dc在训练过程中不会丢弃嘈杂的面部。另外，人脸分类监控信号也可以提高性能，因此，S-16-sc模型比S-16-s具有更高的精度。总之，我们的两流结构可以准确地从教师流中提取有用的信息性知识，并从学生流中恢复丢失的知识。

1. *UCCS上的低分辨率人脸识别*

由于低分辨率人脸验证任务的性能是有前途的，因此我们将在具有挑战性的基准UCCS [61]上进一步研究低分辨率人脸识别任务，并与在[36]中提出的最新方法进行比较

VLRR（非常低分辨率的识别）。

在VLRR中，将裁剪后的面部区域归一化为80×80的高分辨率面部，然后将其下采样5倍为16×16的低分辨率图像。对180个对象的子集逐层进行评估-逐层贪婪

无监督模型训练。他们的模型报告最佳错误率在top1为40.97％，在top5为22.35％。

按照[36]的实验设置，我们通过排名选择了180个原始分辨率图像的子集

根据图像数量拍摄对象。然后将裁剪的脸部区域标准化为16×16，以获得4,825张图像。请注意，这个数字比VLRR声称的数字小一些（即4,500个训练图像和935个测试图像）。之后，为了达到公平的比较，我们随机

按照4：1的比例将图像与训练和测试集分开。最后，我们有3，918张图像用于训练，其余907张用于测试。

图8.各种模型的推理内存占用量。与原来的教师模型相比，可以大大减少学生模型的存储成本。

根据这些数据，我们首先在UCCS的训练集以16×16作为输入直接训练学生模型，然后在其测试集上测试性能。在这种情况下，该模型实现了58.65％的top-1错误率和22.71％的top-5错误率，这比VLRR差。我们怀疑这些模型一旦

如[67]所述，在其他数据集上进行训练后，可以提供有关低分辨率视觉识别问题的有价值的先验知识。为了验证这一点，我们使用在UMDFaces上预先训练的学生模型S-16-sc来微调用于UCCS上人脸识别的新模型。首先，我们在模仿图层之前固定参数，然后将最后一个softmax图层修改为180度。然后，我们用3张918张图像训练特征约简子网。经过微调的模型达到了32.75％的top-1错误率和18.3％的top-5错误率，表明top-1结果中907个测试样本中的610个正确分类，top-5中741个样本中的741个正确分类。这意味着我们的方法比VLRR可以获得更高的准确性，这可能是由于以下事实造成的：即使在脸部分辨率非常低的情况下，来自教师流的选择性监督也可以帮助学生网络学习区分特征。

1. *在SCface上进行低分辨率人脸检索*

我们在SCface [62]上进一步研究了低分辨率人脸检索任务，并与基线（PCA [62]）和4个最新技术进行了比较，其中包括三个基于嵌入的模型（DCA [31]，DAlign [ [68]和LRFRW [70]）和一种基于幻觉的模型（SHSR [69]）。在这里，LRRFW采用深度学习来执行跨域传输。结果显示在表II中，其中分别报告了具有不同距离和总集合的三个子集的精度。在实验中，我们在SCface训练集上微调了两个学生模型，分别是具有默认128D身份特征的S-64-sc和具有1024D身份特征的S-64-sc-1024。

从结果中，我们可以看到一些重要的观察结果。首先，使用基线PCA模型，所有模型的准确度均低于50％，特别是总准确度极低，为4.73％，显示了在

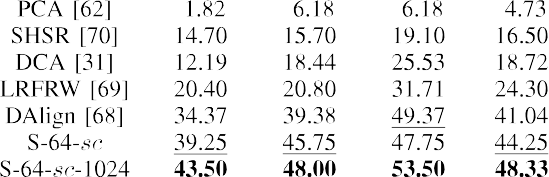
表二

scface上的等级1识别精度（％）。粗体和

下划线表示第一和第二

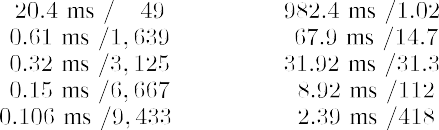
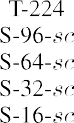
分别

48252;



表三

推断gpu和cpu的时间和速度



SCface是一项非常具有挑战性的任务。其次，随着分辨率随着距离的增加而增大，识别精度逐渐提高，这与预期的一样，这意味着分辨率的确对识别性能产生了重要影响。因此，基于幻觉的SHSR超分辨率低分辨率人脸图像，以馈送到预训练的人脸识别器，从而将总精度提高到16.50％。第三，在不同领域之间传递知识的嵌入模型，例如，通过DCA模型中的判别相关分析和LRFRW模型中的监督判别交叉分辨率学习，将特征从高分辨率的脸部转移到低分辨率的脸部，从近红外传递知识通过DAlign模型中的字典对齐方式对可见图像进行处理，可将总精度分别提高18.72％，24.30％和41.04％。这揭示了从其他域转移的知识的影响。最后，我们的两个模型比其他模型具有更高的总精度，例如，S-64-sc的总准确度比DAlign提高了3.21％，这意味着来自教师流的选择性知识可以促进学生网络的发展。

1. *效率分析*

我们的方法可以大大减少模型参数的数量和内存占用量，而不会显着提高准确性

下降。如图8所示，低分辨率学生模型的内存减少量为22倍，48倍，182倍和620倍。

分别为96×96、64×64、32×32和16×16。

特别是对于分辨率很低的16×16的人脸，推理存储空间仅为0.15MB。

此外，教师模型VGGFace（T-224）包含

1.38亿个参数，而学生网络只有约79万个参数，大大减少了

模型复杂度提高了175倍，但代价是识别精度的下降幅度很小。由于极大地减少了存储成本和参数数量，因此计算

复杂度可以大大降低。如表III所示，高端GPU和低端CPU上的推理运行时间大大减少。使用NVIDIA K80 GPU，人脸的推理时间从使用T-224（VGGFace老师）的20.4ms减少到使用S-96-sc，S-64-sc，0.61ms，0.32ms，0.15ms和0.106ms的时间， S-32-sc和S-16-sc。即使在CPU中，推理时间也显着减少。我们的模型需要0.106毫秒和2.39毫秒才能识别出具有非常低的

GPU和CPU的分辨率分别为16×16，即每秒9张，433张面孔和每秒418张面孔。

1. 结论

目前，基于深度学习的人脸识别模型普遍存在模型参数大，特征尺寸高的问题，这阻碍了它们在资源受限的应用程序（例如嵌入式或移动设备）上的实际部署。针对这一问题，本文提出了一种知识提炼的方法，采用原来的大型模型作为教师网络，通过设计回归和分类项相结合的多任务损失函数，让教师有选择地监督学生网络的训练。我们已经完成了高维深度特征回归和低分辨率人脸分类的结合，从而在保证识别准确率的情况下实现了深度网络和特征维的均匀压缩。实验结果表明，该方法可以将信息知识从教师网络转移到学生模型，从而形成紧凑的人脸识别模型，具有显着的效果和效率。

在未来的工作中，我们将尝试探索循环机制的用途，该机制旨在处理教师流中的失败案例。诸如性别，年龄和化妆等面部属性将被合并到多任务框架中，以进一步增强压缩模型的性能。

参考资料

[1] Y. Taigman，M。Yang，M。Ranzato和L. Wolf，“ DeepFace：弥合人脸验证中与人类水平性能的差距”，Proc。IEEE会议计算可见模式识别。（CVPR），2014年6月，第1701-1708页。

[2] Y. Sun，X。Wang和X. Tang，“通过预测10,000个班级进行深度学习人脸表示”，Proc。IEEE会议计算可见模式识别。（CVPR），2014年6月，第1891-1898页。

[3] F. Schroff，D。Kalenichenko和J. Philbin，Proc。中的“ FaceNet：用于面部识别和聚类的统一嵌入”。IEEE会议计算可见模式识别。（CVPR），2015年6月，第815-823页。

[4] OM Parkhi，A。Vedaldi和A. Zisserman，Proc中的“深脸识别”。英国马赫可见Conf。（BMVC），2015年，第1期。1号3，第6。

[5] B. Amos，B。Ludwiczuk和M. Satyanarayanan，“ OpenFace：具有移动应用程序的通用人脸识别库”，卡内基梅隆学校计算机。科学，匹兹堡，PA，美国，2016。

[6] A. Pentland和T. Choudhury，“智能环境中的人脸识别”，计算机，第一卷。33，没有2，第50-55页，2000年2月。

[7]刘D.程B.王Z.王H.张，和TS黄。（2017）。“通过深层网络在不利条件下增强视觉识别。” [在线]。可用：https://arxiv.org/abs/1712.07732

[8] S. Kolouri和GK Rohde，“过程非常低分辨率的人脸图像基于传输的单帧超分辨率”，Proc。IEEE会议计算可见模式识别。（CVPR），2015年6月，第4876–4884页。

[9]健先生和K.-M。Lam，“基于奇异值分解的幻觉和低分辨率面孔的同时识别”，IEEE Trans。电路系统视频技术。25号2015年11月，第11页，第1761–1772页。

[10] M.-C.杨建平魏永瑞Yeh和Y.-CF Wang，《过程识别》，《远距离识别：极低分辨率的面部识别和幻觉》。诠释Conf。生物识别（ICB），2015年5月，第237-242页。

[11] S. Biswas，KW Bowyer和PJ Flynn，“用于匹配低分辨率面部图像的多维缩放”，IEEE Trans。模式肛门。马赫智力，卷。34号第10页，2019-2030年，2012年10月。

[12] C.-X.任大QDai和H. Yan，“用于低分辨率面部图像识别的耦合内核嵌入”，IEEE Trans。图像处理。21号，第8卷，第3770–3783页，2012年8月。

[13] T．欧元。Conf。计算可见（ECCV）。德国柏林：施普林格，2004年，第469-481页。

[14] SJD Prince和JH Elder，Proc中的“关于身份推断的概率线性判别分析”。IEEE国际Conf。计算可见（ICCV），2007年10月，第1-8页。

[15] Y. Wen，K K. Zhang，Z。Li和Q. Qiao，“区分面部深度识别的特征学习方法”，Proc。欧元。Conf。计算可见（ECCV），2016年，第499-515页。

[16] AT Tran，T。Hassner，I。Masi和G. Medioni，“使用非常深的神经网络回归鲁棒且具有判别力的3D变形模型”，Proc。IEEE会议计算可见模式识别。（CVPR），2017年6月，第1493–1502页。

[17] X. Zhang，Z。Fang，Y。Wen，Z。Li，和Y. Qiao，“使用长尾巴进行深脸识别的距离损失”，Proc。IEEE国际Conf。计算可见（ICCV），2017年6月，第5409–5418页。

[18] G. Hu等人，“采用神经张量融合网络的属性增强人脸识别”，Proc。IEEE国际Conf。计算可见（ICCV），2017年6月，第3764–3773页。

[19] Sun. Y，Chen，X. Wang和X. Tang，“通过联合识别验证进行深度学习人脸表示”，Proc。进阶神经信息处理。系统，卷。27. 2014，pp。1988-1996。

[20]孙Y，梁D.王X.王和唐X。（2015）。“ DeepID3：具有非常深的神经网络的人脸识别。” [在线]。可用：https://arxiv.org/abs/1502.00873

[21] Q. Cao，L。Shen，W。Xie，OM Parkhi和A. Zisserman，“ VGGFace2：用于识别跨姿势和年龄的脸部的数据集”，Proc。IEEE国际Conf。自动面部手势识别。（FG），2018年，第67-74页。

[22] T. Uiboupin，P。Rasti，G。Anbarjafari和H. Demirel，“采用稀疏表示的面部图像超分辨率，可以改善监视监控中的人脸识别”，Proc。IEEE会议信号处理。公社Appl。，2016年5月，第437-440页。

[23] C. Dong，CC Loy，K。He和X. Tang，“学习深度卷积网络以实现图像超分辨率”，Proc。欧元。Conf。计算可见（ECCV）。德国柏林：施普林格，2014年，第184-199页。

[24] C. Ledig等人，“采用生成对抗网络的逼真的单图像超分辨率”，Proc。IEEE会议计算可见模式识别。（CVPR），2017年6月，第105-114页。

[25] WWW Zou和PC Yuen，“非常低分辨率的面部识别问题”，IEEE Trans。图像处理。21号1，第327-340页，2012年1月。

[26] P. Zhang，X。Ben，W。Jiang，R。Yan和Y. Zhang，“低分辨率人脸识别的耦合边际判别映射”，Opt.-Int。J.光学电子，第一卷。126号2015年12月，第23页，第4352–4357页。

[27] J. Jiang，R。Hu，Z。Wang和Z. Cai，“ CDMMA：用于区分低分辨率人脸图像的耦合判别多流形分析”，《信号处理》，第1卷。124，第162-172页，2016年。

[28] X. Wang，H。Hu和J. Gu，“通过基于核的增强型判别分析进行姿态鲁棒的低分辨率人脸识别”，IEEE / CAA J. Autom。Sinica，第一卷3号2，第203–212页，2016年4月。

[29] X. Xing和K. Wang，“具有二分图嵌入的耦合流形判别分析，用于低分辨率人脸识别”，《信号处理》，第1卷。125，第329–335页，2016年8月。

[30] J. Shi和C. Qi，“从局部几何到整体结构：学习用于低分辨率人脸识别的潜在子空间”，IEEE信号处理。Lett。，第一卷22号5，第554–558页，2015年5月。

[31] M. Haghighat和M. Abdel-Mottaleb，“使用判别相关分析的监视系统中的低分辨率人脸识别”，Proc。IEEE国际Conf。自动面部手势识别。（FG），5月/ 6月。2017年，第912–917页。

[32] X. Li，W.-S.。Zheng，X。Wang，T。Xiang，和S. Gong，“多尺度学习用于低分辨率人的重新识别”，Proc。IEEE国际Conf。计算可见（ICCV），2015年12月，第3765–3773页。

[33] K.-H.庞和K.-M.Lam，“用于面部识别的多分辨率特征融合”，Pattern Recognit。，第一卷。47，没有2，第556–567页，2014年2月。

[34] SP Mudunuri和S. Biswas，“姿势和照明变化中的低分辨率人脸识别”，IEEE Trans。模式肛门。马赫智力，卷。38号5，第1034-1040页，2016年5月。

[35] S. Shekhar，VM Patel和R. Chellappa。（2017）。“基于合成的鲁棒低分辨率人脸识别。” [在线]。可用：https://arxiv.org/abs/1707.02733

[36] Z. Wang，S。Chang，Y。Yang，D。Liu和TS Huang，“研究使用深度网络的超低分辨率识别”，Proc.Natl.Acad.Sci.USA，87：3877-5。IEEE会议计算可见模式识别。（CVPR），2016年6月，第4792-4800页。

[37] C. Herrmann，D。Willersinn和J. Beyerer，“用于视频人脸识别的低分辨率卷积神经网络”，Proc。IEEE国际Conf。进阶基于视频信号的监视。（AVSS），2016年8月，第221–227页。

[38] C. Buciluaˇ，R。Caruana和A. Niculescu-Mizil，“模型压缩”，Proc。ACM Conf。知道发现数据挖掘（KDD），2006年，第535-541页。

[39] G. Hinton，J。Dean和O. Vinyals，“在神经网络中蒸馏知识”，载于Proc。神经信息处理。Syst。（NIPS）研讨会，2014年，第1–9页。

[40] A. Romero，N。Ballas，SE Kahou，A。Chassang，C。Gatta和

Y. Bengio，Proc中的“ FitNets：细深网的提示”。诠释Conf。学习。代表。（ICLR），2015年，第1-13页。

[41] T. Chen，I。Goodfellow和J. Shlens，Proc中的“ Net2Net：通过知识转移促进学习”。诠释Conf。学习。代表。（ICLR），2016年。

[42] P. Luo，Z。Zhu，Z。Liu，X。Wang，和X. Tang，“通过从神经元中提取知识来进行人脸模型压缩”，Proc.Natl.Acad.Sci.USA，87：3877-5。AAAI Conf。Artif。智力（AAAI），2016年，第3560–3566页。

[43] Z. Li和D. Hoiem，“学习不忘”，Proc。欧元。Conf。

*计算可见（ECCV），2016年，第614–629页。*

[44] Y. Kim和AM Rush，“序列级知识提炼”，Proc。Conf。经验方法自然郎。处理。（EMNLP），2016年，第1317-1327页。

[45] G.Urban等人，“ Proc中的“深卷积网真的需要深层和卷积吗？”。诠释Conf。学习。代表。（ICLR），2017年，第1-13页。

[46] G. Chen，W。Choi，X。Yu，T。Han，和M. Chandraker，Proc.Natl.Acad.Sci.USA，“利用知识蒸馏学习有效的物体检测模型”。神经信息处理。Syst。（NIPS），2017年，第742–751页。

[47] Y. Chen，N。Wang和Z. Zhang，“ DarkRank：通过交叉样本相似性转移加速深度度量学习”，Proc.Natl.Acad.Sci.USA，87：3877-5877。AAAI Conf。Artif。智力（AAAI），2018年，第2852–2859页。

[48] T. Chen，L。Lin，W。Zuo，X。Luo，和L. Zhang，“学习小波状自动编码器以加速深度神经网络”，Proc。AAAI Conf。Artif。智力（AAAI），2018年，第6722–6729页。

[49] G. Zhou，Fan Y.，Fan R. Cui，W。Bian，X。Zhu和G. Kun，“火箭发射：训练性能良好的光网络的通用有效框架”，Proc。AAAI Conf。Artif。智力（AAAI），2018年，第1-8页。

[50] D. Lopezpaz，L。Bottou，B。Scholkopf和V. Vapnik，“统一蒸馏和特权信息”，载于Proc。诠释Conf。学习。代表。（ICLR），2016年，第1-10页。

[51] J.-C.Su and S. Maji，Proc.Natl.Acad.Sci.USA，“适应模型以利用蒸馏来降解信号”，Proc。英国马赫可见Conf。（BMVC），2017年，第1-14页。

[52] I. Radosavovic，P。Dollár，R。Girshick，G。Gkioxari和K. He，《数据精炼：迈向全监督学习》，载于Proc。IEEE会议计算可见模式识别。（CVPR），2018年6月，第4119–4128页。

[53] K. Simonyan和A. Zisserman，“非常深入的卷积网络，用于大规模图像识别”，Proc。诠释Conf。学习。代表。（ICLR），2015年，第1-14页。

[54] C. Szegedy等人，“ Proc。Theatre of Proc。Physics and Proc。”中的“卷积更深”。IEEE会议计算可见模式识别。（CVPR），2015年6月，第1–9页。

[55] K. He，X X. Zhang，S。Ren和J. Sun，“深度残差学习以进行图像识别”，Proc.Natl.Acad.Sci.USA，87：3877-5。IEEE会议计算可见模式识别。（CVPR），2016年6月，第770–778页。

[56] J. Redmon和A. Farhadi，《过程》中的“ YOLO9000：更好，更快，更强大”。IEEE会议计算可见模式识别。（CVPR），2017年6月，第57263–7271。

[57] A. Krizhevsky，I。Sutskever和GE Hinton，“具有深度卷积神经网络的ImageNet分类”，Proc。神经信息处理。Syst。（NIPS），2012年，第1097-1105页。

[58] Y. Boykov和V. Kolmogorov，“最小切割/最大流量算法用于视觉能量最小化的实验比较”，IEEE Trans。模式肛门。马赫智力，卷。26号9，第1124-1137页，2004年9月。

[59] A. Bansal，A。Nanduri，CD Castillo，R。Ranjan和R. Chellappa，“ UMDFaces：用于训练深层网络的带注释的面部数据集”，Proc。IEEE国际联合会议生物识别（IJCB），2017年10月，第464–473页。

[60] E. Learned-Miller，GB Huang，A。Roychowdhury，H。Li和G. Hua，“人脸检测和面部图像分析的进展”中的“被标记的脸在野外：一项调查”。瑞士湛（Cham）：施普林格（Springer），2016年，第189–248页。

[61] A. Sapkota和TE Boult，Proc中的“大型无约束开放式人脸数据库”。IEEE 6th国际Conf。生物识别，理论，应用。Syst。，9月/ 10月。2014年，第1-8页。

[62] M. Grgic，K。Delac和S. Grgic，“ SCface-监控摄像头人脸数据库”，多媒体工具应用，第1卷。51号3，第863–879页，2011年。

[63] M. Abadi等人，Proc.Natl.Acad.Sci.USA，“ TensorFlow：用于大规模机器学习的系统”。USENIX症状。操作系统设计实现。（OSDI），第1卷。16. 2016年，第265-283页。

[64] S. Ren，X。Cao，Y。Wei和J. Sun，“通过回归本地二进制特征进行人脸对齐”，IEEE Trans。图像处理。25号3，第1233至1245页，2016年3月。

[65] D. Chen，X。Cao，L。Wang，F。Wen和J. Sun，“重新审视贝叶斯面孔：联合表述”，载于Proc。欧元。Conf。计算可见（ECCV），2012年，第566–579页。

[66] L. van der Maaten和G. Hinton，“使用t-SNE可视化高维数据”，J。Mach。学习。水库卷9，第2579–2605页，2008年11月。

[67] B. Cheng，D。Liu，Z。Wang，H。Zhang和TS Huang，“ Proc.Natl.Acad.Sci.USA，Proc.Natl.Acad.Sci.USA，Proc.Natl.Acad.Sci.USA 90：5873-5877：在非常低质量的环境中的视觉识别：深入研究预训练的力量”。AAAI Conf。Artif。智力（AAAI），2018年，第8065–8066页。

[68] SP Mudunuri，S。Venkataramanan和S. Biswas，“字典排序与重新排序以实现低分辨率NIR-VIS人脸识别”，IEEE Trans。Inf。取证安全，第一卷。14号4，pp.886–896，2019年4月。

[69] M. Singh，S。Nagpal，M。Vatsa，R。Singh和A. Majumdar，“用于交叉分辨率面部识别的身份识别合成”，Proc.Natl.Acad.Sci.USA，87：3877-5877。IEEE会议计算可见模式识别。讲习班（CVPRW），2018年7月，第479–488页。

[70] P. Li，L。Prieto，D。Mery和P. Flynn。（2018）。“低分辨率的野外面部识别。” [在线]。可用：https：// arxiv。组织/abs/1805.11529

**葛**世明（M'13–SM'15）获得理学学士和博士学位。中国科学技术大学分别于2003年和2008年获得电子工程学位。他曾是盛大创新公司的高级研究员，三星电子和诺基亚研究中心的研究员。他目前是中国科学院信息工程研究所的副教授。他的研究主要集中于计算机视觉，深度学习和AI安全性，尤其是针对可扩展应用程序的高性能深度模型。



**赵**胜伟于2017年获得武汉大学数学与统计学院的学士学位。他目前正在中国科学院信息工程研究所和中国科学院大学网络安全学院攻读硕士学位。科学。他的主要研究兴趣是深度学习和计算机视觉。

**李**晨雨拥有同济大学电子与信息工程学院的学士学位。她目前正在攻读博士学位。拥有中国科学院信息工程研究所和中国科学院大学网络安全学院的博士学位。她的研究兴趣是计算机视觉和深度学习。



**贾**莉（M'12–SM'15）于2005年获得清华大学的学士学位，并获得博士学位。2011年获得中国科学院计算技术研究所的博士学位。他曾在南洋理工大学，北京大学和盛大创新公司任职。他目前是北京航空航天大学计算机科学与工程学院的副教授。他在TPAMI，TIP，IJCV，ICCV和CVPR等参考期刊和会议上撰写或合着了50多篇技术文章。



他的主要研究兴趣是对可演化算法和模型的认知视野。他的研究兴趣包括计算机视觉和多媒体大数据，尤其是基于深度学习的视觉内容理解。