

基于深度学习的人脸识别表示的综合分析

Mostafa Mehdipour Ghazi 工程与自然
科学学院萨班吉大学，土耳其伊斯坦布尔

mehdipour@sabanciuniv.edu

Hazım Kemal Ekenel 计算机工程系伊
斯坦布尔技术大学，土耳其伊斯坦布尔

ekenel@itu.edu.tr

Abstract

基于深度学习的方法一直在人脸识别领域占据主导地位，因为它们在具有挑战性的野生数据集上提供了显著的性能改进。这些方法已经在此类不受约束的数据集上进行了广泛的测试，例如在 Wild 中的 Labeled Faces 和 YouTube Faces 上。然而，迄今为止，它们处理由头部姿势、照明、遮挡和错位等因素引起的个体外观变化的能力尚未得到彻底评估。在本文中，我们提出了一项综合研究，以评估基于深度学习的人脸表示在多种条件下的性能，包括不同的头姿势角度、上下面部遮挡、不同强度的照明变化以及由于面部特征定位错误而导致的未对准。两个成功且公开的深度学习模型，即 VGG-Face 和 Lightened CNN，已被用来提取人脸表示。获得的结果表明，虽然深度学习为人脸识别提供了强大的表示，但它仍然可以从预处理中受益，例如姿势和照明归一化，以便在各种条件下实现更好的性能。特别是，如果这些变化不包含在用于训练深度学习模型的数据集中，则预处理的作用就变得更加重要。实验结果还表明，基于深度学习的表示对于错位具有鲁棒性，并且可以容忍高达 10% 的眼间距离的面部特征定位误差。

一、简介

人脸识别是计算机视觉中具有多种生物识别应用的一个具有挑战性的问题。这个问题本质上面临的困难是由于照明、表情和配件的部分遮挡等因素导致面部外观的变化。

近年来，基于深度学习的方法越来越多地应用于人脸识别，并取得了良好的效果 [32,30,24,21,37]。这些方法将原始数据作为网络输入，并在多个级别上卷积过滤器，以自动从标记或未标记的数据中发现低级和高级表示，以检测、区分和/或分类其潜在模式 [9, 12、13、31、14]。然而，在深度学习架构中优化数百万个参数以从头开始学习多级权重需要数百万个训练样本并能够访问图形处理单元 (GPU) 等强大的计算资源。因此，迁移学习的方法 [34, 20] 被有效地利用，将先前学到的相关视觉识别问题的知识应用到新的、期望的任务领域。迁移学习可以以两种不同的方式应用，涉及预训练之间的大小和相似性。训练数据集和新数据库。第一种方法是通过反向传播使用新数据集微调预训练的网络权重。这种方法仅建议用于足够大的数据集，因为用很少的训练样本微调预训练网络可能会导致过度拟合 [39]。第二种方法是直接利用所需问题中的学习权重来提取特征并随后进行分类。当新数据集很小和/或存在少量类时，该方案特别有效。根据两个数据集之间的任务相似性，可以决定是否使用较低层的权重（作为通用低级特征提取器）或较高层的权重（作为任务特定主题提取器） [14]。在本文中，较高层部分来自 VGG-Face [21] 和 Lightened CNN [37] 两个深度卷积神经网络 (CNN) 的学习权重，在非常大的人脸识别集合上进行了预训练，已用于提取人脸代表。选择这两个模型是因为它们被发现在公开的情况下能够成功地进行野外人脸识别。前一个网络包括非常深的架构和

后者是一个计算效率高的 CNN。这些深层人脸表征针对不同因素（包括光照、遮挡、姿势和错位）变化的鲁棒性已经使用五个流行的人脸数据集进行了彻底评估，即 AR [17]、CMU PIE [25]、扩展耶鲁数据集[7]、Color FERET [23]和FRGC [22]。这项工作的主要贡献和成果可以概括如下：（i）深度学习的综合评价已经进行了各种条件下的基于表示，包括姿势、照明、遮挡和未对准。事实上，所有提出的基于深度学习的人脸识别方法，如 DeepFace[32]、DeepID[30]、FaceNet[24] 和 VGG-Face[21] 都已经在非常大的野生人脸识别数据集上进行了训练和评估，即 Labeled Faces in the Wild (LFW) [10]、YouTube Faces (YTF) [35] 和 MegaFace [18]。然而，他们处理个体外观变化的表现能力尚未得到评估。（ii）我们已经证明，虽然深度学习为人脸识别提供了强大的表示，但它无法在姿势、照明和遮挡方面取得最先进的结果。为了使深度学习模型取得更好的结果，要么在训练过程中考虑这些变化，要么应将姿态和光照归一化的预处理方法与预训练模型一起使用。（iii）我们发现基于深度学习的面部表示对于错位具有鲁棒性，并且能够容忍高达 10% 的眼间距离的面部特征定位误差。（iv）与 Lightened CNN 模型 [37] 相比，VGG-Face 模型 [21] 具有更高的可迁移性。总的来说，我们认为基于深度学习的人脸识别需要进一步的研究来解决不匹配条件下的人脸识别问题，特别是当手头的任务可用的数据量有限时。本文的其余部分组织为接下来。第 2 节回顾了现有人脸识别深度学习方法。第 3 节描述了两种用于人脸识别的 DeepCNN 模型的细节，并提出了基于这些模型的人脸表示的提取和评估方法。第 4 节解释了所使用的数据集并介绍了设计的实验及其结果。最后，第五部分总结了本文，总结并讨论了所进行的实验以及所获得结果的含义。

2. 相关工作

在深度学习算法出现之前，大多数传统的人脸识别方法首先使用局部二值模式（LBP）、尺度不变特征变换（SIFT）和 Ori 直方图从面部图像中本地提取手工制作的浅层特征-oriented Gradients (HOG)，然后训练特征并分类

通过支持向量机 (SVM) 或最近邻 (NN) [1,6,27,3] 进行身份识别。然而，随着最先进的计算资源的可用性以及对大型数据集的访问激增，深度学习架构已经开发出来，并在包括人脸识别在内的不同视觉识别任务中显示出令人印象深刻的结果[32, 30, 24, 21, 37]。DeepFace [32] 是这些杰出的网络之一，它包含一个带有两个卷积层和超过 1.2 亿个的九层深度 CNN 模型对来自 4,000 多个身份的 400 万张面部图像进行参数训练。该方法通过基于 3D 模型的图像对齐和使用 CNN 集合，可以在 LFW 和 YTF 数据集上分别实现 97.35% 和 91.4% 的准确率。深度隐藏身份特征（DeepID）[30]是另一种成功的深度学习方法，提出用于人脸识别和验证，具有九层网络和四个卷积层。该方案首先通过人脸识别学习权重，并使用最后的隐藏层输出提取特征，然后将它们泛化到人脸验证。DeepID通过基于两个眼睛中心和两个嘴角的相似变换来对齐人脸。该网络在 CelebrityFaces 数据集 (CelebFaces) [29] 上进行训练，并在 LFW 数据集上达到 97.45% 的准确率。FaceNet [24] 是一个基于 GoogLeNet [31] 和 [40] 中提出的网络的深度 CNN，并在包含约 800 万个身份的 100 到 2 亿张图像的面部数据集。该算法使用从在线三元组挖掘方法获得的粗略对齐的面部三元组，并直接学习将面部图像映射到紧致欧几里得空间以测量面部相似度。FaceNet 已在 LFW 和 YTF 数据集上进行评估，准确率分别达到 99.63% 和 95.12%。

3. 方法

在本节中，我们将介绍并描述两种成功的人脸识别 CNN 架构，并讨论基于这些模型的人脸表示。

3.1. VGG-Face Network

VGG-Face [21]是一种使用VGGNet架构[26]用于人脸识别的深度卷积网络。它使用从网络收集的 2,622 个身份的 260 万张面部图像进行训练。该网络包含 16 个卷积层、5 个最大池层、3 个全连接层和一个带有 Softmax activation 的最终线性层。VGG-Face 将大小为 224×224 像素的彩色图像块作为输入，并在全连接层中利用 dropout 正则化 [28]。此外，它将 ReLU 激活应用于所有卷积层。跨越 1.44 亿个参数清楚地表明 VGG 网络是一种计算成本高昂的架构。

VGG network is a computationally expensive architecture.

该方法在LFW数据集上进行了评估，准确率达到98.95%。

3.2. Lightened CNN

该框架是为人脸识别而提出的计算复杂度较低的 CNN [37]。与 ReLU 相比，它使用称为 Max-Feature-Map (MFM) 的激活函数来提取更多抽象表示。Lightened CNN 在两种不同的模型中被引入。第一个网络 (A) 受到 AlexNet 模型 [13] 的启发，包含 3,961K 个参数，其中有四个使用 MFM 激活函数的卷积层、四个最大池化层、两个全连接层和一个使用 Softmax 激活的线性层在输出中。第二个网络 (B) 受到 Network in Network 模型 [16] 的启发，涉及 3,244K 个参数，其中有五个使用 MFM 激活函数的卷积层、四个用于降维的卷积层、五个最大池化层，两个完全连接的层，以及输出中具有 Softmax 激活的线性层。Lightened CNN 模型以大小为 128×128 像素的灰度面部块图像作为网络输入。这些模型在 CASIA WebFace 数据集 [38] 中的 10,575 个身份的 493,456 个面部图像上进行训练。两种 Lightened CNN 模型均在 LFW 数据集上进行了评估，准确率分别达到 98.13% 和 97.77%。

3.3. 使用 CNN 模型进行人脸表示

VGG-Face 和 Lightened CNN 的实现和预训练模型用于 Caffe 深度学习框架 [11]。为了系统地评估上述深度 CNN 模型在不同外观变化下的鲁棒性，每个网络的所有层权重，直到第一个全连接层（在最后一个 dropout 层和具有 Softmax 激活的全连接层之前）都用于特征提取。这些层在 VGG-Face 和 Lightened CNN 模型中分别表示为 FC6 和 FC1。为了分析不同全连接层的效果，我们还部署了 VGG-Face 网络的 FC7 层。VGG-Face 模型提供从大小为 224×224 像素的彩色图像块中提取的 4096 维高级表示，而 Lightened CNN 模型提供从大小为 128×128 像素的灰度图像块中获取的 512 维高级特征向量。128 像素。然后使用具有余弦距离度量的最近邻方法对提取的特征进行分类。尽管我们也测试了其他指标，例如欧几里德距离和互相关，但余弦距离几乎总是取得最佳结果。在需要时应用包括对齐和/或照明归一化和对比度增强在内的预处理步骤。面部对齐是相对于眼睛中心完成的，同时照明标准化和控制



图 1: 来自 AR 数据库的具有不同遮挡条件的样本。左起前三幅图像与会话 1 相关，接下来的三幅图像是从会话 2 中获得的，重复条件为中性、戴墨镜和戴围巾。

使用[33]中提出的方法进行对比度增强。在进行的实验中，数据集中的中性图像（包括在受控照明下从正面姿势捕获的面部图像，没有面部遮挡）用于画廊图像，而探测图像包含多个由于头部姿势、照明变化、面部遮挡和错位而导致的外观变化。

4 实验与结果

在本节中，我们提供所使用的数据集和实验设置的详细信息。此外，我们还介绍了用于评估基于 CNN 的深度人脸表示的场景，并讨论了所获得的结果。

4.1. AR人脸数据库——人脸遮挡

AR人脸数据库[17]包含来自126名受试者的 4,000张彩色、尺寸为 768×576 像素的正面面部图像，具有不同的面部表情、照明和遮挡。每个受试者都参加了两次相隔两周的会议，并且没有头饰限制、化妆、发型、配饰等。由于本实验的目的是对基于 CNN 的深层特征对抗遮挡的鲁棒性进行基准测试，因此每个受试者一张图像第一个会话中的中性表达用于训练。随后，每个受试者每次使用两张图像进行测试，一张戴着太阳镜测试上脸遮挡的影响，另一张戴着围巾测试下脸遮挡的效果。总共，这些样本可以从 110 名受试者中完整获取。随后对每个选定的图像进行对齐、裁剪成方形面部斑块，并缩放至 224×224 或 128×128 像素。最后，从每幅图像中减去从 VGG-Face 训练集中获得的平均图像，以确定应用于预训练模型的相同图像变换的实现。图 1 显示了用于实验的 AR 数据库中与一名受试者相关的图像。在 AR 数据集上进行了四个实验。前两个实验涉及在第一个会话中进行训练和测试，其余实验则使用第一个会话中的样本进行训练，并使用第二个会话中的图像进行测试。表 1 总结了结果

from the second session. Table 1 summarizes the results of

表 1: 针对不同遮挡条件使用深度特征的 AR 数据库的分类结果 (%)

Testing Set	VGG-Face		Lightened CNN
	FC6	FC7	
Sunglasses Session 1	33.64	35.45	5.45 (A)
Scarf Session 1	86.36	89.09	12.73 (A)
Sunglasses Session 2	29.09	28.18	7.27 (B)
Scarf Session 2	85.45	83.64	10.00 (A)

我们使用 AR 人脸数据库进行遮挡变化的实验。从表 1 中可以看出，深度面部表示很难处理由于佩戴太阳镜而导致的上脸遮挡。与最先进的遮挡鲁棒人脸识别算法[36,4,19]相比，深度表示所获得的结果相当低。这些结果表明，除非在大量遮挡数据上进行专门训练，否则深度表示的结果相当低。当存在面部遮挡时，基于 CNN 的表示可能无法正常工作。在相同的实验中，还发现与 Lightened CNN 模型相比，VGG-Face 模型在对抗面部遮挡方面更加稳健。在此表中，仅显示了性能最佳的 Lightened CNN 模型的结果。

4.2. CMU PIE 数据库 - 光照变化

CMU PIE 人脸数据库 [25] 包含 41,368 张彩色、尺寸为 640×480 像素的人脸图像，拍摄自 68 名受试者，在 13 种不同的头部姿势、43 种区分条件和 4 种不同的表情下。由于实验的目标是评估照明变化对基于 CNN 的深度特征性能的影响，因此选择来自 CMU PIE 数据集照明子集的正面图像进行进一步分析。该子集包含每个主题在不同照明条件下拍摄的 21 张图像。一个正面照明的面部图像用于训练，其余 20 张包含不同照明的面部图像用于测试。所有收集到的图像都经过对齐、裁剪成方形面部斑块，最后缩放到 224×224 或 128×128 像素。然后从每个图像中减去 VGG-Face 平均图像。作为示例，图 2 描述了来自 CMU PIE 数据库的一名受试者所使用的样本。光照变化的实验结果如表 2 所示。可以看出，使用 VGG-Face 获得的深度表示对于光照变化具有鲁棒性。然而，与最先进的光照鲁棒人脸识别方法所获得的结果相比，所获得的精度略低 [15,41,33]。这些结果表明深度学习的性能



图 2: 来自 CMU PIE 数据库的不同光照条件的样本。左上角的第一张图片是用于训练的正面图片，其余的用于测试。

面部表示需要使用基于照明的预处理方法进一步改进[33]。

4.3. 扩展耶鲁数据集 - 光照变化

扩展的耶鲁人脸数据集 B [7] 包含从 38 名受试者在 9 种姿势和 64 种照明变化下捕获的 16,128 张图像。这 64 个样本根据光源方向与相机光轴之间的角度分为 5 个子集；子集 1 包含七个光照角度小于 12 度的图像；子集 2 有 12 张角度在 20 度到 25 度之间的图像；子集 3 有 12 张角度在 35 度到 50 度之间的图像；子集 4 有 14 个角度在 60 到 77 度之间的图像；最后，子集 5 包含 19 个角度大于 77 度的图像。换句话说，通过增加子集数量，光照变化会变得更强。为了使用基于深度 CNN 的特征评估光照变化的影响，仅选择该数据集在所有光照变化下的正面图像。第一个具有几乎完美正面光照的子集用于训练，而子集 2 至 5 用于测试。所有获得的图像随后都会被对齐，裁剪成方形面部斑块，最后缩放到 224×224 或 128×128 像素。然后从每个图像中减去 VGG-Face 平均图像。图 3 显示了与扩展耶鲁数据库 B 中的一个受试者相关的一些样本。表 3 报告了使用扩展耶鲁数据集 B 子集进行的照明变化实验的结果。

表 2: CMU PIE 数据库在不同照明条件下使用深层面部特征的分类结果 (%)

	VGG-Face		Lightened CNN
	FC6	FC7	
Accuracy	93.16	92.87	20.51 (A)

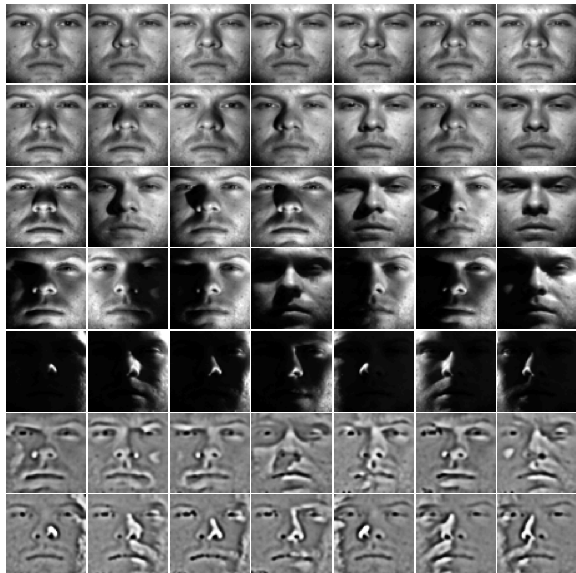


图 3: 来自扩展耶鲁数据集 B 的不同照明条件下的样本。第1到5行分别对应子集1到5, 最后两行是子集4和5的预处理样本。

可以看出, 深度面部表示对于子集 2 和 3 中存在的小光照变化具有鲁棒性。然而, 随着光照变化强度的增加, 性能会显著下降。造成这种结果的主要原因可以归因于这样一个事实: 深度面部模型主要是根据从网络上获得的名人图片进行训练的, 这些图片通常是在相对明亮的条件下收集的。因此, 他们没有学会处理强烈的光照变化。解决这个问题一个方法是在应用深度 CNN 模型进行特征提取之前进行预处理。为了评估图像预处理的贡献, 我们通过照明归一化和对比度增强对子集 4 和 5 的人脸图像进行预处理,

表 3: 扩展耶鲁数据库 B 使用针对各种照明条件的深度表示的分类结果 (%)

Testing Set	VGG-Face		Lightened CNN
	FC6	FC7	
Subset 2	100	100	82.43 (A)
Subset 3	88.38	92.32	18.42 (B)
Subset 4	46.62	52.44	8.46 (B)
Subset 5	13.85	18.28	4.29 (B)
Preprocessed Subset 4	71.80	75.56	26.32 (A)
Preprocessed Subset 5	73.82	76.32	24.93 (A)



图 4: 来自颜色 FERET 数据库的具有不同姿势条件的样本。左侧第一张图像是用于训练的正面图像 (fa), 其余图像 (ql、qr、hl、hr、pl、pr) 用于测试。

对这些新获得的子集进行相同的实验。相应的结果如表3的最后两行所示。可以看出, 预处理有助于提高获得的精度。这些结果证明, 尽管深度 CNN 为人脸识别提供了强大的表示, 但它们仍然可以从预处理方法中受益。当测试集中可用的变化未考虑到预训练时, 尤其如此, 因此必须对这些变化进行标准化。

4.4. 颜色 FERET 数据库 - 姿势变化

彩色 FERET 数据库 [23] 包含 11,338 张尺寸为 512×768 像素的彩色图像, 这些图像是在半受控环境中捕获的, 来自 994 名受试者的 13 种不同姿势。为了衡量深层特征针对姿势变化的鲁棒性, 我们使用常规正面图像集 (fa) 进行训练, 每个受试者一张正面图像。然后在六个非正面姿势上测试网络, 包括两个四分之一左 (ql) 和四分之一右 (qr) 姿势, 头部向左和向右倾斜约 22.5° , 两个半左 (hl) 和半右 (hr) 姿势头部倾斜约 67.5° , 以及两个侧向左 (pl) 和侧向右 (pr) 的姿势, 头部倾斜约 90° 。所有使用的图像都被裁剪成正方形面部补丁并缩放至 224×224 或 128×128 像素。然后从每个图像中减去 VGG-Face 平均图像。图 4 显示了 FERET 数据库中 与一名受试者相关的样本。表 4 报告了在相同数据集上获得的姿势变化的准确性。如结果所示, VGG-Face 模型能够

表 4: 针对不同姿态条件使用深度特征的 FERET 数据库的分类结果 (%)

Testing Set	VGG-Face		Lightened CNN
	FC6	FC7	
Quarter Left	97.63	96.71	25.76 (A)
Quarter Right	98.42	98.16	26.02 (A)
Half Left	88.32	87.85	6.08 (B)
Half Right	91.74	87.85	5.98 (A)
Profile Left	40.63	43.60	0.76 (B)
Profile Right	43.95	44.53	1.10 (B)



图 5: 来自 FRGC 数据库的样本与不同的配准错误对齐。前三行是从 FRGC1 获取的, 显示训练样本 (row1) 和测试样本分别与零 (row2) 和 10% (row3) 配准误差对齐。第二三行与 FRGC4 相关, 并描述了分别与零 (第 5 行) 和 20% (第 6 行) 配准误差对齐的训练样本 (第 4 行) 和测试样本。

处理高达 67.5 度的姿势变化。尽管如此, 通过采用姿势标准化方法可以进一步改善结果, 该方法已被发现可用于人脸识别[5,2,8]。当使用侧面图像测试系统时, 性能显著下降。除了正面到侧面的人脸匹配是一个具有挑战性的问题之外, 深度 CNN 人脸模型的训练数据集中缺乏足够的侧面图像可能是性能下降的原因。

4.5. FRGC 数据库 - 错位

人脸识别大挑战 (FRGC) 数据库[22]包含在受控和非受控环境下在两种不同照明条件下拍摄的具有中性或微笑面部表情的正面人脸图像。受控图像子集是在工作室环境中拍摄的, 而不受控照片是在走廊、中庭或户外拍摄的。为了评估基于 CNN 的深度特征对抗错位的鲁棒性, 我们利用并划分了 2003 年秋季和 2004 年春季的照片集分为受控子集和非受控子集以获得四个新子集, 每个子集包含 120 名受试者的照片, 每个受试者十张图像。2003 年秋季子集是用于画廊, 而 2004 年春季的那些被用作探测图像。换句话说, 图库图像来自 2003 年秋季的受控 (非受控) 子集, 探测图像来自 2004 年春季的受控 (非受控) 子集。我们将在受控条件下进行的实验命名为 FRGC1, 将在不受控条件下进行的实验命名为 FRGC1。 , FRGC4。与之前的任务类似, 图库图像相对于手动注释的眼睛中心坐标进行对齐,

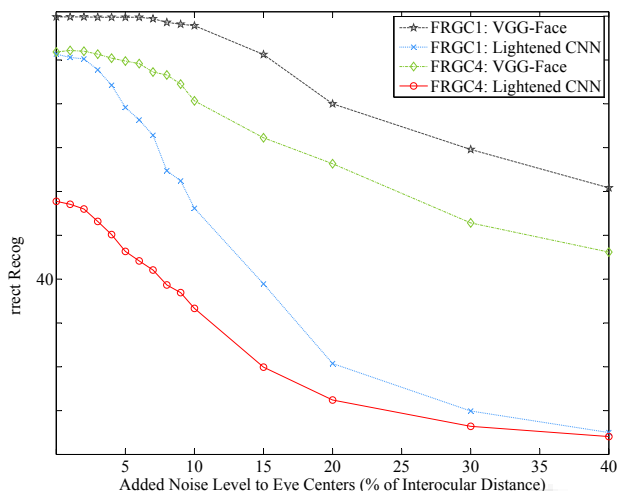


图 6: 使用深度面部表征针对不同面部特征定位错误级别的 FRGC 数据集的分类结果

裁剪成方形面部补丁, 并缩放至 224×224 或 128×128 像素。还从每个图像中减去 VGG-Face 平均图像。为了模拟由于错误的面部特征定位而导致的未对准, 在对齐探针图像时, 将高达眼睛之间距离的 40% 的随机噪声添加到手动注释的眼睛中心位置。图 5 显示了来自 FRGC 数据库不同子集的与一名受试者相关的示例面部图像。所使用的深度模型针对不同程度的面部特征定位误差的分类结果如图 6 所示。请注意, 此任务的 VGG-Face 特征均从 FC6 中获得。对图 6 中显示的结果的分析表明, 深度模型基于 CNN 的面部表示对于错位具有很强的鲁棒性, 即它可以容忍面部特征定位系统高达 10% 的眼间距离误差。这是一个非常重要的属性, 因为已知传统的基于外观的人脸识别算法对未对准很敏感。

4.6. 面部边界框扩展

作为我们的最后一个实验, 我们评估了深度面部表征与较大面部边界框的对齐情况。为此, 所使用的数据集的每个图像都被对齐并裁剪成扩展的方形面部补丁, 以包括头部的所有部分, 即耳朵、头发和链条。然后将这些图像缩放为 224×224 或 128×128 像素, 并从每个图像中减去 VGG-Face 平均图像。表5显示了不同人脸数据集上较大边界框的对齐结果。将表5中获得的结果与表1至表3的结果进行比较表明, 使用从

Tables 1 to 3 shows that using deep features extracted from

表 5: 使用深度特征对与较大边界框对齐的不同人脸数据集进行分类的结果 (%)

Training Set	Testing Set	VGG-Face (FC6)
AR Neutral Set 1	AR Sunglasses Set 1	44.55
AR Neutral Set 1	AR Scarf Set 1	93.64
AR Neutral Set 1	AR Sunglasses Set 2	39.09
AR Neutral Set 1	AR Scarf Set 2	91.82
CMU PIE Train	CMU PIE Test	97.72
Ext. Yale Set 1	Ext. Yale Set 2	100
Ext. Yale Set 1	Ext. Yale Set 3	94.52
Ext. Yale Set 1	Ext. Yale Set 4	56.58
Ext. Yale Set 1	Ext. Yale Set 5	27.56

整个头部显着提高了性能。对于这一观察结果的一个可能的解释是，VGG-Face 模型是在包含所有头部而不仅仅是面部图像的图像上进行训练的；因此，扩展面部边界框通过包含从整个头部提取的有用特征来提高分类准确性。

五、总结与讨论

在本文中，我们对基于深度学习的人脸识别表示在各种条件下（包括姿势、照明、遮挡和未对准）进行了综合评估。两个成功的深度 CNN 模型，即 VGG-Face [21] 和 Lightened CNN [37]，在非常大的面部数据集上进行了预训练，用于提取面部图像表示。这些实验使用了五个著名的人脸数据集，即 AR 人脸数据库 [17] 用于分析遮挡的影响，CMU PIE[25] 和扩展耶鲁数据集 B [7] 用于分析照明变化，Color FERET 数据库 [23] 评估姿势变化的影响，FRGC 数据库 [22] 评估未对准的影响。事实证明，基于深度学习的表示提供了有希望的结果。然而，所达到的性能水平并不像文献中报告的这些数据库的最先进方法那么高。对于训练数据集中测试条件稀缺的情况，性能差距非常显着。CNN 模型。我们建议使用预处理方法进行姿势和照明归一化以及预先训练的深度学习模型或在训练期间考虑这些变化可以基本上解决这个弱点。除了这些重要的观察结果之外，这项研究还揭示了以下优点：基于深度学习的人脸表示的特点是其对未对准的鲁棒性，因为它们可以容忍由于高达 10% 的眼间距离的面部特征定位误差而导致的未对准。VGG-Face 模型表现出了更好的性能。可转移性

与 Lightened CNN 模型相比。这可能归因于其更复杂的架构，从而产生了更抽象的表示。另一方面，顾名思义，Lightened CNN 模型是一种更快的方法，它使用不常见的激活函数 (MFM) 而不是 ReLU。此外，从 FC6 层获得的 VGG-Face 特征对姿态变化表现出更好的鲁棒性，而从 FC7 层获得的 VGG-Face 特征对光照变化具有更好的鲁棒性。总体而言，尽管近年来深度学习取得了显着的进展，但基于人工智能的方法，不匹配条件下的人脸识别——特别是当手头的任务可用的数据量有限时——仍然是一个具有挑战性的问题。

Acknowledgements

这项工作得到了 TUBITAK 项目 no.113E067 和第七欧盟框架计划内的居里夫人 FP7 整合补助金的支持。

References

[1] T. Ahonen、A. Hadid 和 M. Pietikäinen。具有局部二进制模式的人脸识别。计算机科学讲义，第 469-481 页。2004年.[2] A. Asthana、T. K. Marks、M. J. Jones、K. H. Tieu 和 M. Ro-hith。通过 3dpose 归一化实现全自动姿势不变人脸识别。2011年计算机视觉国际会议，2011.[3] O. Déniz、G. Bueno、J. Salido 和 F. D. la Torre。使用定向梯度直方图进行人脸识别。模式识别快报，32(12):1598–1603, 2011.[4] H.K. Ekenel 和 R. Stiefelhagen。为什么面部遮挡是一个具有挑战性的问题？《生物识别技术进展》，第 299-308 页。2009.[5] H. 高、H. K. 埃克内尔和 R. Stiefelhagen。基于局部外观的人脸识别的姿势归一化。《生物识别技术进展》，第 32-41 页。2009.[6] C.耿和X.江。使用 sift 特征进行人脸识别。第 16 届 IEEE 国际图像处理会议 (ICIP)，2009 年。[7] A. Georghiades、P. Belhumeur 和 D. Kriegman。从少到多：用于可变照明和姿势下人脸识别的照明锥模型。IEEE 模式分析和机器学习汇刊，23(6):643–660, 2001.[8] T. Hassner、S. Harel、E. Paz 和 R. Enbar。无约束图像中的有效脸部正面化。IEEE 计算机视觉与模式识别会议，2015 年。[9] G. E. Hinton、S. Osindero 和 Y.-W. 嗯。深度信念网络的快速学习算法。神经计算，18(7):1527–1554, 2006.[10] G. B. Huang、M. Ramesh、T. Berg 和 E. Learned-Miller。野外标记面孔：用于研究无约束环境中人脸识别的数据库。技术报告，技术报告 07-49，马萨诸塞大学阿默斯特分校，2007 年。

Amherst, 2007.

- [11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama 和 T. Darrell. Caffe: 用于快速特征嵌入的卷积架构。第 22 届 ACM 国际多媒体会议论文集, 第 675-678 页, 2014 年。[12] K. Kavukcuoglu, P. Sermanet, Y.-L. 布罗, K. 格雷戈尔, M. 马蒂厄和 Y. LeCun. 学习视觉识别的卷积特征层次结构。神经信息处理系统进展, 第 1090-1098 页, 2010 年。[13] A. Krizhevsky, I. Sutskever 和 G. E. Hinton. 使用深度卷积神经网络进行 ImageNet 分类。神经信息处理系统 (NIPS) 进展, 第 1106-1114 页, 2012 年。[14] Y. LeCun, Y. Bengio 和 G. Hinton. 深度学习。自然, 521 (7553): 436-444, 2015。[15] K.-C. Lee, J. Ho 和 D. Kriegman. 获取可变照明下人脸识别的线性子空间。IEEE 模式分析和机器智能交易, 27(5):684-698, 2005。[16] M. Lin, Q. Chen 和 S. Yan. 网络中的网络。计算研究知识库 (CoRR), 2013. arXiv:1312.4400。[17] A.M. 马丁内斯. ar 人脸数据库。CVC 技术报告, 1998 年 24 月。[18] D. Miller, I. Kemelmacher-Shlizerman 和 S. M. Seitz. Megaface: 用于大规模识别的百万张面孔。计算研究知识库 (CoRR), 2015. arXiv: 1505.02108。[19] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, 和 Z. Zhu. 通过遮挡字典学习进行鲁棒人脸识别。模式识别, 47(4):1559-1572, 2014。[20] S.J. 潘和 Q. 杨. 迁移学习综述。IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359, 2010。[21] O. M. Parkhi, A. Vedaldi 和 A. Zisserman. 深度人脸识别。英国机器视觉会议, 2015 年。[22] P. 菲利普斯, P. 弗林, T. 斯克鲁格斯, K. 鲍耶, J. 张, K. 霍夫曼, J. 马克斯, J. 敏和 W. 沃雷克. 人脸识别大挑战概述。IEEE 计算机学会计算机视觉和模式识别会议, 2005 年。[23] P. Phillips, H. Wechsler, J. Huang 和 P. J. Rauss. 人脸识别算法的 FERET 数据库和评估程序。图像与视觉计算, 16(5):295-306, 1998。[24] F. Schroff, D. Kalenichenko 和 J. Philbin. FaceNet: 用于人脸识别和聚类的统一嵌入。见 IEEE 计算机视觉和模式识别会议, 2015 年。[25] T. Sim, S. Baker 和 M. Bsat. CMU 姿态、照明和表情 (PIE) 数据库。第五届 IEEE 自动人脸手势识别国际会议, 2002 年。[26] K. 西蒙扬和 A. 齐瑟曼. 用于大规模图像识别的非常深的卷积网络。计算研究知识库 (CoRR), 2014. arXiv: 1409.1556。[27] J. Sivic, M. Everingham 和 A. Zisserman. “你是谁?” - 从视频中学习特定于个人的分类器。IEEE 计算机视觉和模式识别会议, 2009 年。[28] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever 和 R. 萨拉胡季诺夫. Dropout: 防止神经网络过度拟合的简单方法。机器学习研究杂志, 第 1929-1958 页, 2014 年。[29] 孙 Y. 王 X. 和唐 X. 用于面部验证的混合深度学习。IEEE 国际计算机视觉会议, 2013 年。[30] 孙 Y. 王 X. 和唐 X. 深度学习通过预测 10,000 个类来表示人脸。IEEE 计算机视觉和模式识别会议, 2014 年。[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke 和 A. Rabinovich. 深入研究卷积。IEEE 计算机视觉和模式识别会议, 2015 年。[32] Y. Taigman, M. Yang, M. Ranzato 和 L. Wolf. DeepFace: 缩小人脸验证方面与人类水平表现的差距。IEEE 计算机视觉和模式识别会议, 2014 年。[33] X. Tan 和 B. Triggs. 增强的局部纹理特征集, 用于困难照明条件下的人脸识别。IEEE 图像处理交易, 19(6):1635-1650, 2010。[34] L. 托里和 J. 沙夫利克. 迁移学习。机器学习应用和趋势, 第 242-264 页。2009。[35] L. Wolf, T. Hassner 和 I. Maoz. 具有匹配背景相似性的无约束视频中的人脸识别。见 IEEE 计算机视觉和模式识别会议, 2011 年。[36] J. Wright, A. Ganesh, Z. Zhou, A. Wagner 和 Y. Ma. Demo: 通过稀疏表示进行鲁棒人脸识别。第 8 届 IEEE 自动人脸和手势识别国际会议, 2008 年。[37] X. 吴, R. 何, Z. 孙. 用于深度面部表示的轻量化 CNN。计算研究知识库 (CoRR), 2015. arXiv: 1511.02683。[38] D. Yi, Z. Lei, S. Liao 和 S. Z. Li. 从头开始学习人脸表示。计算研究知识库 (CoRR), 2014. arXiv: 1411.7923。[39] J. Yosinski, J. Clune, Y. Bengio 和 H. Lipson. 深度神经网络中的特征可转移性如何? 《神经信息处理系统进展》, 第 3320-3328 页, 2014 年。[40] M. D. Zeiler 和 R. Fergus. 可视化和理解卷积网络。计算机视觉 - ECCV 2014, 第 818-833 页。施普林格, 2014。[41] S. Zhou, G. Aggarwal, R. Chellappa 和 D. Jacobs. 线性朗伯物体的外观表征、广义光度立体和光照不变的面部识别。IEEE 模式分析和机器智能汇刊, 29(2):230-245, 2007 年。