改进的知识蒸馏技术，可训练快速低分辨率人脸

识别模型

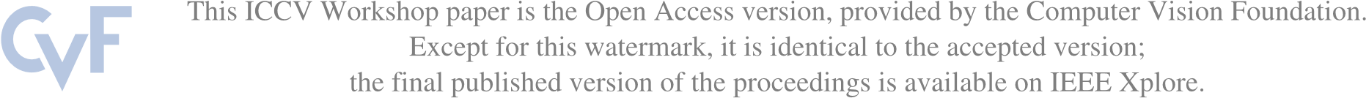
王梦娇[[1]](#footnote-0)[1]，刘如杰富士通研究开发中心有限公司，中国北京[[2]](#footnote-1)

{ wangmengjiao，rjliu } @ cn.fujitsu.com

Nada Hajime，Narishige Abe，内田秀久（Hidetsugu Uchida），Tomoaki Tomaki Matsunami Fujitsu Laboratories Ltd.日本川崎

{ nada.hajime，abe.narishige，u.hidetsugu，t.matsunami } @ fujitsu.com

# 摘要

*低分辨率（LR）人脸识别（FR）对于FR任务是一个具有挑战性但又常见的问题，尤其是对于监视场景。此处解决的问题不仅仅是建立LR-FR模型，更重要的是使它快速运行。在这里，知识蒸馏方法用于我们的任务，通过指导学生的训练过程，可以将教师的知识“蒸馏”成小的学生模型。对于LRFR任务，原始知识蒸馏方案将首先通过使用LR增强训练集对其进行调整来更新教师的权重，然后在更新的教师指导下使用同一训练集来训练学生模型。这种方法的问题是大型教师模型的权重调整非常耗时，特别是对于大型数据集。在本文中，我们提出了一种改进的方案，使我们能够避免教师模型的再训练，并且仍然能够为LR-FR任务训练小型模型。在这里，与原始方案不同，教师和学生模型的训练集变得不同，其中教师模型的训练集保持不变，而一个学生是LR增强的。因此，由于训练组不变，因此不再需要更新教师模型。只有小学生模型需要在原始教师模型的指导下进行训练。这可以加快整个训练过程，特别是对于大型数据集。针对教师和学生的不同训练设置将增加数据分布的差异。为了解决这个问题，我们限制了输出之间的多核最大平均差异以减小这种影响。实验结果表明，我们的方法可以将训练过程加快约5倍，同时保持准确性。就LFW和SCFace而言，我们的学生模型具有最先进的准确性。它可以实现与教师模型相比，加速度为* 3倍*，并且在CPU上运行仅需35ms。*

# 简介

对于人脸识别（FR）任务，特别是对于监视视频，低分辨率（LR）是一个常见问题。在此任务中，探头面部图像通常为低分辨率（LR），而画廊图像通常为高分辨率（HR）。我们在此解决的问题是如何训练快速LR-FR模型。在各种模型压缩技术中[1、2、3、4、5、6]，我们采用了知识蒸馏（也称为师范训练）[5、6]。这样做的目的是在训练过程中强制小型学生模型的输出与大型教师模型的输出相同。该方案可以用公式（1）表示，其中*S*代表学生模型，*T*代表教师模型，*x*代表训练样本。

*S 输出*（*x*）= *T 输出*（*x*） （1）

由于缺乏来自监视场景的LR训练数据，常见的方法是通过模拟LR图像来增强现有的训练集。如果对LR-FR任务采用原始的知识蒸馏方案，则通常包括两个步骤：（1）使用LR增强训练集更新教师模型的权重；（2）在教师模型的指导下，使用同一组训练集来训练学生模型。这两个步骤如图1.A所示。该方案的问题是大型教师模型的参数调整很复杂，这使整个过程很耗时，尤其是对于大型训练集。另外，还有其他情况--如果教师模型是由某些第三方提供的，则意味着我们没有有关训练细节的任何信息。这可能会使教师模型的再训练更加困难，有时甚至无法重现。

我们的目标是研究如何避免麻烦的教师模型再训练，以及如何仍然能够为LR-FR任务训练快速的学生模型。我们提出了一种改进的知识蒸馏方案，其中教师模型仍然使用原始训练集，而学生使用LR增强训练集。由于教师训练集不变，因此无需更新教师模型。我们只需要执行一个步骤，即在原始教师模型的指导下训练快速学生模型。图1.B说明了该方案。我们的改进方案可以更快地为LRFR任务建立小型学生模型，因为轻量级学生的训练要简单得多。该方案可以公式化为方程式（2），其中*S*，*T*，*x*和与方程（1）具有相同的含义，而∆ 表示用于数据增强的LR方差。

*S 输出*（*x* + ∆）= *T 输出*（*x*） （2）

改进方案带来的问题是，由于教师模型和学生模型使用不同的训练集，因此两个训练集之间的分布差异可能会损害准确性。因此，我们采用域相似性度量多核最大平均差异（MK-MMD）[7，8]作为我们的损失函数，以减少域差异并提高性能。

我们在本文中的贡献可以概括为两个方面：

1. 提出了一种改进的知识蒸馏方案，以加快快速LR-FR模型训练的全过程。在我们的方案中，只有学生的训练集是LR增强的，而教师模型的输入保持不变。这种方法可以避免教师模型耗时的权重更新，从而极大地加快了训练过程。
2. 约束MK-MMD度量以减少由不同的教师和学生训练集引入的分布差异。

# 方法

在这一部分中，我们首先将介绍有关原始知识蒸馏方案的详细信息，其中将详细说明损失函数的选择，学生模型网络结构和有用的技巧。然后介绍了改进的知识蒸馏方案，包括有关LR模拟训练集准备，MK-MMD损失整合等方面的详细信息。

## 原始知识蒸馏方案

在知识蒸馏方案中，小的学生模型的输出被迫与教师模型的输出相等。这样，我们可以将教师模型的表示能力转移到学生模型。在这里，我们使用不同损失函数的组合来实现这个想法。损失函数包括：软损失；硬损失和 L2丢失。

让我们表示最终得分输出为*Ž*，教师模型*Ť的*软标签可以被定义为*X τ Ť* = *SOFTMAX*（*ż Ť /τ*）其中*τ*是温度参数。类似地，对于学生网络S的软标签是*X τ s* = *SOFTMAX*（*ż S /τ*）。软损失是*X τ Ť*和*X τ 小*之间的交叉熵*号*：

*大号软* = *ħ*（*X τT ，X τS*） （3）

硬损失是未软化类概率*X S*和ground truth *y*之间的交叉熵：

*L 硬* = *H*（*X S，y*） （4）

在此，*H*（*。*）表示交叉熵。

对于提示学习，我们使用特征层作为提示来训练学生模型。提示损失实际上是特征L2损失：

 （5）

*F S*和*F T*是学生和教师模型的特征。

在这三种损失中，软损失可以通过使用软标签将知识从教师转移到学生，硬损失可以使学生发展自己的分类能力，而提示学习则可以提高学习成绩并加速收敛[6]。 ]。这里，在计算L2损失之前将特征标准化。根据[9]，这将提高整体性能。损失组合可表示为公式（6），其中

*L特征规范*表示L2损失归一化特征 *λ* 1 ，*λ* 2 ，和*λ* 3 表示用于硬，软和归一化特征损失的权重。

*大号总体* = *λ* 1 *大号硬* + *λ* 2 *大号软* + *λ* 3 *大号特征范数*（6）

在本文中，我们的教师模型是64层ResNet [10]模型。为了选择用于学生训练的网络体系结构，可以使用几种轻量级的网络体系结构，包括SqueezeNet [11]，MobileNet [12]和ShuffleNet [13]；最新的体系结构模型，包括DenseNet [14]；以及Inception-ResNet [15]，以及更薄/更深的模型。在[9]中，得出的结论是，当学生网络具有类似的网络体系结构但通道较少和/或深度较大时，更薄/更深的模型将产生最佳性能。在我们的工作中，我们采用了36层更薄的ResNet网络作为学生模型。下一部分将演示详细的网络体系结构。

|  |
| --- |
| 图1.用于训练快速低分辨率（LR）人脸识别（FR）模型的原始和改进的知识蒸馏方案之间的比较。原始方案显示为1.A，其中教师和学生模型的训练数据集必须相同。对于使用LR FR模型的训练，原始方案包括两个步骤：第一步骤：使用LR增强训练数据集更新教师；第二步：用相同的LR增强训练集训练学生。我们改进的方案如1.B所示，其中避免了教师模型的更新。它只有一个步骤：用LR增强数据训练学生，而教师模型的训练数据集不变。 |

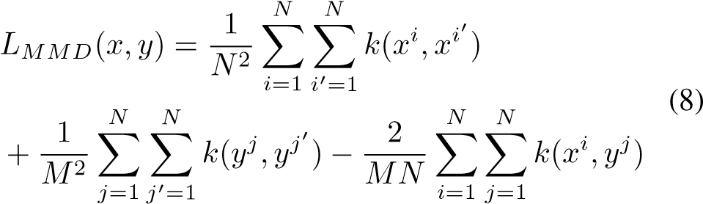
## LR-FR任务的改进知识蒸馏方案

为了使用LR方差来增强训练集，我们可以将高斯模糊添加到原始训练集。高斯模糊可以通过将每个像素与高斯滤波器卷积来应用。高斯核的大小决定了降采样比例。为了与其他作品进行公平比较，将CASIA Webface数据集[16]用作训练集。MTCNN [17]用于检测面部区域并将其对齐为112 × 96图像。在这些人脸图像中，随机选择训练集的40％进行缩小作为LR样本。LR样本的大小包括8 × 8、12 × 12、16 × 16、20 × 20和30 × 30。

在我们改进的知识蒸馏方案中，用于教师模型的训练集不变，而用于学生模型的训练集为LR增强。此方案可能会增加教师和学生的训练集之间的分布差异。在我们的工作中，我们采用多核最大平均差异（MK-MMD）[7，18，8]作为我们的损失函数，以减少数据集差异。MMD被广泛用作测量两个域之间差异的分布距离。它比较了复制内核希尔伯特空间（RKHS）[18]中的分布。MMD的公式可以表示为：



在等式（7）中，*φ*（*。*）是显式映射函数。*x i*和*y j*代表两个样本，分别从师生训练数据集中得出。通过扩展方程式（7），该方程式可重新公式化为：



从等式（8），我们可以看到MMD使用核方法*k*（*。，。*）将样本向量投影到更高维。在这里，我们使用高斯核*k*（*x，y*）=

。所述*σ* 2被设置为对的平方距离的平均值。在MK-MMD中，我们通过分别设置高斯分布方差[5]来考虑五个高斯核。

为了将MK-MMD指标集成到整体损失函数中，我们替换了在方程（6）归一化特征L2损失L*特征范数*

，用与归一化的特征MK-MMD。在此，等式（8）中的*x*和*y*分别表示由教师模型和学生模型提取的特征。在我们的方法中，代替原始*x*和*y*，将等式（8）中使用的特征被标准化为和，以使它们具有相同的大小。我们也尝试将MK-MMD应用于软标签，但会稍微影响准确性。因此，我们仅将MK-MMD用于归一化特征。新的总损失函数可以用公式（9）表示，其中 *L MK Feature* - *MMD Norm* 表示

用于归一化的特征的MK-MMD。*λ* ˉ 1，*λ* ˉ 2，和*λ* ˉ 3表示用于硬，软，和归一化的特征MK-MMD损失的权重。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | 转换1.x | Conv2.x | Conv3.x | Conv4.x | 足球俱乐部 | |  |  |  |  | 512 |   表1.教师模型的网络配置。在这里[3 × 3，128] × 4，*S* 2表示4个级联卷积层，其中128个过滤器的大小为3 × 3，步幅设置为2。默认步幅设置为1。“ FC”是完全连接的层。   |  |  |  |  |  | | --- | --- | --- | --- | --- | | 转换1.x | Conv2.x | Conv3.x | Conv4.x | 足球俱乐部 | |  |  |  |  | 512 |   表2.学生模型的网络配置。在这里[3 × 3，128] × 4，*S* 2表示4个级联卷积层，其中128个过滤器的大小为3 × 3，步幅设置为2。默认步幅设置为1。“ FC”是完全连接的层。 |

*大号* ˉ *整体* = *λ* ˉ 1 *大号硬* + *λ* ˉ 2 *大号软* + *λ* ˉ 3 *大号MKFeature* - *MMDNorm*（9）

# 实验与结果

## 实验设定

在这里，测试集包括LFW [19]和SCFace [20]。LFW数据集是用于人脸识别引擎的广泛使用的测试集。使用MTCNN检测LFW中的面部并将其对齐为112 × 96分辨率。为了模拟LR-FR的设置，我们修改了原始LFW数据集，其中画廊图像保持不变，并且通过与高斯核卷积将探针图像预处理为低分辨率图像。探测图像的分辨率可以为8 × 8、12 × 12、16 × 16、20 × 20，原始图像为112 × 96。此新数据集称为低分辨率LFW（LR-LFW）。

SCFace是用于评估LR人脸识别引擎的广泛使用的基准[21]。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。对于每个对象，监视摄像机在三个不同的距离（4.20m，2.60m和1.00m）上总共拍摄了15张图像，而数码相机则拍摄了一张正面的面部照片。在此，将正面面部照片用作图库图像，并将监视摄像机在不同距离处拍摄的图像用作探测图像。检测和对齐的设置与LFW数据处理相同。

## 深度神经网络架构配置

在我们的实验中，我们使用Sphere-Face Loss预先训练64层ResNet网络，并将其作为教师模型。训练数据是CASIA Webface数据集[16]。教师的网络架构与[22]中的描述相同。在这里，学生模型使用了一个36层更薄的网络。与教师模型相比，它可以达到3倍的加速度。表1和表2分别显示了教师模型和学生模型的网络体系结构。在表1和表2中，Conv1.x，Conv2.x，Conv3.x和Conv4.x表示多个卷积层和残差单元，在双列括号中显示。例如，[3 × 3，128] × 4，*S* 2表示4个级联卷积层，其中128个过滤器的大小为3 ×3，跨步设置为2。默认跨步设置为1。“ FC”是完全连接的层。

## 评估改进的知识蒸馏方案

在这一部分中，首先，我们比较原始方案和改进方案的整个训练时间，然后评估MK-MMD的性能。最后，将该方法与其他最新方法进行了比较。

### 对不同知识蒸馏方案的评估

在这一部分中，我们进行了实验，以评估两种不同的知识蒸馏方案在LFW和SCFace的整体训练时间和性能：

KD ：原始知识蒸馏方案，包含2个步骤。步骤1，使用LR增强数据更新教师模型的权重；步骤2，在更新的教师模型指导下，使用相同的训练集来训练学生模型。

KD *im*：改进的知识蒸馏方案，包含一个步骤：使用LR增强训练集在原始教师模型模型的指导下训练学生模型。

在这两种不同的策略中，KD 采用了原始的知识蒸馏方案。KD *im*是我们改进的知识蒸馏方案，其中仅在原始教师模型的指导下使用LR增强训练组训练学生模型。这里使用CASIA-webface数据集进行训练。在这里，我们使用4个Nvidia Titan Xp GPU，批量大小为512。采用Caffe进行训练，该过程将持续20个纪元。教师模型的训练大约需要2天才能完成，而学生模型的训练则只需0.5天，因为它的规模要小得多。因此，KD中的两个步骤总共大约需要2.5天才能完成，而KD *im*方案的单步过程仅需0.5天。由于改进的方案（KD *im*）避免了耗时的教师再训练，因此整个训练过程可以加快大约5倍。这两种不同方案的训练时间如表3所示。

对于大型训练集而言，减少训练时间将更为重要。例如，如果采用DeepGlint [23]训练集，其中包含181K ID和6.7M图像，则使用相同的训练参数和硬件设置进行KD 方案的训练可能需要一周以上的时间。通过使用KD *im*方案，我们可以避免耗时的教师模型的权重调整，并且仍然能够在大约1.5天内获得快速的LR-FR模型。

|  |  |
| --- | --- |
| 训练计划 | 训练时间 |
| KD | 2.5天 |
| KD *IM* | 0.5天 |

表3.不同训练策略的训练时间。“ KD”代表原始知识蒸馏方案。“ KD *im* ”代表我们改进的知识蒸馏方案。

除了缩短训练时间外，还在LR-LFW和SCFace数据集上比较了这两种知识蒸馏方案的性能。结果如表4和表5所示。从表中可以看出，用KD *im*方案和KD 方案训练的学生模型具有相同水平的性能。因此，我们可以得出结论，采用改进的知识蒸馏方案的学生模型可以在保持准确性的同时，将训练过程加快5倍。

在表4和表5中，仅在112 × 96 LFW和d = 1.0m SCFace图像上测试了教师模型。这两个图像集可以视为高分辨率（HR）图像。原因是此处使用的教师模型尚未使用LR增强数据集重新训练。因此，评估LR图像的准确性是不公平的。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 聚苯乙烯 | 8 × 8 | 12 × 12 | 16 × 16 | 20 × 20 | 112 × 96 |
| Ť |  |  |  |  | 99.42 |
| S *KD* | 93.95 | 95.08 | 97.00 | 97.10 | 99.17 |
| 小号*KDim* | 94.05 | 95.20 | 96.74 | 97.13 | 99.03 |

表4. LR-LFW数据集上不同模型的准确性。在此，“ PS”是“ probe size”的缩写。“ T”代表教师模型。“ S *KD* ”代表采用原始知识蒸馏方案训练的学生模型。'S *KDim* '代表采用我们改进的知识蒸馏方案训练的学生模型。

|  |  |  |  |
| --- | --- | --- | --- |
| 距离 | *d*＝ 4 *。*2 *米* | *d*＝ 2 *。*6 *米* | *d*＝ 1 *。*0 *米* |
| Ť |  |  | 99.13 |
| S *KD* | 73.88 | 93.50 | 98.34 |
| 小号*KDim* | 73.20 | 93.95 | 98.03 |

表5. SCFace数据集上不同模型的人脸识别率。在这里，“ T”代表教师模型。“ S *KD* ”代表采用原始知识蒸馏方案训练的学生模型。'S *KDim* '代表采用我们改进的知识蒸馏方案训练的学生模型。

### MK-MMD损失的表现

在这一部分中，我们主要集中在改进的知识蒸馏方案中对MKMMD损失函数的评估。我们比较了两个实验：

**L2损失**：在改进的知识蒸馏方案中，使用方程式（6）作为损失函数，其中采用特征L2损失代替MK-MMD损失。

**MK-MMD损失**：在改进的知识提纯方案中，使用公式（9）作为损失函数，其中采用了MK-MMD损失特征。

表6和表7显示了两个数据集（LR-LFW，SCFace）的结果。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 聚苯乙烯 | 8 × 8 | 12 × 12 | 16 × 16 | 20 × 20 | 112 × 96 |
| 小号*大号* | 90.36 | 91.75 | 94.66 | 96.83 | 99.00 |
| 小号*中号* | 94.05 | 95.20 | 96.74 | 97.13 | 99.03 |

表6.对于使用不同损失组合训练的学生模型的LR-LFW评估。“ PS”是“ probe size”的缩写。“ S *L* ”代表受L2损失组合训练的学生模型。“ S *M* ”代表受MKMMD损失组合训练的学生模型。

|  |  |  |  |
| --- | --- | --- | --- |
| 距离 | *d*＝ 4 *。*2 *米* | *d*＝ 2 *。*6 *米* | *d*＝ 1 *。*0 *米* |
| 小号*大号* | 69.37 | 91.54 | 97.98 |
| 小号*中号* | 73.20 | 93.95 | 98.03 |

表7.使用经过不同损失训练的学生模型评估SCFace人脸识别率。“ S *L* ”代表受L2损失组合训练的学生模型。S *M*代表受MK-MMD损失组合训练的学生模型。

从表6和表7，可以看出MK-MMD损失的结果不断优于在两个试验组L2损失，尤其是在较小的尺寸探头图像场景中，如8 × 8个，12 × 12 LR-LFW图像和*d* = 4 *。*2 *米*SCFace图像。原因是，当使用原始特征L2损失时，由于l2损失函数的约束较弱，教师未更改的训练集与学生的LR增强训练集之间的数据分布差异可能会损害性能。当探针图像尺寸越来越小时，这个问题将变得更加严重。通过采用特征MK-MMD损失函数可以减少此问题的影响，可以将其视为内核技巧和L2损失的组合，并且在训练过程中会对特征施加更强的约束。

### SCFace上的性能比较

在本部分中，将SCFace的结果与三种最新方法进行了比较，包括深耦合树脂模型（DCR）方法[24]，多维缩放（MDS）[25]和判别多维缩放（DMDS）[26]。 ，（表8）。包括DCR和我们的学生模型在内的最佳性能方法的推理时间在Intel Core 2.5GHz CPU上进行了测试，以模拟嵌入式系统的硬件性能。推理时间如表9所示。从结果中，我们可以安全地得出一个结论，即我们的模型可以实现可比的，有时甚至更好的性能，同时能够将推理时间减少到35ms。

|  |  |  |  |
| --- | --- | --- | --- |
| 距离 | *d*＝ 4 *。*2 *米* | *d*＝ 2 *。*6 *米* | *d*＝ 1 *。*0 *米* |
| 教师模型 |  |  | 99.13 |
| 小号*KDim* | 73.20 | 93.95 | 98.03 |
| DCR [24] | 73.30 | 93.50 | 98.00 |
| DMDS [26] | 62.70 | 70.70 | 65.50 |
| MDS [25] | 60.30 | 66.00 | 69.50 |

表8. SCFace上不同距离的不同模型的面部识别率。“教师”表示64层Resnet教师模型。“ S *KDim* ”代表采用我们改进的知识蒸馏方案训练的学生模型。

|  |  |
| --- | --- |
| 模型 | 推论时间 |
| 小号*KDim* | 35 *毫秒* |
| DCR [24] | 132 *毫秒* |

表9.我们的模型和DCR模型在CPU上的推理时间。S *KDim*代表使用我们改进的知识蒸馏方案训练的学生模型。

更好的性能背后的原因可以归结为两个方面：

1. 更好的教师模型：64层教师模型在HR面部图像上具有更高的性能，即*d* = 1 *。*与DCR和其他方法相比，SCFace为0 *m*（表8）。较大规模的网络结构和出色的Sphereface损失特征是教师表现更好的主要原因。得益于改进的知识蒸馏方案，与高分辨率图像集上的教师模型相比，我们的学生模型的性能仅略有下降。
2. 改进的知识蒸馏方案仅对学生模型使用LR增强训练集，这迫使学生模型的输出与教师的输出相同，而与LR差异无关。无论输入是LR还是HR，都可以将教师更好的表示能力转移到学生模型上，从而提高了学生模型的LR图像表现。

# 结论

在本文中，我们提出了一种改进的知识蒸馏方案，用于快速LR-FR模型训练。为了避免教师模型的耗时训练过程，我们保持教师模型的训练集不变，而仅将LR增强添加到学生模型的训练集中。这可以使我们避免更新教师模型模型的权重，并且仍然能够训练LR-FR学生模型，这将减少整个训练过程的时间成本。仅将LR增强添加到学生模型的训练集中将增加教师和学生的训练输入之间的分布差异。可以通过最小化MK-MMD损失函数来减少这种差异。结果表明，我们的方法可以将训练时间减少约5倍，同时保留了学生模型的准确性。我们的学生模型可以实现与教师模型相比，加速度为3倍，并且在CPU上运行仅需35ms。改进后的方案也可以推广到其他数据方差，例如照明，姿势等。

# 参考文献

1. Chen Wenlin，James T. Wilson，Stephen Tyree，Killian Q. Weinberger和Yixin Chen。用哈希技巧压缩神经网络。*CoRR，abs / 1504.04788*，2015年。
2. 龚云超，刘柳，杨明和卢博米尔·布尔德夫。使用向量量化压缩深度卷积网络。*arXiv预印本arXiv：1412.6115，2014*。
3. Song Han，Huizi Mao和William J. Dally。深度压缩：通过修剪，训练有素的量化和霍夫曼编码来压缩深度神经网络。*CoRR，abs / 1510.00149*，2015年。
4. Alexander Novikov，Dmitry Podoprikhin，Anton Osokin和Dmitry P. Vetrov。张紧神经网络。*《神经信息处理系统的进展》，NIPS*，2015年。
5. 杰弗里·欣顿（Geoffrey Hinton），奥里尔·维尼亚尔斯（Oriol Vinyals）和杰夫·迪恩（Jeff Dean）。在神经网络中提取知识。*在深度学习与表示学习研讨会上，NIPS*，2014年。
6. 罗平，朱振耀，刘子伟，王小刚和唐小鸥。通过从神经元中提取知识来压缩人脸模型。*第30届AAAI人工智能会议*，第3560--3566页，2016年。
7. 龙明胜，曹Yue，王建民和迈克尔·乔丹。通过深度适应网络学习可转让的特征。*arXiv预印本arXiv：1502.02791，2015*。
8. 罗自孟，胡佳妮，邓卫红和沉海峰。深度无监督领域自适应，可用于人脸识别。*2018年第13届IEEE国际自动人脸和手势识别会议（FG 2018）*，2018。
9. 王梦娇，刘如杰，安倍成茂，内田秀久，松南智明和山田茂史。通过改进师生训练，发现有效的人脸识别模型压缩策略。*在2018年第25届IEEE国际图像处理会议（* ICIP *）中*，第2416-2420页，2018年。
10. 何开明，张向云，任少青和孙健。深度残差学习，用于图像识别。*计算机视觉与模式识别》，CVPR*，2016年。
11. Forrest N. Iandola，Song Han，Matthew W. Moskewicz，Khalid Ashraf，William J. Dally和Kurt Keutzer。Squeezenet：Alexnet级别的精度，参数减少了50倍，模型尺寸只有1mb。*arXiv预印本arXiv：1602.07360，2016*。
12. 安德鲁·霍华德，朱梦露，陈博，德米特里·卡列尼琴科，王卫军，托比亚斯·韦恩德，马可·安德烈托和哈特维格·亚当。Mobilenets：用于移动视觉应用的高效卷积神经网络。*arXiv预印本arXiv：1704.04861，2017*。
13. 张向云，周新宇，林孟小和孙健。Shufflenet：一种用于移动设备的极其高效的卷积神经网络。*arXiv预印本arXiv：1707.01083，2017*。
14. 高煌，刘壮和基利安·温伯格。密集连接的卷积网络。*计算机视觉与模式识别，CVPR*，2017年。
15. 克里斯蒂安·塞格迪，谢尔盖·艾菲和文森特·凡豪克。Inception-v4，inception-resnet以及残余连接对学习的影响。*CoRR，abs / 1602.07261*，2016年。
16. 董毅，甄磊，廖圣才和史丹利。从头开始学习面部表情。*CoRR，abs / 1411.7923*，2014年。
17. 张开鹏，张展鹏，李志峰和于乔。使用多任务级联卷积网络进行联合人脸检测和对齐。*在IEEE Signal Processing Letters，23（10）*，1499– = 1503页，2016中。
18. 黄泽浩和王乃艳。像您喜欢的那样：通过神经元选择性转移提取知识。*arXiv预印本arXiv：1707.01219，2017*。
19. Gary B. Huang，Marwan Ramesh，Tarama Berg和Eric Learned-Miller。带有标签的野外面孔：用于在不受限制的环境中研究面部识别的数据库。*在马萨诸塞大学阿默斯特分校，《技术报告》*，2007年。
20. Mislav Grgic，Kresimir Delac和Sonja Grgic。Scface-监控摄像机的人脸数据库。*多媒体工具应用* ，第863--879页，2011年。
21. 黄志武，山士光，王如萍，张海宏，老士宏，阿里夫·库尔班和陈锡林。scface人脸数据库中基于视频的人脸识别的基准测试和比较研究。*IEEE Transactions on Image Processing*，2015年。
22. 刘未央，温彦东，于志定，李明，比克莎·拉杰和乐松。Sphereface：深层超球面嵌入，用于人脸识别。*CVPR*，2017年。
23. http://trillionpairs.deepglint.com/overview。
24. 吕泽，江旭东和亚历克斯·科特。深度耦合的Resnet用于低分辨率人脸识别。*IEEE信号处理。Lett。，*第526--530页，2018年。
25. Sivaram P. Mudunuri和Soma 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* ，第1034--1040页，2016年。
26. 杨福伟，杨文明，高日强和廖庆明。区分性多维缩放，用于低分辨率人脸识别。*IEEE信号处理。来吧* ，第388--392页，2018年。
27. 原文

提供更好的翻译建议

1. 作者：wangmengjiao@cn.fujitsu.com [↑](#footnote-ref-0)
2. [↑](#footnote-ref-1)