

端到端深度人脸识别的要素:最新进展综述

中国石海林上海大学人工智能研究所

丹增, 上海大学, 中国梅涛, 京东人工智能研究所, 中国

人脸识别是计算机视觉领域最基本和最长期的课题之一。随着深度卷积神经网络和大规模数据集的发展, 深度人脸识别取得了显著的进步, 并被广泛应用于现实世界。给定自然图像或视频帧作为输入, 端到端深度人脸识别系统输出人脸特征用于识别。为了实现这一点, 整个系统通常由三个关键元素构建: 人脸检测、人脸预处理和人脸表示。人脸检测在图像或帧中定位人脸。然后, 进行人脸预处理, 以将人脸校准到标准视图, 并将它们裁剪到标准像素大小。最后, 在人脸表示阶段, 从预处理后的人脸中提取识别特征进行识别。所有这三个要素都由深度卷积神经网络实现。本文综述了端到端深度人脸识别的各个方面的最新进展, 因为蓬勃发展的深度学习技术极大地提高了它们的能力。首先, 我们介绍了端到端深度人脸识别的概述, 如上所述, 它包括人脸检测、人脸预处理和人脸表示。然后, 我们回顾了基于深度学习的每个元素的进展, 分别涵盖了许多方面, 如最新的算法设计, 评估指标, 数据集, 性能比较, 现有的挑战, 以及未来研究的方向。我们希望这项调查能给人们带来有益的思考, 以便更好地理解端到端人脸识别的全貌, 并以系统的方式进行更深入的探索。

附加关键词和短语: 深度卷积神经网络、人脸识别、人脸检测、人脸预处理、人脸表示。

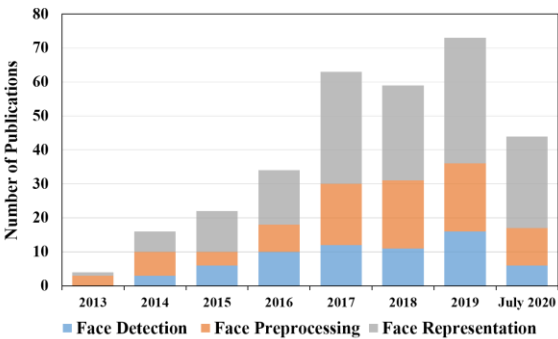
一 正式介绍

人脸识别是计算机视觉中广泛研究的课题。在现有的人体生物特征识别技术中, 人脸识别在现实世界中的应用最为广泛, 如身份认证和监控系统。根据数据的形态, 人脸识别可以分为基于 2D 图像的方法和基于三维扫描的方法, 这两种方法在发展和应用上都有很大的不同。此外, 随着深度卷积神经网络的巨大进步, 基于深度学习的方法在包括人脸识别在内的各种计算机视觉任务中取得了显著的性能改进。本文主要研究基于 2D 图像的端到端深度人脸识别, 它以自然图像或视频帧为输入, 提取每个人脸的深度特征作为输出。我们全面回顾了端到端深度人脸识别的最新进展。具体来说, 端到端深度人脸识别系统由三个关键要素组成: 人脸检测、人脸预处理和人脸表示。下面, 我们对每个元素进行简单介绍。

人脸检测是端到端人脸识别的第一步。它旨在定位自然图像或视频帧中的人脸区域。在深度学习时代之前, 其中一个先驱

[§] Equal contribution. This work was performed at JD AI Research.

[†] Corresponding author.



图。1.端到端深度人脸识别的要素从 2013 年到 2020 年 7 月的出版号。

人脸检测的工作是 Viola-Jones [232]人脸检测器，它利用具有哈尔特征的 AdaBoost 分类器来构建级联结构。随后，后续方法探索有效的手工特征[7, 165, 172]和各种分类器[16, 127, 155]以提高检测性能。此外，一些方法[59, 281]采用可变形零件模型(DPM)进行人脸检测。人们可以参考[304]来全面了解传统的人脸检测方法。近年来，随着分布式神经网络的发展，基于深度学习的人脸检测得到了广泛的研究。通过使用 DCNN 从大规模数据中学习，人脸检测器对各种条件变得更加鲁棒，例如大的面部姿态和遮挡。

接下来，人脸预处理是指将自然人脸校准到一个规范的视图，并将其裁剪到一个规范化的像素大小，以便于后续的人脸表示计算任务。这是人脸识别系统必不可少的中间步骤。在这篇综述中，我们介绍了两种主要的人脸预处理方法，。面部对齐和正面化。一般来说，面部对齐利用空间变换将面部扭曲到参考面部标志的标准位置。因此，面部标志点定位对于人脸对齐是必要的。大多数传统的面部标志定位工作集中在生成方法[36, 37]或鉴别方法[158, 354]，并且有一些关于它们的详尽调查[100, 249, 371]。一些方法不是利用面部标志，而是直接从输入的面部标志生成对齐的面部。此外，人脸正面化研究从非正面输入合成正面人脸，这通常用于处理大姿态人脸识别。

在人脸表示阶段，从预处理后的人脸图像中提取识别特征进行识别。这是人脸识别最后也是最核心的一步。在早期的研究中，许多方法通过将人脸图像投影到低维子空间来计算人脸表示，例如特征脸[[229]和鱼眼脸[[12]。后来，更多手工制作的基于局部描述符的方法 [3, 137]在人脸表示中占了上风。关于这些传统方法的详细评论，可以参考[6, 233, 312]。近年来，人脸表示得益于离散余弦神经网络的发展，在高性能人脸识别方面取得了很大的进步。

这项调查的重点是回顾和分析最近的进展，在每个元素的端到端的深度人脸识别。一个重要的事实是，人脸识别的性能取决于所有元素的贡献(即，。人脸检测、预处理和表示)。换句话说，任何一个元素中的劣势都会成为木桶中最短的一块，损害最终的性能。为了建立一个高性能的端到端人脸识别系统，有必要讨论整体框架的每个元素及其相互影响。在过去的二十年里，已经发表了许多人脸识别调查。表 1 总结了我们的调查和现有调查之间的主要差异。

表 1。人脸识别的代表性调查

Title	Year	Description
-------	------	-------------

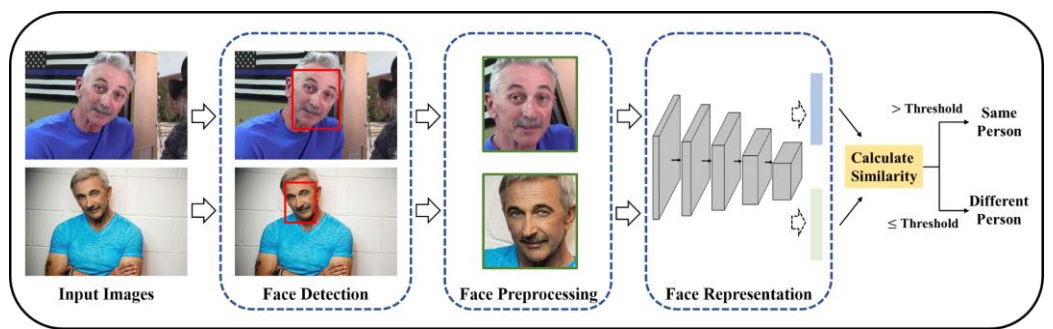
人脸识别:文献综述[233]	2003	人脸识别中基于图像和视频的传统方法。不覆盖深度人脸识别。
根据每个人的单个图像进行人脸识别:调查[312]	2006	该方法解决了人脸识别中的单样本问题，不覆盖深度人脸识别。
3D 和多模式 3D+2D 人脸识别方法和挑战综述[15]	2006	3D 和多模态人脸识别综述，不涉及深度人脸识别。
光照不变的人脸识别:综述[369]	2007	专注于光照不变的人脸识别任务，不覆盖深度人脸识别。
人脸识别技术综述[6]	2009	传统的人脸识别方法基于不同的模态人脸数据，没有覆盖深层人脸识别。
姿态不变人脸识别综述[48]	2016	专注于姿态不变的人脸识别任务。
三维人脸识别的局部特征方法综述[206]	2017	基于特征提取的三维人脸识别方法综述。
理解面孔的深度学习[181]	2018	提供端到端深度人脸识别的简要概述，不覆盖近期作品。
深度人脸识别:综述[246]	2018	专注于深度面孔表征学习。
人脸识别的过去、现在和未来:综述[2]	2020	回顾 2D 和 3D 人脸识别，不包括端到端深度人脸识别。

具体来说，有一些关于人脸识别的调查[6, 233, 312]，但不包括基于深度学习的方法，因为它们是在深度学习时代之前早期发表的；此外，一些调查侧重于 3D 人脸识别[15, 206]和特定任务[48, 369]。相反，我们关注的是实际应用中最需要的 2D 人脸识别。Ranjan 等人。[181]简要概述了这三个要素，但没有涵盖过去几年迅速发展的最新技术。如图 2 所示。1，这些年来发表的作品数量急剧增加。王等。[246]提出了一个关于深度人脸识别的系统综述，其中他们主要集中于深度人脸表征学习，而训练损失的分类是次优的。例如，他们通过基于欧几里德距离的损失、基于角度/余弦边缘的损失、基于最大值的损失及其变化来对深度人脸表示的监督学习进行分类，然而，几乎所有基于角度/余弦边缘的损失都是作为最大值损失的变化而不是单个集合来实现的。相反，我们建议用三个子集对训练监督进行更合理的分类，即，。分类、特征嵌入和混合方法(见第 5.2 节)。最近，Insaf 等人。[2]回顾了 2D 和 3D 人脸识别从传统到深度学习的时代，而在人脸表示方面的范围仍然有限。总之，面部识别技术需要在覆盖端到端管道的所有元素的广泛范围内进行系统地审查，而现有的调查很少完成这项工作。

因此，我们分别系统地回顾了端到端人脸识别中基于深度学习的各种方法。对每个元素的回顾涵盖许多方面:算法设计、评估指标、数据集、性能比较、剩余挑战和未来研究的有希望的方向。我们希望这项调查能给人们带来有益的思考，以便更好地理解端到端人脸识别的全貌，并以系统的方式进行更深入的探索。

具体而言，主要贡献可归纳如下:

- 我们全面综述了端到端深度人脸识别的最新进展，包括人脸检测、人脸预处理、人脸表示。
- 我们从算法设计、评估指标、数据集和性能比较等多个方面讨论了这三个要素。



图。2.端到端深度人脸识别系统的标准管道。首先，人脸检测阶段旨在定位输入图像上的人脸区域。然后，进行人脸预处理以将检测到的人脸归一化为规范视图。最后，人脸表示致力于提取人脸识别的判别特征。

- 我们进一步收集每个元素的现有挑战和有希望的方向，以促进未来的研究，并从整体框架的角度讨论未来的趋势。

2 概述

一个典型的端到端深度人脸识别系统包括三个基本元素:人脸检测、人脸预处理和人脸表示，如图。2.首先，人脸检测定位输入图像上的人脸区域。然后，进行人脸预处理，以将检测到的人脸归一化为标准布局。最后，人脸表示致力于从前置人脸中提取有区别的特征。这些特征用于计算它们之间的相似性，以便决定这些人脸是否属于同一身份。

我们根据这三个要素来构建身体部分(第 3、4、5 节)，其中的每一个都是一个涵盖了大量计算机视觉文献的研究课题。在这一节中，我们简要地概述了这三个元素，并在接下来的正文部分深入探讨了它们。

2.1 人脸检测

人脸检测是人脸识别系统的第一步。给定一幅输入图像，人脸检测的目的是找到图像中的所有人脸，并给出包围盒的坐标

有信心得分。人脸检测的主要挑战包括不同的分辨率、比例、姿态、光照、遮挡等.传统方法侧重于设计区分面部和背景区域的手工特征。随着深度学习的发展，深度特征已经广泛应用于人脸检测。在第三节中，我们从多个维度对基于深度学习的人脸检测方法进行了分类，包括多阶段、单阶段、基于锚、无锚、多任务学习、中央处理器实时和面向问题的方法。通常，多阶段和单阶段方法的分类标准依赖于面部检测器是否生成候选框，然后接下来的一个或多个阶段进一步细化候选以进行准确预测。大多数基于锚的方法在特征图上预设若干锚，然后对这些锚进行分类和回归。主播在这个节目中起着至关重要的作用。最近，另一个程序，即，。无锚设计由于其灵活性和高效性，在目标检测中受到越来越多的关注。因此，我们还讨论了无锚方法，并与基于锚的方法进行了比较。另外，由于人脸检测是人脸识别系统中的首要步骤，人脸检测器的计算效率

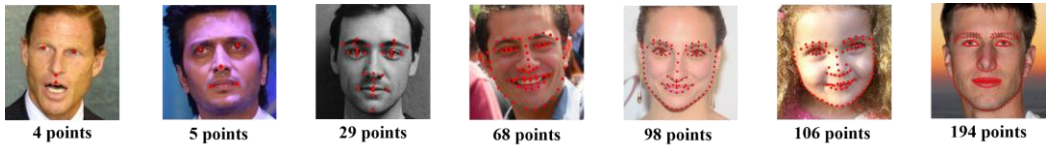


图. 3.不同版本面部标志的可视化。4 点和 5 点标志通常用于面部对齐。

在现实应用中很重要。虽然检测器可以通过 DCNNs 实现良好的性能，但是部署重负载网络是不切实际的，尤其是在非 GPU 设备上。因此，我们介绍了实际应用的中央处理器实时方法。当然，我们不应该忽略另一组面向问题的人脸检测方法，因为它们有明确的动机来应对特定的挑战。从上述角度出发，我们在第三节对现有的深度人脸检测方法进行了深入的讨论。值得注意的是，类别之间存在技术重叠，因为如上所述，分类是从多个角度建立的。这将有助于我们更好地识别基于深度学习的人脸检测方法。

2.2 人脸预处理

在第二阶段，人脸预处理旨在将检测到的人脸校准到标准视图(即。*面部对齐或正面化*)，这是提高面部识别的端到端性能的基本过程。由于人脸呈现出规则的结构，其中面部部分(眼睛、鼻子、嘴巴等。在有恒定排列的情况下，人脸的对齐对于后续的人脸识别特征计算有很大的好处。通常，面对齐利用空间变换技术将面校准到标准化布局。对于大多数现有的面部对齐方法，面部标志或所谓的面部关键点

(如图. 3)，是必不可少的，因为它们作为相似变换或仿射变换的参考而涉及的。因此，人脸标志点定位是人脸对齐的前提。基于离散余弦神经网络的人脸标志点定位方法可分为三类:基于坐标回归的方法、基于热图回归的方法和基于三维模型拟合的方法。基于坐标回归的方法以地标坐标为回归目标，旨在从输入人脸图像到地标坐标的非线性映射。此外，基于热图回归的方法分别输出对应于每个地标的似然响应图。此外，基于 3D 模型拟合的方法从 2D 图像预测 3D 面部形状，然后将其投影到图像平面上以获得 2D 地标。在不依赖面部标志的情况下，几种方法可以通过学习变换参数直接从输入中输出对齐的面部。此外，人脸正面化技术还可以应用于人脸预处理，通过从非正面视图合成保持身份的正面人脸来处理大的姿态变化。面对齐和面正面化都是将不受约束的面校准到规范视图并促进后续面表示的常见做法。我们将在第 4 节回顾这套方法。

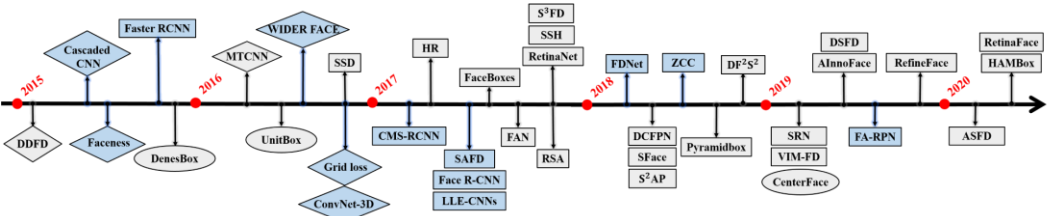
2.3 面部表征

人脸表示作为人脸识别系统的关键步骤，致力于学习深度人脸模型，并利用深度人脸模型从预处理后的人脸中提取特征进行识别。这些特征用于计算匹配人脸的相似度。在第 5 节中，我们回顾了基于深度学习的人脸识别方法。我们从网络结构和训练监督两个学习人脸表示的重要方面来回顾这些方法。对于网络体系结构，我们介绍了为广泛的计算机视觉任务而设计的一般体系结构，以及专门用于人脸表示的特殊体系结构。对于训练监督，我们主要介绍了四种方案，包括分类、特征嵌入、混合和半监督方案。具体来说，分类方案将人脸表征学习作为一个分类问题(每个 ID 视为一个类)，一般采用 softmax 损失及其变种作为训练监督。特征嵌入方案通过根据样本的身份优化样本之间的距离来学习表示。混合方案是指联合使用分类和特征嵌入来训练深层人脸模型。这三种方案都侧重于监督训练。近年来，深度半监督人脸表示学习由于能够利用大量未标记的人脸数据来提高人脸表示的学习

效果而受到越来越多的关注。此外，我们还提出了几个具体的人脸识别场景，包括跨领域，低镜头学习和基于视频的场景。

3 人脸检测

人脸检测是端到端人脸识别系统的第一步，其目的是从输入图像中定位人脸区域。近年来，随着深度卷积神经网络的发展，深度人脸检测得到了广泛的研究。在这一部分，我们首先对现有的人脸检测深度学习方法进行分类和比较。接下来，我们介绍了几种流行的人脸检测数据集和常用的评价指标。最后，我们描述了一些现有的挑战和有希望的未来方向。



图。4.代表性人脸检测方法的发展。蓝色和灰色代表多阶段和单阶段方法；根据锚的用法，矩形、椭圆形和菱形表示基于锚、无锚和其他方法。这些方法的参考文献可参见表 2。

3.1 人脸检测的分类

为了给当前的人脸检测方法一个清晰的分类，我们将它们分成七组，即。多阶段、单阶段、基于锚点、无锚点、多任务风格、CPU 实时和面向问题的方法。这些集合不一定是排他的，因为我们从多个角度建立分类。例如，多阶段和单阶段方法通过检测建议和阶段学习的证据来区分，而基于锚和无锚的方法根据锚的使用来区分。因此，检测方法可以是单阶段的，也可以同时基于锚。这并不妨碍我们的演示，但有助于读者确定他们感兴趣的方法。

3.1.1 多阶段方法。按照由粗到细的方式或由建议到细化的策略，基于多阶段的检测器首先生成多个候选框，然后细化

表 2。深度人脸检测方法的分类

类别	描述	方法
多阶段	检测器首先生成候选框，然后接下来的一个或多个阶段细化用于人脸检测的候选框。	Faceness [288], HyperFace [180], STN [26], ConvNet-3D [128], WIDER FACE [290], [76], CMS-RCNN [357], 万等[234], 人脸更快 RCNN [98], DeepIR [212], 电网损耗[173], Face R-CNN [94], Face R-FCN [255], ZCC [356], FDNet [309], FA-RPN [169], 级联 CNN [123], MTCNN [319], 秦等[178], -CNN[64], PCN [197], [306]
单级	检测器一次完成从特征图的人脸分类和包围盒回归。	DDFD [58], HR [86], face box[327], SSH [168], S3FD [328], DCFPN [329], RetinaFace [41], FAN [244], FANet [318], RSA [149], S2AP [207], PyramidBox [224], DF2S2 [225], SFace [243], DSFD [124], RefineFace [323], SRN [31], PyramidBox++ [130], VIM-FD [337], ISRN [326], AlnooFace [313], ASFD [308], HAMBox [151], DenseBox [90], UnitBox [298], CenterFace [280]

基于锚点	检测器在特征地图上部署多个密集锚，然后对这些锚进行分类和回归。	万等[234]，人脸更快 RCNN [98]，RSA [149]，人脸 R-CNN [94]，FDNet [309]，DeepIR [212]，[76]，SSH [168]，S3FD [328]，DCFPN [329]，face box[327]，FAN [244]，FANet [318]，PyramidBox [224]，ZCC [356]，S2AP [207]，DF2S2 [225]，SFace [243]，视网膜 Face [41]，DSFD [124]，refineeface[323]，SRN [31]，VIM-FD [337]，PyramidBox+[130]，FA-RPN [169]，ISRN [326]
无锚	探测器直接找脸，不用预设锚点。	DenseBox [90]，UnitBox [298]，CenterFace [280]
多任务学习	检测器联合学习分类和包围盒回归 在一个框架中具有其他任务(例如，地标定位)。	STN [26]，ConvNet-3D [128]，HyperFace [180]，MTCNN [319]，Face R-CNN [94]，视网膜面[41]，DF2S2 [225]，FLDet [367]，金字塔盒++ [130]，中心面[280]
CPU 实时	检测器可以在单个中央处理器内核上实时运行，以获得视频图像分辨率。	Cascade CNN [123]，STN [26]，MTCNN [319]，DCFPN [329]，face box[327]，PCN [197]，retainface[41]，FLDet [367]，FBI [99]，PPN [306]，CenterFace [280]
面向问题	检测器专注于解决人脸检测中的特定挑战，如微小人脸、遮挡人脸、旋转人脸和模糊人脸。	HR [86]，SSH [168]，S3FD [328]，白等[8]，PyramidBox [224]，网损 [173]，FAN [244]，-CNNs [64]，PCN [197]

一个或多个附加阶段的候选人。第一阶段使用滑动窗口以给定的比例提出候选包围盒，而后一阶段拒绝误报并以更高的分辨率细化剩余的包围盒。在这种情况下，级联结构[123，197，306，319]自然是从粗到细人脸检测的有效解决方案。

人脸检测可以看作是一般物体检测的一个具体目标。因此，许多作品[26，64，94，98，128，169，180，212，309，357]继承了一般物体探测器的显著成就。例如，fast R-CNN[185]是一个经典而有效的检测框架，它采用一个区域建议网络(RPN)在第一阶段用一组密集的锚盒生成区域建议，然后在第二阶段对建议进行细化。基于提出的改进方案，CMS-RCNN [357]提出了一种基于上下文多尺度区域的 CNN，利用人脸和人体周围的特征来实现小人脸检测。几部著作[94，98，212，309]从改进的损失设计、在线硬例挖掘、多尺度训练和测试策略、特征拼接等多个方面改进了用于人脸检测的 fast R-CNN。

像细化阶段的努力一样，建议阶段[26，128，169]中的改进也引起了很大兴趣，例如使用辅助面部信息，或者在不同比例的锚之间共享分类参数。此外，由于 CNN 缺乏尺度不变性的能力，它需要额外的参数和计算成本来处理人脸尺度变化，这是人脸检测的一个关键挑战。因此，估计面部比例是一种合理的做法[76，207]，有助于在适当的比例下检测面部。

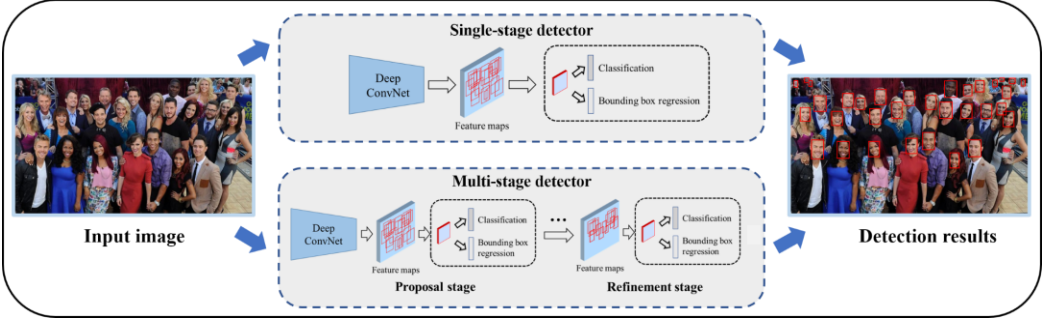


图. 5.单级和多级人脸检测器的说明。单阶段检测器直接从整个特征图中完成人脸检测，而多阶段检测器采用一个建议阶段来生成候选，并采用一个或多个阶段来细化这些候选。

除了建模，如何训练多级探测器是另一个有趣的话题。多级检测器通常是逐级训练的，因为每一级都受其自身目标的监督。这可能会导致劣质优化。为了解决这个问题，为级联 CNN [123 和更快的 R-CNN 设计了联合训练策略[178，以实现端到端优化和更好的人脸检测性能。

3.1.2 单阶段方法。单阶段方法直接从整个特征图中完成候选分类和包围盒回归，而不涉及建议阶段。

单级的经典结构来自于一个名为单次多箱探测器(SSD) [142]的通用物体探测器。与 RPN 类似，SSD 在要素地图上预设了不同比例和比例的密集锚框。SSD 是对象检测中的一个流行框架，因为它在保持可比精度的同时，运行速度比 R-CNN 快得多。因此，许多开发人员在应用程序中使用固态硬盘进行人脸检测。然而，固态硬盘对大规模变化不够鲁棒，尤其是对小人脸。后来，研究了许多方法[99, 224, 327–329]来修改用于人脸检测的 SSD。比如张等人。[328]设计了一个比例公平的版本，以从不同比例的表面获得足够的特征。

许多最先进的人脸检测器采用特征金字塔网络(FPN) [132]，该网络由自上而下的体系结构组成，具有跳跃连接，并合并高级和低级特征进行检测。高层特征图有更多的语义信息，而低层特征图有更小的感受野但有更详细的局部信息。特征融合保留了两者的优点，在检测目标方面取得了很大的进步

范围很广。因此，许多单级人脸检测器[31, 41, 124, 130, 168, 224, 225, 244, 318, 326]是利用 FPN 的优势开发的。这些方法不仅通过 FPN 处理了人脸检测中的尺度问题，而且试图解决原 FPN 的固有缺点，如感受野的冲突。特殊特征融合操作[124, 130, 224]也适用于处理人脸检测的困难情况，如模糊和遮挡人脸。

虽然单阶段法具有效率高的优点，但其检测精度低于两阶段法。部分原因是密集锚带来的正负失衡问题，而提案细化方案能够缓解这个问题。因此，RefineDet [325]在其网络中建立了一个锚点细化模块，以消除大量的负面影响。受 RefineDet 的启发，SRN [31]提出了一种选择性的两步分类和回归方法；在低层进行两步分类以减少分类器的搜索空间，在高层进行两步回归以获得准确的位置。后来，VIM-FD [337]、ISRN [326]、AlnoFace [313]和 RefineFace [323]利用训练数据增强、改进的特征提取器和训练监督、锚分配和匹配策略、多尺度测试策略等多种有效技术改进了 SRN。

前面提到的大多数方法都需要预设锚点进行人脸检测，而一些有代表性的单级检测器，如 DenseBox [90]、UnitBox [298]和 CenterFace [280]则是在没有预设锚点的情况下完成检测的。我们将在下一小节中以无锚类型的形式介绍它们。

3.1.3 基于锚和无锚的方法。如表 2 所示，由于长期开发和卓越的性能，目前大多数人脸检测器都是基于锚的。通常，我们在特征图上预设密集锚，然后对这些锚进行一次或多次分类和包围盒回归，最后输出接受的锚作为检测结果。因此，锚的分配和匹配策略对检测精度至关重要。比如 S3FD [328]提出的锚匹配的尺度补偿，可以有效提高微小和外人脸的召回率。此外，S3FD 使用最大输出标记机制来减少大量否定，这也是基于锚的机制中经常出现的问题。朱等。[356]引入了预期最大重叠分数(EMO)来评估匹配锚的质量，并提出了几种技术来鼓励真正的肯定达到高 EMO 分数。由于人脸的比例分布在训练数据集中是不平衡的，组采样[164]根据其比例对锚框进行排序，并在训练期间为每个组保持相同数量的样本。最近，HAMBox [151]提出了一种在线锚补偿策略，以帮助检测外表面，利用不匹配的锚，但提供有利的回归。

基于锚点的方法在人脸检测领域占据了主导地位，但它们也有一些弱点。超参数(例如，比例、步幅、比率、数量)需要针对每个特定数据集进行仔细调整，这限制了检测器的泛化能力。此外，密集锚增加了计算成本，带来了正负锚的不平衡问题。

无锚方法[120, 226, 355]在一般目标检测中引起越来越多的关注。关于人脸检测，近年来出现了一定的开创性工作。DenseBox [90]和 UnitBox [298]试图预测像素边界框和置信度得分。此外，中心人脸[280]将人脸检测视为一个广义的关键点估计任务，它预测人脸中心点和特征图中包围盒的大小。简而言之，无锚检测器摆脱了预设锚，实现了更好的泛化能力。在检测精度方面，需要进一步探索对误报的鲁棒性和训练过程的稳定性。

3.1.4 多任务学习方法。多任务学习在计算机视觉领域得到了广泛的研究。一般来说，基于多任务学习的方法被设计为通过共享视觉表示来与其他相关任务一起解决问题。在这里，我们引入了多任务学习方法，将人脸检测器与相关的人脸任务或辅助监督分支一起训练，以丰富特征表示和检测鲁棒性。

许多多任务学习方法[26, 90, 128, 280, 310, 319, 367]探索了人脸检测和面部标志点定位的联合学习。其中，MTCNN [319]是最具代表性的一个，它通过一个三级级联网络，利用了人脸包围盒与地标之间的内在相关性。随后，HyperFace [180]融合了低级特征和高级特征，同时执行四项任务，包括人脸检测、面部标志定位、性别分类和姿态估计。基于表 3。CPU 实时人脸检测器运行效率。“准确度(%)”表示 Fddb 1000 个假阳性的真阳性率。

方法	中央处理器型号	速度(FPS)	准确度(%)
面板盒[327]	E5-2660v3@2.60GHz	20	96.0
STN [26]	i7-4770K	30	-
DCFPN [329]	2.60GHz	30	-
联邦调查局[99]	E5-2660v3@2.60GHz	20	96.8
PCN [197]	3.40GHz	29	-
PPN [306]	i5	60	-
视网膜面[41]	i7-6700K	60	-
中心面[280]	i7-6700@2.60GHz	30	98.0

视网膜网[133]，视网膜面[41]在一个框架中集成了人脸检测，人脸标志定位和密集三维人脸回归。

从多任务例程中，我们可以看到面部检测器可以从相关的面部任务中受益。此外，某些方法[94, 130, 225, 243]利用了辅助监督分支，如分段分支、无锚分支等.这些分支用于增强人脸检测的训练。

3.1.5 CPU 实时方法。虽然最先进的人脸检测器在准确性方面取得了巨大的成功，但在现实世界的应用中，尤其是在非 GPU 设备上，它们的效率还不够。根据推理速度对 CPU 的要求，我们在这里收集 CPU 实时人脸检测器[26, 41, 197, 280, 306, 327, 329, 367]，方便检索。这些检测器能够在一个具有 VGA 分辨率输入图像的单个 CPU 上每秒运行至少 20 帧。表 3 显示了它们的运行效率。

通常，当输入大小或通道变大时，卷积运算会消耗更多时间。为了加快速度，轻量级主干[41, 41]和快速消化的卷积层[327, 329]是关于网络体系结构的常见实践。知识蒸馏是提

升轻量级人脸检测器性能的另一选择[99]。此外，引入了感兴趣区域(RoI)卷积[26]，仅计算 RoI 区域的卷积。

3.1.6 面向问题的方法。在这一小节中，我们强调了一些面向问题的方法，这些方法是针对人脸检测中的各种特定挑战而设计的。

检测大范围的人脸是人脸检测中长期存在的挑战。许多方法[86, 164, 168, 224, 328]被设计用于尺度不变的人脸检测，包括尺度选择、多尺度检测、密集锚点设置、尺度平衡策略等.此外，生成清晰的超分辨率图像[8]是定位模糊和微小人脸的可行方法。

部分可见的面(即。遮挡)损害了传统面部检测器的性能。许多方法[64, 173, 244, 288]利用了特定的技术来检测遮挡人脸。例如，面部度[288]根据面部部分的出现和空间排列来计算置信度得分，因此被遮挡的面部将被高置信度地召回。FAN [244]为数据增加生成遮挡人脸，并引入锚级注意力算法来强调来自面部区域的特征。

同样，面内旋转是阻碍人脸检测的现有因素。为了解决这个问题，PCN [197]针对朝向直立的旋转逐步校准候选物。

3.2 评估指标和数据集

表 4。人脸检测的流行数据集。

数据集	年	#图像	#脸	每个图像的人脸数量	描述
训练					
阿尔夫[114]	2011	21,997	25,993	1.18	人脸检测训练源。
宽面[290]	2016	16k	199k	12.43	最大的人脸检测数据集。
测试					
Fddb [231]	2010	2,845	5,171	1.82	一个经典的人脸检测基准。
AFW [363]	2012	205	473	2.31	多重面部标注。
PASCAL 面孔[281]	2014	851	1,335	1.57	面部变化大。
MALF [282]	2015	5,250	11,931	2.27	细粒度评估。
宽面[290]	2016	16k	194k	12.12	最大的人脸检测数据集。
MAFA [64]	2017	30,811	35,806	1.16	假面检测。

3.2.1 指标。与一般的目标检测算法一样，平均精度是一种广泛用于评估人脸检测性能的指标。由检测精度回忆曲线导出。为了获得精度和召回率，联合交集(IoU)用于测量预测包围盒(Boxp)和地面真实包围盒(Boxdt)的重叠，其公式如下：

面积(Boxp ∪ Boxdt)

IoU = _____ . (1)

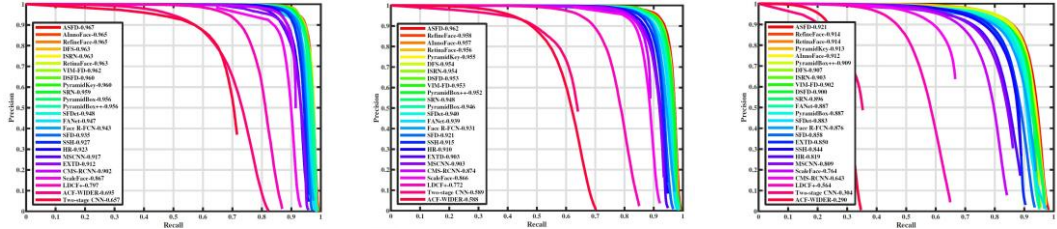
面积(Boxp ∩ Boxdt)

人脸检测器的输出包含一个置信度得分和一个预测的包围盒。具有置信阈值的置信分数用于确定是否接受该预测。当 IoU 大于预设阈值(对于人脸检测通常为 0.5)时，可接受的预测可被视为真阳性(TP)。否则，将被视为假阳性(FP)。在确定了目标点和目标点后，通过改变置信阈值可以得到精度-召回率曲线。在一系列均匀间隔的离散回忆水平上，平均精确度被计算出来[56]。

接收机工作特性(ROC)曲线也用于评估 Fddb [231]中的人脸检测性能。Fddb 提出了两个度量标准(即,。离散和连续)来绘制真阳性率相对于假阳性率的 ROC 曲线。对于离散度量,如果 IoU 大于 0.5,则预测的边界框将被视为真正;对于连续度量,分数等于匹配的 IoU,反映了预测与基本事实的吻合程度。此外,每秒帧数(FPS)用于衡量实际应用中的运行效率。

3.2.2 数据集。我们介绍了几种广泛使用的人脸检测数据集。它们的统计数据见表 4.其中,Fddb-[231]是无约束人脸检测的经典数据集,包括低分辨率人脸、遮挡人脸和困难的姿态变化。值得注意的是,Fddb 使用椭圆而不是矩形框作为人脸标注。AFW [363]是从 Flickr 收集的,包括杂乱的背景和大的变化,如年龄、太阳镜、化妆品和表情。Pascal 人脸数据集[281]中的图像取自 PASCAL 人布局数据集[57]。MALF [282]是为野外人脸检测的细粒度评估而设计的。MAFA [64]是一个掩蔽人脸检测基准,具有不同的方向和遮挡程度。

以上数据集用于性能评估,ALFW [114]数据集仅用于训练。此外,WIDER FACE 数据集[290]具有用于训练、验证和测试的子集。每个子集都有三个难度级别:简单、中等和困难,用 EdgeBox [368]进行评估。在过去的几年里,WIDER FACE 促进了人脸检测的发展,它提供了大量的训练数据和一个具有挑战性的测试基准



测试:简单测试:中等 测试:硬

图. 6.在 WIDER FACE 测试子集上的精度-召回曲线[289]。排名由[308]更新。

变化。图. 图 6 显示了最先进的方法在 WIDER FACE 测试子集上的精度-召回曲线。

3.3 挑战和未来工作

在这一部分中,我们从多个方面对基于深度学习的人脸检测进行了综述。我们可以看到,人脸检测技术近年来取得了很大的进步。人脸检测的进步也促进了其他人脸任务,如人脸识别、人脸属性分析等。

然而,仍然存在某些困难和挑战。

- 运行效率:最先进的检测器已经取得了很大的进步,但仍然需要在检测精度和效率之间进行权衡。例如,在许多应用程序中,调整输入图像的大小是检测器加速的一种常见做法,但它也会损害小脸的召回。
- 图像变化:在不受约束的条件下,例如监控视频,检测器往往会错过姿态和遮挡变化较大的人脸,而不同的图像背景往往会导致误报。此外,检测大范围的人脸也是一个很大的挑战。

人脸检测是深度人脸识别系统中最先进的技术。考虑到剩下的问题和现有技术,我们收集了几个有前途的未来工作方向的深度人脸检测。

- 有效统一的锚点设置:现有的基于锚点的方法从多方面设计锚点设置，如赋值匹配策略[124, 130, 151, 224, 328]，属性调优[31, 327, 356]，采样策略[164]。调谐良好的锚可能会限制人脸检测器的泛化能力。因此，探索一种有效的、统一的、可用于不同应用需求的锚定设置是值得的。
 - 无锚人脸检测框架:无锚检测器[120, 226, 355]在一般物体检测中受到越来越多的关注，因为它们显示出灵活的设计和更大的泛化能力潜力。然而，有少量的作品[90, 280, 298]探索了人脸检测的无锚机制。无锚框架的优势可以进一步推动人脸检测的发展。
 - 更高效的检测框架:由于人脸检测是人脸识别系统中的首要步骤，人脸检测器的计算效率对于实时性至关重要
- 世界应用。许多人脸检测器在基于重型骨干网的基础上实现了很高的检测精度，而轻量级检测器的效率在移动和嵌入式设备上更为重要。因此，在保持检测精度的同时，设计一个更高效的检测框架至关重要。

表 5。人脸预处理方法的分类。

Category		Description	Method
Landmark-based Face Alignment	Coordinate regression	Take the landmark coordinates as the target of regression, and learn the nonlinear mapping from the input face image to the landmark coordinates.	DCNC [216], EFLL [353], CFAN [317], TCDCN [339], RAR [275], MDM [228], TSR [154], JFA [279], RDN [139], SIR [273], TCNN [270], DSRN [163], SBR [52], Wing loss [61], AAN [301], Lai <i>et al.</i> [119], ODN [359], HyperFace [180], MTCNN [319], RetinaFace [41], FLDet [367], CenterFace [280]
	Heatmap regression	Output the likelihood response maps of each landmark.	CALE [17], RED [176], Yang <i>et al.</i> [102], JMFA [44], FAN [18], LAB [266], SAN [50], FALGCN [162], PCD-CNN [116], ELT [82], HRNet [242], Zhang <i>et al.</i> [316], SA [153], FHR [222], Awing loss [251], DeCaFA [38], HSLE [370], FAB [211], KDN [28], Dong <i>et al.</i> [51], Robinson <i>et al.</i> [186], LUVLi [117], PropagationNet [92]
	3D model fitting	Infer a 3D face shape from a 2D face image, and then project it back to the image plane to obtain 2D landmarks.	LPFA [103], 3DDFA [360], FacePoseNet [24], PIFASCNN [104], DeFA [147], RDR [274], Bhagavatula <i>et al.</i> [13], Zhang <i>et al.</i> [314], PR-Net [60], PAFA [125]
Landmark-free Face Alignment		Directly output aligned faces without the explicit use of landmark.	Hayat <i>et al.</i> [77], E2e [349], ReST [265], GridFace [352], Wei <i>et al.</i> [258], RDCFace [340]
Face Frontalization		Synthesize frontal faces from nonfrontal views.	FIP [364], Zhang <i>et al.</i> [336], Zhu <i>et al.</i> [366], MVP [365], SPAE [106], CPF [96], Yang <i>et al.</i> [284], HPEN [361], Cole <i>et al.</i> [35], DR-GAN [227], FF-GAN [295], TP-GAN [91], PIM [342], CAPG-GAN [88], CR-GAN [299], UV-GAN [43], 3D-PIM [343], PW-GAN [324], A3FCNN [338], HF-PIM [21], FNM [177]

4 人脸预处理

给定检测到的人脸区域，人脸预处理旨在将无约束的人脸校准到规范的布局，以便于识别和分析的下游任务，这是端到端人脸识别系统中必不可少的中间步骤。在这一节中，我们回顾了两种主流的人脸预处理方法，包括人脸对齐和人脸正面化。为了消除缩放、旋转和平移的变化，人脸对齐使用空间变换来在面部标志的帮助下将人脸校准到预定义的规范布局。在不依赖面部标志的情况下，其他几种方法仍然可以生成对齐的面部。因此，我们将人脸对齐分为基于标志点的方法和无标志点的方法。此外，人脸正面化旨在从非正面视角合成正面人脸，可用于大姿态人脸识别和人脸数据增强。图。图 7 展示了许多人脸预处理方法的发展。

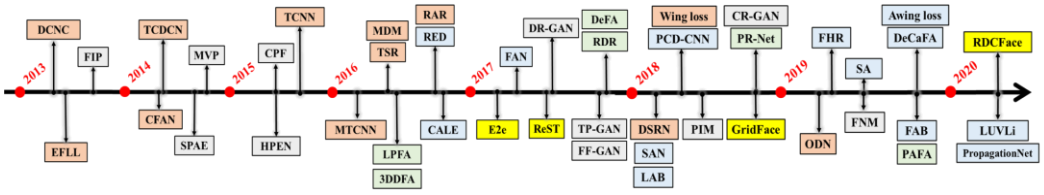


图. 7.人脸预处理典型方法的发展。橙色、蓝色、绿色、黄色和灰色分别代表坐标回归、热图回归、3D 模型拟合、无标志人脸对齐方法和人脸正面化。这些方法的参考文献可参见表 5。

4.1 基于地标的人脸对齐

基于标志点的面部对齐利用空间变换，通过将面部标志点作为参考，将面部校准到预定义的规范布局。因此，准确的人脸标志点定位是基于标志点对齐的核心任务。根据现有的地标定位方法，我们将基于地标的对齐方法分为三个子类别。基于坐标回归的方法、基于热图回归的方法和基于 3D 模型拟合的方法。

4.1.1 坐标回归。基于坐标回归的方法通过神经网络将地标坐标作为回归的目标。换句话说，他们专注于学习从人脸图像到地标坐标向量的非线性映射。

按照由粗到细的方式，大多数方法[154, 216, 317, 353]采用级联回归来逐步细化地标坐标的先前结果。此外，由于递归神经网络(RNN)能够在级联细化过程中建模历史信息，RAR [275]和 MDM [228]一起使用 CNN 和 RNN 来提取全局特征和细化预测。

多任务学习也是一种常见的例程，以促进与面部任务相关的地标定位。许多方法[41, 180, 280, 319, 367]被设计来联合检测面部和预测面部标志。事实上，这些方法最初是为人脸检测任务设计的，并同时输出一组面部标志。TCDCN [339]和 JFA [279]专为面部标志定位而设计，受益于面部辅助属性，如表情、性别和头部姿势。

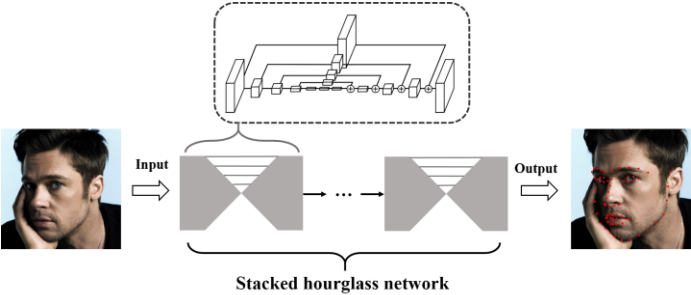
基于级联回归的方法在解决人脸标志点定位问题上显示出巨大的优势。然而，由于它们采用多阶段回归来改进预测，所以性能很大程度上取决于初始预测，这意味着不适当的形状初始化将可能导致不准确的预测。此外，多元回归增加了计算成本。为了解决这些缺点，一些方法[139, 163, 270, 273, 301, 359]开发了新的用于面部标志定位的回归方法，例如自迭代回归、直接形状回归和推理-决策回归等。

坐标回归的学习主要采用 L1、L2 或平滑的 L1 损失函数。这些目标驱动学习过程更加关注大误差样本，促进向精确预测收敛；另一方面，它对异常值具有很高的敏感性。为了解决这个问题，冯等人。[61]通过放大具有小范围或中等范围误差的样本的影响，改进了损失函数，即机翼损失。另一个问题是，关于地标的欧几里德距离的优化可能导致训练和测试之间的差距，因为测试度量通常采用归一化平均误差(NME)。为了解决这个问题，赖等人。[119]提出了增强的归一化平均误差损失，以优化地标定位网络。

上述方法研究了静止图像上的人脸标志点定位。对于视频人脸地标定位，如何利用跨帧的时间信息变得十分必要。TSTN-[[138]开发了一个双流架构，它从单个帧中定位地标，并捕捉时间一致性以进行细化。此外，丁苯橡胶[52]提出，鼓励光流一致性的检测标志时，培训与视频数据。

4.1.2 热图回归。与坐标回归相反，基于热图回归的方法输出每个地标的似然响应图。热图回归的早期探索[17]研究了如何用 DCNNs 聚集得分图和细化预测。后来，纽厄尔等人。[171]设计了堆叠沙漏(HG)网络，为人体姿态估计生成热图。沙漏是一种自下而上和自上而下的

体系结构，与中间监管一起在瓶颈块的深层堆栈中发挥着重要作用。图。图 8 是堆叠沙漏网络的示意图。堆叠沙漏网络在人体姿态估计方面取得了巨大的成功。由于面部标志定位任务类似于人的姿态估计，



图。8.用于面部标志定位的堆叠沙漏网络[171]的图示。在每个沙漏结构中，宽度(即，。特征通道)是一致的，并且方框代表剩余的模块。

许多最近的作品[18, 44, 92, 102, 251, 316]采用了堆叠沙漏网络进行面部标志定位，大大提高了最先进的性能。

全卷积网络的密集像素分类为热图回归任务提供了一种有效的方法。沙漏结构可以看作是全卷积网络的一个例子。除了沙漏结构，许多有效的网络架构[38, 50, 116, 162, 242]都是为热图回归而新设计的。其中，DecaA[38]利用堆叠的全卷积 U 网来保持空间分辨率，并利用地标方向的注意力图来提取当前估计周围的局部信息。高分辨率网络(HRNet) [242]旨在保持高分辨率表示，并显示其在地标类任务中的优势。

由于前景和背景的像素数量不平衡，上述为坐标回归设计的机翼损失不能保证热图回归的收敛性。为了解决这个问题，王等人。[251]对前景像素处罚较多，对背景像素处罚较少；类似地，传播网[92]提出了一个焦点机翼损失其调节每个小批量样品的失重。

一些面部标志的定义不明确，比如脸颊上的标志，导致不同标注者的标注不一致。此外，被遮挡的面部区域中的界标也会导致不精确的标注。这两个问题导致语义偏差，从而降低地标定位的性能。许多方法[28, 117, 153, 153, 266, 370]致力于缓解这些问题。面部边界热图[266]是提供面部几何结构以减少语义歧义的好选择。刘等认为语义歧义是一种干扰性注释。[153]提供了用概率模型估计真实地标位置的另一种途径。最近，KDN [28]和鲁夫里[117]提出同时估计面部标志和预测的不确定性。这种不确定性可用于识别人脸对齐失败的图像。

考虑到构建具有精确标注的大规模面部标志数据集的昂贵成本，一些方法[51, 82, 186]探索了用于面部标志定位的半监督学习。Honari 等人。[82]提出了一种等变地标变换损失，以使预测与同一图像上的不同变换一致。基于对抗性学习机制，罗宾逊等人。[186]应用了一个生成器来为未标记的数据生成热图，并应用了一个鉴别器来区分生成的热图和真实的热图。此外，在未标记的数据[51 上分配伪界标标记]是用于促进界标定位的半监督学习的又一个有前途的例程。

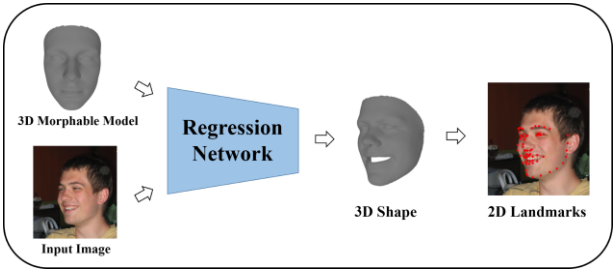


图. 9.面向面对齐的三维模型拟合过程。密集的 3D 可变形模型用于将 2D 面建模为 3D 网格。回归网络估计三维形状和投影矩阵的参数，然后将三维形状投影到图像平面上以获得 2D 界标。

对于视频帧中的人脸对齐，设计了几种方法[[176, 211, 222]来解决视频中的具体问题。例如，为了减轻高分辨率视频中调整大小的热图的量化误差，分数热图回归[222]通过对热图中的多个点进行采样来估计分数坐标。为了解决视频中的运动模糊问题，FAB [211]利用结构感知去模糊模块，通过保持相邻帧之间的结构一致性来恢复清晰的人脸。

4.1.3 3D 模型拟合。考虑到 2D 面部标志和三维人脸形状之间的显式关系，基于三维模型拟合的方法从 2D 图像重建三维人脸形状，然后将其投影到图像平面上以获得 2D 标志。与估计一组界标的常规 2D 方法相比，基于 3D 模型拟合的方法能够用成千上万个顶点的 3D 模型来拟合面部，并且将它们与大姿态对准。

由于级联回归是估计模型参数的一种有效方式，LPFA [103]和 3DDFA [360]将级联 CNN 回归器与密集 3D 可变形模型[14]相结合来估计 3D 面部形状参数。此外，他们都设计了特殊的功能，使回归稳健的姿态变化。此外，DeFA [147]不仅使用地标作为回归约束，还使用 3D 形状的投影轮廓和局部描述符。尽管有许多优点，级联的神经网络经常缺乏端到端的训练。作为迂回，Jourabloo 等人。[104]试图通过单个 CNN 拟合 3D 人脸模型，该模型由几个可视化块组成，以根据先前块的特征和预测来调整 3D 形状和投影矩阵。

虽然上述方法充分利用了三维模型的优势，但是不同的面部形状会导致 2D 标志点定位不准确，尤其是在三维形状系数稀疏的情况下。为了解决这个问题，RDR [274]提出通过动态表达模型来拟合 3D 人脸，并使用循环 3D-2D 双重学习模型来交替细化 3D 人脸模型和 2D 地标。除了对 3D 人脸形状的参数进行回归外，法斯特-TRFA [13]和法斯特珀森[24]还估算了渲染一般 3D 人脸模型的不同视图时的扭曲参数。一些方法[60, 314]从面部形状的 3D 坐标回归地标。例如，PR-Net [60]采用紫外位置图来记录具有语义对应的面部形状的 3D 坐标，其中包括预定义的界标。

4.2 无地标面对齐

无标志点的面对齐方法将对齐变换处理集成到 DCNNs 并输出对齐的人脸，而不依赖于面部标志。这套方法一般

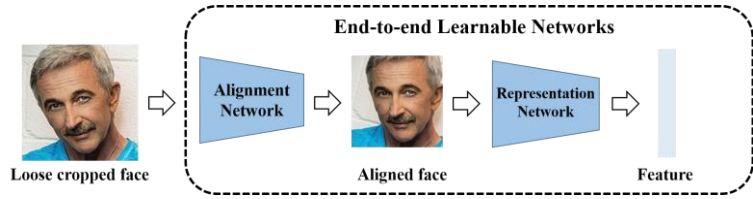


图. 10.无标志面对齐框架的图示。面部对齐和表示形成了一个完整的可训练网络。

将空间变换网络(Spatial-TN) [97]用于几何扭曲，其中变换参数通过端到端训练来学习。基于 Spatial-TN，Hayat 等。[77]和钟等。[349]利用随后的面部表示模块来联合优化面部对齐。由于面部变化因各种因素而相当复杂，因此设计了一些方法[265，352]来提高空间总氮的变形能力。例如，在 ReST [265]中，在每次迭代中基于先前变换的面部执行进一步的变换，逐步处理大的面部变化。此外，人脸图像的径向失真是另一个常见的问题，这是由广角相机带来的。[340]提出了一种级联网络

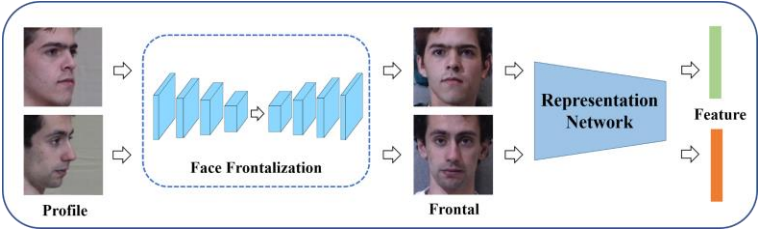
其以端到端的方式学习针对径向透镜失真、面部对准变换和面部表示的校正。最近，魏等人。[258]提供了对面部对齐效果的综合分析。结果表明，过多的对齐会影响后续的人脸识别，而人脸识别对特征映射上的对齐具有鲁棒性。因此，他们提出在人脸识别的共同监督下，在特征地图上学习人脸对齐。

4.3 正面化

在不受控制的环境中，姿态变化是人脸识别的一个严重问题。为了消除姿态的影响，人脸正面化旨在从非正面视图中合成身份保持正面人脸。图. 图 11 是具有面部表征下游任务的面部正面化的图示。在前一节中，我们介绍了一些基于 3D 模型拟合的方法[60，103，104，147，360]，这些方法可以通过旋转 3D 人脸模型并将其投影回 2D 平面来构建正面化的 3D 人脸。除了 3D 程序之外，许多方法[96，106，336，364–366]使用具有编码器-解码器架构的深度学习神经网络来恢复规范视图中的的人脸。对他们来说，保证下游任务识别性能的身份保持特性并不是一个容易实现的目标。

最近，由于生成对抗网络，高质量的图像生成取得了很大进展[68]。许多人脸正面化方法[91，177，338，342]受益于氮化镓的合成。其中，黄等人。[91]开发了一个双向生成对抗网络(TP-GAN)来分别推断正面视图中的全局和局部面部结构。走向认同型的一代，赵等。[342]提出了一种姿态不变模型，它由一个人脸正面化子网络和一个区分学习子网络组成，用于相互学习人脸正面化和人脸表示。除了从侧面视图恢复正面视图，几种基于 GAN 的方法[88，227，299]以各种姿势渲染旋转的面。比如 Tran 等人。[227]提出了一种解纠缠的表征学习遗传神经网络，它生成身份相关表征，并利用姿态编码合成任意姿态下的身份保持人脸。

此外，一些方法[21，43，295，324，343]将 GAN 与 3D 人脸模型相结合，以利用人脸先验知识进行人脸合成。求助于 3D 人脸模型，他们利用



图。11.人脸正面化旨在从侧面视图合成正面人脸，以方便后续任务，如人脸表示计算。面部图像取自多张照片[69]。

全局形状和外观信息，以提高正面图像的质量，特别是当从大姿态视图合成时。

4.4 评估指标和数据集

介绍了人脸预处理中常用的评价指标和数据集，特别是基于标志点的对齐。如本小节的下一部分所述，大多数基于陆地标记的方法采用定量度量，如标准化平均误差；而人脸正面化通常研究合成正面的视觉质量，缺乏标准的评价指标。此外，一些方法采用面向大姿态人脸识别的评价，我们将在人脸表示部分描述它们的度量。

4.4.1 度量。对于面部标志点定位，广泛使用的评估度量是通过归一化平均误差(NME)来测量点到点欧几里德距离，归一化平均误差可以定义为：

$$NME = \frac{1}{M} \sum_{k=1}^M \frac{\|p_k - d_k\|_2}{d_k} \tag{2}$$

其中 M 是界标的数量，p_k 和 d_k 表示面部界标的预测和地面真实坐标，k 表示界标的指数，d 表示由眉眼距离或瞳孔间距离定义的归一化距离。d 用于缓解由于不同的人脸比例和较大的姿态而导致的测量异常。小 NME 意味着该方法具有更好的性能。

累积误差分布曲线也用作评估标准。CED 是 NME 的一个分销功能。CED 的纵轴代表误差值小于或等于横轴误差值的测试图像的比例。曲线下面积(AUC)也提供了算法在给定误差下如何执行的参考：

$$AUC^\alpha = \int_0^\alpha f(e) de, \tag{3}$$

其中，α 是对应于积分计算上限的给定误差，e 是渐进归一化误差，f(e)是指 CED 曲线。AUC 越大，表示性能越好。基于 CED 曲线，失败率可以用来衡量算法的性能和鲁棒性，它表示测试集中 NME 大于阈值的样本的百分比。

4.4.2 数据集。面部标志数据集可以根据约束条件和野生条件进行排序。表 6 给出了这些数据集的统计数据。

CMU 多姿态，照明和表情(多派)[69]是在约束条件下最大的面部数据集，它为 337 个对象提供 15 个预定义的姿态，19 个照明

表 6。人脸标志数据集的统计。“-”指的是分割训练和测试集的非官方协议。

数据集	年	#总计	#培训	#测试	#点	描述
-----	---	-----	-----	-----	----	----

多重 PIE [69]	2008	755,370	-	-	68	可控条件下最大的人脸数据集。
LFPW [11]	2010	2,845	-	-	35	从不受控制的环境中拍摄的图像。
阿尔夫[114]	2011	24,386	20,000	4,386	21	一个大规模的面部标志数据集。
AFW [363]	2012	473	-	-	6	多重面部标注。
海伦[121]	2012	2,330	2,000	330	194	提供密集的地标注释。
COFW [19]	2013	1,852	1,345	507	29	包含遮挡的脸。
300-W [189]	2013	3,837	3,148	689	68	最常用的面部标志数据集。
300-大众[195]	2015	114	50	64	68	一个视频面部标志数据集。
Menpo [303]	2017	28,273	12,014	16,259	68	包含半正面和侧面。
WFLW [266]	2018	10,000	7,500	2,500	98	多注解，大变奏。
JD-landmark [150]	2019	15,393	13,393	2,000	106	覆盖大的面部变化。

条件和 6 个面部表情。带注释的面部标志是正面面部的 68 点和侧面面部的 39 点。因为它包含了多种多样的姿态变化，所以多 PIE 通常作为人脸正面化的数据集。

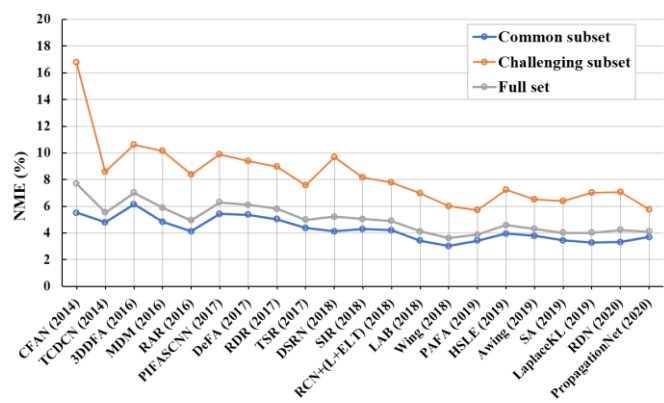
此外，更多的野生数据集[11, 19, 114, 121, 150, 189, 195, 266, 303, 363]被提议用于面部标志定位。其中，300-W [188, 189]是最常用的数据集，它遵循 Multi-PIE 的标注配置，重新标注 LFPW、AFW、HELEN 中的图像，以及新采集的数据集 iBug。图 12 显示了不同地标定位算法在 300 瓦测试集上的性能比较。除此之外，Menpo [303]是一个大规模的人脸地标数据集，人脸地标定位的案例比较困难。JD-landmark [150]用 106 个人脸界标对人脸图像进行标注，提供更多的人脸成分结构信息。

上述数据集侧重于静态图像，而 300-VW [195]提供 50 个视频剪辑用于训练，64 个用于视频中面部标志定位的测试。

4.5 挑战和未来工作

在本调查中，人脸预处理(即。面对齐和面正面化)指的是将不受约束的面标准化为规范视图，以便于下游任务。人脸对齐的目的是将人脸在空间上转换到一个规范的位置，而人脸正面化的重点是从非正面视角合成保持身份的正面人脸。这两种方法都可以作为提高人脸识别性能的中间步骤。尽管他们的发展取得了显著的进步，但他们的研究仍在继续改进。以下是人脸对齐和人脸正面化的主要挑战。

- 面部变化:当在各种极端变化下工作时，例如运动模糊、严重遮挡、大姿态、低照度等，面部标志定位仍然不够鲁棒。
- 运行时效率:在许多实际应用中，人脸识别系统为中间过程分配较低的运行时预算，尤其是在移动和嵌入式设备上的部署。
- 标注模糊性:由于某些面部标志点的位置模糊，如脸颊上的标志点，标志点的标注模糊是面部标志点数据集中的常见问题。
- 标注粒度:现有的大部分人脸地标数据集提供 68 或 106 点的标注。一般来说，我们希望注释中有更多的标志点来描绘丰富的面部结构。

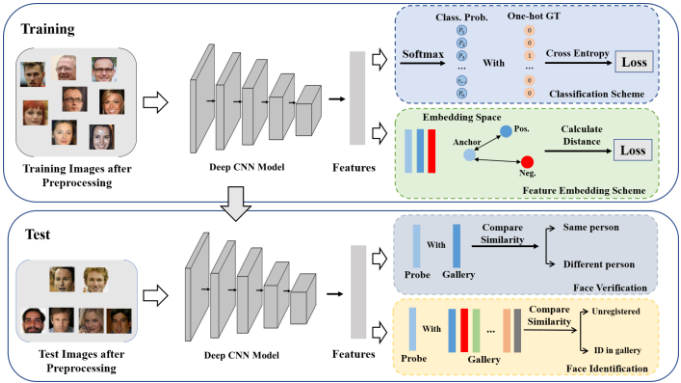


图。12.不同地标定位方法在 300W 测试集上的性能比较。度量单位是 NME (%), 瞳孔间归一化。NME 越低, 表现越好。

- 高保真人脸正面化:高保真人脸正面化要求高分辨率、保持身份的输出, 从侧面看这是一个病态问题。

为了应对这些挑战, 我们提出了一些有希望的未来工作方向。

- 高鲁棒性和高效率:在现实世界的应用中有大量的面部变化, 这需要预处理方法对各种输入面部具有鲁棒性。面部标志定位的效率也需要作为系统中的中间步骤。
- 密集地标定位:大多数数据集使用 68 或 106 个关键点作为标注配置。它们足以进行面部对齐(通常需要 5 个关键点), 但不足以完成复杂的面部分析任务, 如面部运动捕捉。此外, 密集的地标配置将有助于定位更精确的对准所需的关键点。因此, 密集的人脸地标定位数据集和算法对于许多人脸分析任务来说都是值得探索的。
- 基于视频的地标定位:大多数现有的方法在静止图像上完成这项工作。有几种方法[52, 138, 176, 211, 222]侧重于基于视频的面部标志点定位, 这仍然是一个很有前途的研究方向。如何更好地利用时间信息是基于视频的地标定位的主要挑战。其他问题, 如运动模糊, 低分辨率和检测效率, 也是有趣的话题。
- 半监督地标定位:地标定位的广泛研究属于监督学习的范畴, 需要精确标注地标。然而, 获取带有精确标注的大规模数据集既昂贵又低效。正如开创性的工作[51, 52, 82, 186]所探索的那样, 半监督例程对于面部标志定位是一种可行且有价值的解决方案。
- 高保真人脸正面化及其度量:从侧面合成高保真正面仍然是一个具有挑战性的任务。为了评价人脸正面化方法, 目前的实践是通过测量正面化人脸识别的准确性来证明其身份保持能力。视觉质量的度量也需要发展。



图。13.人脸表征训练阶段和测试阶段的流水线。在训练阶段，有两种方案，即，。分类和特征嵌入通常用于学习人脸表示。在测试阶段，人脸验证和人脸识别是主要任务。

5 FACE REPRESENTATION

人脸预处理之后，在人脸表示阶段，目标是将对齐的人脸图像映射到一个特征空间，在该特征空间中，相同身份的特征很接近，而不同身份的特征相距很远。在实际应用中，人脸识别有两个主要任务，即，。人脸验证和人脸识别。人脸验证是指预测一对人脸图像是否属于同一身份。人脸识别可以看作是人脸验证的扩展，其目的是确定人脸的具体身份(即，。探测)在一组身份(即，。画廊)；此外，在开放集人脸识别的情况下，需要一个优先任务，其目标是预测人脸是否属于图库身份之一。

无论是人脸验证还是人脸识别，人脸表示都是用来计算人脸图像之间的相似度。因此，如何学习有区别的人脸表示是人脸识别系统的核心目标。随着分布式神经网络先进的特征学习能力，人脸表示取得了很大的进步。在下文中，我们从两个主要方面对人脸表征的学习方法进行了系统的回顾，。网络体系结构和培训监督。

5.1 网络架构

最近面部表现的改进部分得益于深层建筑设计的进步。因此，我们首先回顾了用于人脸表征学习的网络体系结构的文献。根据设计目的，我们将其分为通用架构和专用架构。通用体系结构首先是普通视觉识别任务的基本和通用设计，然后应用于人脸表征学习。专门的架构包括面向人脸识别的修改或集成设计。

5.1.1 通用架构。随着深度卷积神经网络[33， 65， 78， 85， 115， 203， 221， 238]的高级特征学习能力，深度人脸表示有了很大的进步。其中，AlexNet [115]获得了2012年ImageNet竞赛(ILSVRC)第一名[40]，与传统方法相比取得了显著的提高。然后，VGGNet [203]提出了一个更一般的网络，它用堆叠的 3×3 核代替了大的卷积核，使网络能够在深度上增长。为了扩大网络而不

表 7。人脸表征的分类。

Category		Description	Method
Network Ar- chitectures	General architectures	The basic and universal designs for common visual recognition tasks.	AlexNet [115], VGGNet [203], GoogleNet [221], ResNet [78], Xception [33], DenseNet [65] AttentionNet [238], SENet [85],

			SqueezeNet [95], MobileNet [84], ShuffleNet [334], MobileNetV2 [190], Shufflenetv2 [156]
	Specialized architectures	The modified or ensemble designs oriented to face recognition.	ConvNet-RBM[217], DeepID [213, 215, 219], MM-DFR [47], B-CNN [34], ComparatorNet [276], Contrastive CNN [75], PRN [109], AFRN [108], FANface [283], Sparse ConvNet [220], Light-CNN [267, 268], Mobile-FaceNet [29], Mobiface [55], ShuffleFaceNet [159], Hayat <i>et al.</i> [77], E2e [349], ReST [265], GridFace [352], RDCFace [340], Wei <i>et al.</i> [258], Co-Mining [252], GroupFace [112], DB [20]
Training Supervision	Classification	Considering the face representation learning as a classification task.	DeepFace [223], DeepID series [213, 215, 218, 219], NormFace [239], L2-softmax [179], COCO loss [148], Ring loss [348], L-softmax [144], SphereFace [143], AM-softmax [237], CosFace [241], ArcFace [42], AdaptiveFace [140], Fair loss [136], MV-softmax [253], ArcNeg [300], CurricularFace [93], Adacos [332], P2SGrad [333], NTP [87], UT [351], Co-Mining [252], Shi <i>et al.</i> [201], GroupFace [112], DB [20], RCM loss [272], PFE [199], DUL [25]
	Feature embedding	Optimizing the feature distance according to the label of sample pair.	DeepID2 [213], DeepID2+ [219], DeepID3 [215], FaceNet [192], N-pair loss [205], Lifted structured [208], Smart mining [157], Doppelganger mining [204]
	Hybrid	Applying classification and feature embedding together as the supervisory signals.	DeepID2 [213], DeepID2+ [219], DeepID3 [215], TUA [141], Doppelganger mining [204], Center loss [261], Range loss [331], UniformFace [54], RegularFace [345], UT [351], LBL [362], Circle loss [214]
	Semi-supervised	Exploiting labeled and unlabeled faces for representation learning.	CDP [307], GCN-DS [287], GCN-VE [286], UIR [297], Shi <i>et al.</i> [198], RoyChowdhury <i>et al.</i> [187]
Specific Tasks	Cross-age	Identifying faces across a wide range of ages.	LF-CNNs [260], CAN [278], AFRN [53], DAL [240], AE-CNN [230], OE-CNN [254], IPCGANs [256], LMA [4], Dual CGANs [209], AIM [341]
	Cross-pose	Identifying faces across a wide range of poses.	TP-GAN [91], PIM [342], DREAM [22], DA-GAN [344], DR-GAN [227], UV-GAN [43], CAPG-GAN [88], PAMs [160], AbdAlmageed <i>et al.</i> [1], MvDN [107]
	Racial bias	Addressing the imbalance race distribution of training datasets.	IMAN [248], RL-RBN [247]
	Cross-modality	Performing face recognition on a pair of images captured by different sensing modalities.	Reale <i>et al.</i> [184], HFR-CNNs [191], TRIVET [146], IDR [80], DVR [269], MC-CNN [46], WCNN [81], NAD [122], ADHFR [210], CFC [79], Mittal <i>et al.</i> [166], ForensicFR [62], TDFL [235], E2EPG [320], CASPG [311], DualGAN [293], PS2-MAN [245], DTFS [322], Cascaded-FS [321], PTFS [315]
	Low-shot	Training and testing with the data that have a small number of samples per identity.	SSPP-DAN [83], Guo <i>et al.</i> [73], Choe <i>et al.</i> [32], Hybrid Classifiers [271], Cheng <i>et al.</i> [30], Doppelganger mining [204], Yin <i>et al.</i> [296],
	Video-based	Performing face recognition with video sequences.	TBE-CNN [49], NAN [285], C-FAN [67], FANVFR [152], MARN [66], Rao <i>et al.</i> [182], CFR-CNN [174], ADRL [183], DAC [145]

计算预算的额外增加，GoogleNet [221]开发了一个初始架构来连接由不同感受野的卷积生成的特征图。不久，GoogleNet 被应用于人脸表征学习，即 FaceNet [192]。最近，ResNet [78]提出了一个剩余映射框架，使训练具有数百层的深层网络成为可能。ResNet 是一个现代网络，已经广泛应用于许多视觉任务，包括人脸识别。此外，注意力网[238]将注意力模块引入剩余网络，以在特征推理中利用空间注意力。塞内特[85]提出了挤压和激励块来融合通道方向和空间信息，这也可以看作是一种渠道维度的注意机制。此外，提出了几种轻量级神经网络[84， 95， 156， 190， 334]来实现速度和精度的折衷。在人脸识别文献中，所有这些结构在设计完成后都被用作表征学习的骨干网络。

5.1.2 专业架构。上述架构最初是为一般的视觉任务而提出的。此外，许多作品开发了用于人脸表示学习的专门架构。起初，许多研究[47， 213， 217， 218]试图将多个卷积网络组装在一起，以便从一组面部斑块中学习多个局部特征。给定人脸出现时面部部分(眼睛、鼻子、嘴巴等)有规律的排

列。相对于面部部分，多个网络的这种组合比单个网络更可靠。后来基于双线性 CNN [134]，乔杜里等人。[34]利用双线性结构进行人脸表示学习。再说谢等人。[276]设计了一种端到端结构，即比较器网络，以测量两组可变数量的人脸图像的相似性。类似于多网络组装，比较器网络使用局部面部护理来促进集合式表示学习。韩等。[75]提出了一种对比 CNN，通过生成用于卷积的对比核来处理人脸验证任务，使得特征适应于输入人脸对。康等。[109]引入了成对关系网络来捕获一对局部外观面片之间的关系。此外，AFRN [108]用注意机制改进了成对关系网络。最近，FANFace [283]集成了人脸表示网络和人脸地标定位网络，因此地标的热图将增强识别特征。

为了实现速度-精度的平衡，一些研究[29, 55, 159, 220, 267, 268]集中于开发轻量级架构。为了减少深层网络的参数，稀疏 ConvNet [220]提出了稀疏神经网络连接，它可以从以前学习的密集模型中迭代学习稀疏结构。此外，Light-CNN [267, 268]引入了最大特征映射(MFM)激活函数，以获得比 ReLU 更好的人脸识别泛化能力；基于 MFM，作者开发了在速度和模型尺寸方面具有优势的轻量级架构。MobileFaceNet [29]用全局深度卷积层替换了原始 MobileNet [190]中的全局平均汇集层，因此输出要素可以通过最后一层的空间重要性得到改善。Mobiface [55]通过使用快速下采样和带有扩展层的瓶颈剩余块，对 MobileFaceNet 进行了修改。ShuffleFaceNet [159]通过使用全局深度卷积层和用于实时人脸识别应用的参数校正线性单元(PReLU)，扩展了 ShuffleNetV2 [156]。

值得一提的是，在我们已经在人脸预处理部分介绍的一些无标志点的人脸对齐方法[77, 258, 265, 340, 349, 352]中，可以针对人脸表示学习和人脸对齐的共同目标来优化网络。大部分都是利用空间变换网络[97]标定人脸，后面是表示模块；空间变换网络和表示模块是关于面部识别目标联合学习的。

以下架构开发面向一些特定的目标。为了处理训练数据集中的标签噪声问题，联合挖掘 [252]使用两个对等网络来协作区分噪声标签样本，并获取剩余的干净样本用于训练。Kim 等人。[112]提出了一种称为组脸(GroupFace)的体系结构，它可以学习脸的潜在分组方案，并通过组感知表示来促进识别。为了处理长尾域问题，曹等人提出。[20]引入了剩余平衡映射块，以将面部表示与域相关特征相结合。

5.2 培训监督

除了网络架构之外，训练监督对于人脸表征的学习也起着关键作用。人脸表征学习的监督目标是鼓励相同身份的人脸在特征空间中靠近，不同身份的人脸在特征空间中远离。

按照表示学习的惯例，我们将现有的人脸表示训练监督方法分为监督模式、半监督模式和非监督模式。虽然最近在深度无监督学习方面取得了一些进展

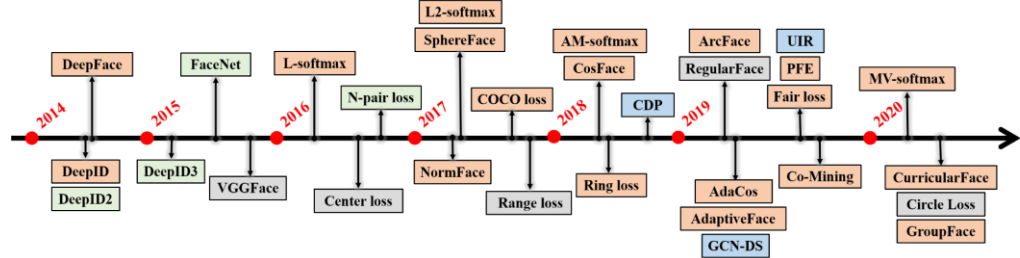


图. 14.人脸表征学习训练监控的发展。橙色、绿色、灰色和蓝色分别代表分类、特征嵌入、混合和半监督方法。详细参考资料可参考表 7。

方法[72, 135, 200, 257]对于人脸聚类，在这篇综述中，我们集中在监督的和半监督的方法，它们构成了最先进的人脸识别的主要文献。图 14 显示了人脸表征学习训练方法的发展。在有监督的方案中，我们可以进一步将现有的作品分成三个子集，即，。分类、特征嵌入和混合方法。分类方法以一个 N 向分类目标完成人脸表示学习，将 N 个类中的每一个看作一个身份。特征嵌入方法旨在优化样本之间相对于身份标签的特征距离，即最大化人与人之间的距离和最小化人与人之间的距离。此外，一些作品采用分类和特征嵌入程序来联合训练表示网络，即混合方法。对于半监督方案，也有一些研究利用标记和未标记的人脸进行表示学习。

5.2.1 分类方案。基于分类的深度人脸表示学习源于一般的对象分类任务。每个类对应于一个身份，该身份包含同一个人的多个面孔。软最大训练损失是分类任务中应用最广泛的监督，它由全连通层、软最大函数和交叉熵损失组成。对于人脸表征学习，DeepFace [223]和 DeepID [218]是利用 softmax 预测大量训练数据同一性概率的先驱。它们的训练损失函数可以表述如下：

普通

$$-\log \sum_c \exp(W_{ci} x_i + b_c)$$

$$-\log \sum_c \exp(W_{ci} x_i + b_c) = -\log \sum_c \exp(W_{ci} x_i + b_c) \quad (4)$$

其中， N 是批量大小， c 是类(恒等式)的数量， W_{ci} 是 FC 层中样本 x_i 的基本真值权重向量， b_c 是偏差项。对数内的项是地面真实类的预测概率。训练目标是最大化这个概率。

在软最大训练损失的基础上，研究了归一化对特征向量和权向量的影响，并用它们之间的余弦相似度重构目标。L2-softmax [179]首次提出将特征向量归一化为位于固定半径的超球面上。此外，NormFace [239]和 COCO loss [148]研究了归一化操作的必要性，并在省略偏置项 b_j 的情况下对特征和权重应用了 L2 归一化约束。为了有效地训练归一化的特征，它们使用比例因子来重新缩放特征和权重之间的余弦相似性。此外，环损失[348]引入了软归一化，可以逐渐将特征的范数约束到目标范数值，而不是直接对特征使用 L2 归一化。总之，归一化软最大值可以重新表述为：

普通

$$-\log \sum_c \exp(s \cos(\theta_{ci}))$$

$$-\log \sum_c \exp(s \cos(\theta_{ci})) = -\log \sum_c \exp(s \cos(\theta_{ci})) \quad (5) \quad e_i + j = 1, j, y_i$$

—— 其中， $\cos(\theta_j)$ 源自加权 L2 归一化的内积 $W_j^T x_i$ ， $w_{jj} = \sum W W_{jj} \sum 2$ ，特征为 x_i
 $= \sum x x_{ii} \sum 2$ ， s 为标度参数， y_i 为样本 x_i 的基本真值标签。.

为了进一步提高类内紧性和类间分离性，几种方法在损失函数中引入了余量。l-softmax[144]用 $\psi(\theta_{yi})$ 代替了基本事实逻辑 $\cos \theta_{yi}$ ， $\psi(\theta_{yi})$ 定义为

$$\psi(\theta_{yi}) = (1 + k \cos(m\theta_{yi}))^{2k}, \quad \theta_{yi} \in [k\pi, (k+1)\pi], \quad (6)$$

其中 m 是正整数的角裕度， k 也是 $k \in [0, m]$ 的整数

修改后的 \logit 使学习目标变得更加困难。类似于 L-softmax，SphereFace [143]在地面真值逻辑 $\cos \theta_{yi}$ 中应用了一个角裕度，以使学习的人脸表示在超球面流形上更具区分性。然而， $\cos m\theta_{yi}$ 中的乘法角裕度在训练期间导致潜在的不稳定收敛。为了克服这个问题，AM-softmax [237]和 CosFace [241]对 \logit ， $\cos \theta_{yi} + m1$ 提出了一个加性裕度惩罚，这带来了更稳定的收敛性。随后，ArcFace [42]在余弦 $\cos \theta_{yi} + m2$ 内引入了一个附加角裕度，它对应于超球流形上的测地距离裕度罚分。以下是 AM-softmax、CosFace 和 ArcFace 的统一表述：

$$L = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{yi} + m2) + m1)}}{e^{s(\cos(\theta_{yi} + m2) + m1)} + \sum_{j \neq i} e^{s(\cos(\theta_{yi} + m2) + m1) + \alpha \cos \theta_{ij}}} \quad (7)$$

其中 $m1 < 0$ 表示 α -软最大值和共面的加性余弦裕度， $m2 > 0$ 表示弧面的加性角裕度。它们易于实现，并且可以实现比原始 softmax 更好的性能。在基于边缘的监督的基础上，AdaptiveFace [140]给出了一个可学习的边缘，该边缘适应于训练数据中的每个身份。目的是解决训练数据集中的不平衡分布问题，这意味着身份具有不同的数量和不同的样本多样性。类似地，公平损失[136]引入了自适应边际策略，该策略应用强化学习来针对不平衡分布问题选择适当的边际。

MV-softmax [253]借助硬样本挖掘策略[133，202]的优势，提出对负(非背景真实)逻辑重新加权，以强调对误分类样本的监督，从而从非背景真实的角度改进人脸表征学习。ArcNeg [300]在 softmax 中用距离感知高斯函数重构负 \logit ，进行硬负挖掘，减弱标签噪声的影响。考虑到训练过程中易样本和难样本变化的相对重要性，课程空间[93]将课程学习的理念引入到人脸表征学习中，前期更加强调简单样本，后期更加强调硬样本。

阿达科斯·[[332]研究了上述基于边际的损失函数公式中规模和边际参数的影响，发现它们对预测概率有很大影响。因此，Adacos 提出了一种统一的自适应方法来重构逻辑和预测概率之间的映射，而不需要预先设置参数。P2SGrad[[333]从训练梯度的角度分析了基于边缘的软最大值损失的影响，并提出在后向传播中用余弦相似度代替分类概率，以达到更好的优化。

高质量的标签标注对于大规模的人脸数据集来说是昂贵的。因此，噪声标签是一个不可避免的问题，某些训练方法追求噪声鲁棒的人脸表示学习。胡等。[87]提出了一种耐噪范式，根据角度分布对监督训练样本重新加权，并在不同训练阶段调整权重。钟等。[351]分析了错误标签和预测之间的差异，并提出了一个抗噪声损失来处理噪声标签问题。假设训

训练数据集上的噪声率是先验知识，联合挖掘[252]使用两个对等网络来找到干净样本并基于损失值丢弃有噪声的样本，并强调在训练过程中对干净样本的监督。

最近，一些方法在人脸表示学习的分类监督方面做得更进一步。为了提高对各种测试条件的泛化能力，石等。[201]提出了一种置信度感知的软最大损失，以强调硬样本，并将特征表示分成子嵌入，以学习互补信息。为了学习固有的分组信息，即，每个组包含一组面部具有共同特征的人，GroupFace [112]引入了一种特征聚集方法来从两个角度组合特征。第一个来自可以看作是个体脸的原始表征的分支；第二个来自自我分组的分支，它指出这张脸最有可能属于哪个组。这样的方案有利于表征学习

分类目标是广泛的身份，包括各种特征。而且，曹等人。[20]提出了一种域平衡机制来解决长尾域分布问题。具体来说，他们提出了一个特征增强模块来提取领域相关的特征，以及一个领域平衡余量来优化尾部领域的特征。为了缓解低比特量化人脸表示模型的性能下降，吴等。[[272]将量化误差视为类误差和个体误差的组合，并提出了旋转一致性裕度损失来减小后者，后者对于人脸表示更为关键。最近的一些著作，如 PFE·[[199]和 DUL·[[25]，提出考虑数据的不确定性来建模深度人脸表示，以解决低质量人脸图像引起的不确定性问题。

$$\frac{1}{2} \max(0, m_d - \|f(x_i) - f(x_j)\|_2)^2 \left\{ \begin{array}{l} \frac{1}{2} \|f(x_i) - f(x_j)\|_2^2 \end{array} \right.$$

5.2.2 特征嵌入方案。特征嵌入方案旨在

根据样本对的标签优化特征距离。如果这对属于相同的身份，即，正对，目标是最小化距离或最大化相似性；否则，即，负对，以最大化距离或最小化相似性。像传统的度量学习方法[193, 259, 277]一样，用于人脸表示的特征嵌入方案获得了相同的目标。对比损失[213, 215, 219, 292]直接优化了成对距离，其余量鼓励正对靠近，负对远离。要最小化的损失函数写为

如果 $y_i = y_j$,

$$L_c = (8) \text{ if } y_i, y_j,$$

其中 $y_i = y_j$ 表示 x_i, x_j 为正对， y_i, y_j 表示负对， $f()$ 为特征嵌入函数， m_d 为非负距离余量。因此，对比损失驱动对所有正对和那些距离小于裕度的负对的监督。余量可以通过固定值来设置，或者根据沿着训练过程的距离分布来更新。

FaceNet [192]首次将三重损失[193, 259]用于深度人脸表征学习。与对比损失不同，三重损失促使正对相对于边缘具有比负对更小的距离，

普通

$$L_t = \frac{1}{2} \|f(x_a) - f(x_{ip})\|_2^2 + \frac{1}{2} \|f(x_a) - f(x_{in})\|_2^2 + m_d, \quad (9)$$

其中 m_d 是距离余量， x_a 表示锚样本， x_{ip} 和 x_{in} 分别表示阳性样本和阴性样本。

对比损失和三重损失每次只考虑一个负面例子，而负面对在训练数据中是丰富的，值得全面参与训练监督。因此，N-对损失[[205]将三重损失推广到多负对形式，在人脸验证和识别上得到了进一步的改进。

与监督分类相比，特征嵌入可以节省 softmax 中 FC 层的参数，尤其是当训练数据集规模较大时。但是训练样本的批量限制了特征嵌入的性能。为了缓解这个问题，[[157, 204,

208]提出了硬样本挖掘策略来丰富每一批中的有效信息，这对于提高特征嵌入的性能至关重要。

5.2.3 混合方法。混合方法是指将分类和特征嵌入一起作为监控信号的方法。DeepID2 [213], DeepID2+ [219], DeepID3 [215]联合利用最大软损失和对比损失来学习人脸表征。另外，刘等人。[[141]提出了一种两阶段训练法，第一阶段是最软损失，第二阶段是三重损失。后者，通过利用类内或类间约束，例如中心损失[261]，均匀面[54]和规则面[345]，几种方法改进了混合方案中的特征嵌入部分。

许多混合方法[331, 351, 362]显示了处理长尾分布式数据的优势，长尾分布式数据是人脸识别中广泛存在的问题。一般来说，分类方案在头部数据上效果很好，但在尾部数据上效果很差，因为它要求每个类都有足够的训练样本。与分类方案相比，特征嵌入方案能够对尾部数据提供补充监督。因此，分类和特征嵌入的结合可以提高对长尾分布式数据的训练。按照这一路径，Range loss [331]在一个小批量中优化了最大类内距离和最近类间距离，以有效地利用尾部数据，钟等人。[351]提出减少尾部数据的类间相似性。而且，朱等人。[362]引入了三阶段(即，。分类-验证-分类)策略，以解决大规模身份识别与每个身份只包含两个样本的点人脸数据的训练问题。孙等。[214]从分类和嵌入学习的统一角度提出了一种圈损失，它将三重损失与交叉熵损失相结合，同时学习具有成对标签和类标签的深层特征。

5.2.4 半监督方案。前述方法集中于监督学习。构建标记数据集需要大量的标注工作，而大量的未标记数据很容易获得。因此，将已标记和未标记的数据一起用于训练深度模型是一个很有吸引力的方向。对于半监督人脸表示学习，假设未标记数据的身份与标记数据不相交，一些现有的工作[286, 287, 297, 307]集中于为未标记数据生成伪标签或最小化未标记数据相对于标记身份的最大软分类概率。此外，考虑到已标记和未标记数据之间的领域差距，师等人。[198]开发了一个领域概括框架，以减少它们在特征空间中的差距。

然而，这些方法假设未标记和标记数据之间的身份不重叠，这在现实场景中通常是不切实际的。因此，通过伪标记方法，重叠身份的未标记样本将被错误地聚类为新类。此外，伪标记数据中的标记噪声是另一个问题。为了解决这些问题，RoyChowdhury 等人。[187]提出通过分布外检测算法将未标记数据分成不相交和重叠类别的样本。此外，他们还设计了一种改进的基于不确定性的训练损失来减轻伪标记数据的标记噪声。

5.3 特定的人脸识别任务

5.3.1 跨域人脸识别。这里，跨域一词指的是一个广义的定义，包括各种因素，如跨年龄和跨姿态的人脸识别。由于深度学习是一种数据驱动的技术，深度网络通常在训练领域工作良好，但在看不见的领域工作不佳。在人脸识别的实际应用中，提高人脸表示跨不同领域因素的泛化能力是至关重要的。下面，我们讨论跨领域人脸识别的某些方面，包括跨年龄、跨姿势、跨种族和跨模态；此外，我们还回顾了目前专门研究跨域人脸识别的方法。

跨年龄:随着年龄的增长，面部外观有很大的类内差异，因此识别不同年龄的人脸是一项具有挑战性的任务。对于这样的跨年龄人脸识别，目前的工作遵循两个方向。在第一个方向，许多方法[53, 230, 240, 254, 260, 278]旨在通过将深层人脸特征分解为与年龄相关和与身份相关的部分来学习年龄不变的人脸表示。第二个方向是基于生成机制。以这种方式，几种人脸老化方法[5, 111, 250]试图合成目标年龄的人脸，但是它们在老化的人脸中

表现出对原始身份的不完全保留。因此，更多的方法[4, 209, 256, 341]集中于改善面部老化期间的身份保持能力。

交叉姿态:在不受约束的条件下，如监控视频，摄像机不能总是捕捉每个出现的对象的正面人脸图像。因此，从正面视图到侧面视图，捕获的人脸具有很大的姿态变化。如前面人脸预处理部分所述，将侧面人脸图像转换为正面姿态是交叉姿态人脸识别的一种可行方法，如 TP-GAN [91]、FF-GAN [295]、PIM [342]和 FNM [177]。然而，生成正面人脸会增加人脸识别系统的负担。曹等。[22]通过将侧面的表示转换为特征空间中的正面视图来缓解这个问题。另一个问题是，训练数据中侧面的数量比正面的数量少得多。因此，一些生成方法[43, 88, 227, 344]提出合成任意姿态的身份保持人脸，以丰富训练数据。此外，某些方法[1, 107, 160]开发了多个特定姿势的深度模型来计算多视图面部表示。

种族偏见:种族偏见是人脸识别中的另一个问题。由于训练数据中不同种族分布的不平衡，深度人脸特征表现出良好的识别性能

在训练数据中，大部分种族比小部分种族更受青睐。最近有几部著作[247, 248]研究过这个问题。王等。[248]构建了具有身份和种族注释的野外人脸数据集(RFW)，该数据集由四个种族子集组成，即高加索人、亚洲人、印度人和非洲人。此外，他们还提出了一个信息最大化适应网络来缓解人脸识别中的种族偏见。后来，在 RL-RBN(基于强化学习的种族平衡网络)[247]的工作中，他们为大比例种族设定了固定的余量，并为小比例种族自动选择最佳余量，以实现针对种族偏见问题的平衡性能。

跨模态:跨模态人脸识别通常是指异构人脸识别，它是通过一对由不同传感模态捕获的输入人脸图像来执行的，例如红外 vs. 可见或草图 vs. 照片。许多传统方法[63, 101, 129, 196, 291]以前已经全面研究过这个课题。如何缩小不同模态之间的领域差距是深度人脸识别方法面临的主要挑战。此外，与常规可视人脸图像的大规模训练样本相比，可用的红外或素描人脸图像数量非常有限。现有的作品主要处理这两个问题。对于红外-可见人脸识别，[146, 184, 191]中的几种方法采用了转移学习机制，即，对大量可见光图像进行预处理，利用近红外数据进行微调。另一组方法[46, 80, 81, 269]旨在将近红外和可见光表示分解为模态特定和模态不变的分量，并将后者用于识别任务。此外，许多方法([79, 122, 210])研究从近红外输入合成可见人脸，然后在可见域执行常规人脸识别算法。

类似地，为了减少照片和草图之间的域差异，在草图-照片人脸识别中采用了转移学习[62, 166, 235]。除此之外，素描面合成是另一个方向。为了将照片转换成草图，一些作品[[311, 320]试图学习具有生成损失的完全卷积网络(FCN)，该网络提供草图面的密集像素合成。然而，合成的草图面经常由于严重的噪声而退化。受图像生成中基于甘的方法[[293, 358]的启发，许多方法[[245, 315, 321, 322]开发了基于甘的框架来恢复真实的面部结构并为草图面部合成保留身份相关信息。

5.3.2 低镜头人脸识别。人脸识别中的低镜头学习侧重于人脸身份的低镜头识别任务，每个身份都有少量的人脸样本。这些方法中的大多数试图解决这种低射问题，主要是在 MS-Celebe-1M 低射学习基准[73]上，该基准在一个基本集合中对于每个身份具有大约 50 到 100 个训练样本，而在一个新集合中对于每个身份只有一个训练样本。目标是识别基本集和新集中的标识。关键的挑战是正确识别每个身份只有一个训练样本的小说集中的主题。

要处理这个问题，Choe 等人。[32]和洪等。[83]提出通过人脸合成增加具有不同属性和姿态的低镜头训练样本的数量。吴等。[271]开发了由 CNN 和最近邻模型组成的混合分类器。Smirnov 等人。[204]提出通过对相似的相似恒等式进行抽样来构造更好的训练小批量。此外，一些方法[30, 73, 296]通过更好的训练监督改进了低镜头人脸识别。通常，在 softmax

分类器中，低镜头类的权值范数小于常规类的权值范数，导致对低镜头类的弱区分。据此，郭等人提出。[73]提出通过对齐低射和规则类的权重向量的范数来优化分类器。尹等。[296]发现由于训练样本不足，低射类的特征分布不足。因此，他们提出了一种特征转移方法来丰富低射类的特征空间，以模拟常规类的特征空间。

5.3.3 视频人脸识别。上述算法集中于基于静止图像的人脸识别。对于视频人脸识别，一种常见的方法[27, 49]是同等地考虑每一帧的重要性，并简单地平均一组深层特征，以获得模板人脸表示。然而，该例程不考虑帧的不同质量和跨帧的时间信息。如何在视频中获得最佳的模板人脸表示是基于视频的人脸识别的主要挑战。许多方法[66, 67, 152, 285]提出用注意力权重或质量分数来聚集帧级特征。再说饶等人。[182]聚集多个帧以合成代表性面部图像。帕尔恰米等人。[174]使用自动编码器生成高质量的标准面来处理低质量帧的问题。

大多数方法独立地利用每个帧的空间信息，而不考虑跨帧的时间信息。因此，一些方法[145, 183]利用顺序注意机制对时空信息进行建模，以利用丰富的相关性并找到视频帧的焦点。

5.4 评估指标和数据集

5.4.1 指标。人脸识别的性能通常在验证和识别两个任务上进行评估，每个任务都有相应的评估指标。具体地，两组样本，即，。画廊和探针。图库是指在人脸识别系统中注册的一组已知身份的人脸，而探针是指在验证或识别中需要识别的一组人脸。在讨论常用的评估指标之前，我们先介绍一些基本概念。人脸识别系统确定

是否通过将它们的相似性与给定阈值进行比较来接受探测面和画廊面的匹配，所述相似性是通过它们的特征之间的一些测量来计算的。具体来说，当探测面和画廊面是相同的身份时，真接受(TA)表示它们的相似度高于阈值，假拒绝(FR)表示它们的相似度低于阈值；如果他们是不同的身份，一个真正的拒绝(TR)意味着他们的相似性低于阈值，一个错误的接受(FA)代表他们的相似性高于阈值。以下是构建评估指标的基本概念。更多细节可以参考[70, 71]。

验证任务:人脸验证经常应用于身份认证系统，用于测量人脸对的相似度。一个人展示他或她的脸，并在画廊中声明登记的身份。然后，系统通过计算所呈现的面部和所声明的面部之间的相似性来确定它是否接受该人是所声明的身份中的同一个人。换句话说，给定一对照片，系统比较两张照片中的人脸，以确定它们是否是同一身份。因此，验证任务可以被视为一对一的人脸匹配过程。假接受率(FAR)和真接受率(TAR)用于评估验证性能。FAR 是相似度在阈值以上的视点替用特效对的分值，可以用 $FAFA+TR$ 来计算；TAR 代表相似度在阈值以上的正品对的分值，可以用 $TATA+FR$ 计算。然后，通过改变阈值，ROC 曲线可以由许多操作点绘制，每个操作点由一对 TAR 对确定。FAR。ROC 曲线(在选定的远场具有焦油值)及其 AUC(即，。曲线下面积)被广泛用于评估面部验证任务的性能。

识别任务:人脸识别任务确定探测人脸是否属于图库集中已注册的身份。为此，探针面需要与画廊组中的每个人进行比较。因此，识别任务也可以称为一对一的人脸匹配。通常，人脸识别面临两个任务，即，。开集和闭集识别。开集识别任务是指探头面不一定是图库集包含的身份，这是实践中的一般情况。真阳性识别率(TPIR)和假阳性识别率(FPIR)是以下两种情况下最常用的指标。第一种情况是指探针对应于画廊集中的注册身份。这种情况叫交配搜索，探针叫交配探针。匹配搜索成功表示真正匹配的等级高于目标等级，同时相似

度高于阈值。在这种情况下，配偶探针被正确地识别为其真实身份，配偶搜索由代表配偶搜索成功试验比例的 **TPIR** 测量。第二种是非配偶搜索，其中探针不对应于任何登记的身份(即，。非配对探针)。非配偶搜索由 **FPIR** 测量，它报告了被错误地识别为登记身份的非配偶探针的比例。通过固定等级和改变阈值，**ROC** 曲线可以由许多操作点绘制，每个操作点由一对 **TPIR** 曲线决定。**FPIR**。**ROC** 曲线
(给定 **FPIR** 下的 **TPIR** 值)用于评估开放集人脸识别任务的性能。

在封闭集场景中，每个探测面的身份都包含在图库集中。累积匹配特征曲线用于评价封闭集人脸识别。临界胶束浓度曲线由一对识别率对确定的操作点绘制。排名。识别率是指探针表面被正确识别为真实身份的比例，因此 **CMC** 曲线报告与给定等级匹配的真实比例，等级 **1** 的识别率是最常用的性能指标。值得注意的是，当我们放宽门槛时，中央军委是 **TPIR** 的一个特例。

5.4.2 数据集。近年来，随着深度人脸识别的发展，另一个促进人脸表征学习的关键因素是不断增长的训练和测试数据集。在过去的几年里，人脸数据集已经变得大规模和多样化，测试场景已经接近现实世界的无约束条件。这里，我们回顾了用于深度人脸识别训练和测试的数据集。它们的统计数据如表 9 所示.训练数据:大规模训练数据集对于学习深度人脸表示至关重要。早期的工作通常使用私有数据集，如 **Deepface** [223]，**FaceNet** [192]，**DeepID** [213]。为了使公平比较成为可能，易等。[292]发布了 **CASIA-WebFace** 数据集，该数据集包含 10, 575 个主题和 494, 414 幅图像，是使用最广泛的训练数据集之一。随后，发布了更多的公共训练数据集，为深度人脸模型的训练提供了丰富的人脸图像。其中，**VGGFace** [175]和 **VGGFace2** [23]包含每个科目的多个训练样本。相比之下，**MS-celebe-1M**[74]、**MegaFace** [110]和 **IMDbFace** [236]为大量受试者提供了每个受试者有限的训练样本。标签噪声是采集大规模人脸数据集时常见的问题。**IMDb-法斯**[236]估计了现有数据集中的噪声分布，表明它们存在严重的噪声问题。他们还发现清理数据可以有效提高人脸识别的性能。

测试数据:就测试而言，野外标记人脸(**LFW**) [89]是无约束环境中人脸识别的经典和最广泛使用的基准。**LFW** 的原始协议包含 3000 个真实的和 3000 个冒名顶替的人脸对，并评估这 6000 对人脸的平均验证精度。到目前为止，最先进的精确度已经在 **LFW** 达到饱和，而 **LFW** 的总样本比原始协议中的要多。基于此，**BLUFR** 等人。[131]提出利用 **LFW** 的所有人脸图像进行大规模无约束人脸识别评估；**SLLFW** [45]用更易变的 **LFW** 代替了负对。此外，**CFP** [194]、**CPLFW** [346]、**CALFW** [347]、**AgeDB** [167]和 **RFW** [248]利用 **LFW** 的相似评价度量来测试具有各种挑战的人脸识别，例如交叉姿势、交叉年龄和多个种族。**MegaFace** [110, 170]和兆对 [39]专注于严格错误接受率下的性能(即，。1e-6 和 1e-9)关于具有百万级干扰物的面部验证和识别任务。表 8 显示了 **LFW** 和 **MegaFace** 上许多方法的性能比较。

上述测试数据集侧重于图像到图像的人脸识别，而 **YouTube 人脸(YTF)** [264]、**IJB-A** [113]、**IJB-B** [262]、**IJB-C** [161]、**IJB-S** [105]和 **QMUL-surface**[236]则充当

表 8。 **LFW** 和 **MegaFace** 挑战赛的表现(%)比较。“训练数据”表示由相应方法使用的训练人脸图像的数量。对于 **MegaFace** 上的评价，“Id.指的是用 1M 干扰物和“Veri”进行的 1 级人脸识别的准确性。指 1e-6 FAR 的面部验证 TAR。带“*”的性能是指对精版 **MegaFace** 的评价。“-”表示作者没有报告相应协议的性能。

Method	Training Data	Architecture	LFW	MegaFace	
				Id.	Veri.
DeepFace [223]	4M	CNN-8	97.35	-	-

DeepID2 [213]	0.3M	CNN-8	99.15	65.21	78.86
FaceNet [192]	400M	GoogleNet	99.63	-	-
VGG Face [175]	2.6M	VGGNet	98.95	64.79	78.32
Center Loss [261]	0.7M	CNN-11	99.28	65.49	80.14
L-Softmax [144]	0.5M	VGGNet-18	99.10	67.12	80.42
SphereFace [143]	0.5M	ResNet-64	99.42	72.72	85.56
Ring loss [348]	3.5M	ResNet-64	99.50	74.93	-
AM-softmax [237]	0.5M	ResNet-20	98.98	72.47	84.44
CosFace [241]	0.5M	ResNet-64	99.42	77.11	89.88
ArcFace [42]	0.5M	ResNet-50	99.53	77.50	92.34
RegularFace [345]	3.1M	ResNet-20	99.61	75.61	91.13
UniformFace [54]	3.8M	ResNet-34	99.80	79.98	95.36
Fair Loss [136]	0.5M	ResNet-50	99.57	77.45	92.87
PFE [199]	4.4M	ResNet-64	99.82	78.95	92.51
TURL [201]	0.6M	ResNet-100	99.78	78.60	95.04
AdaCos [332]	2.35M	ResNet-50	99.73	97.41 [®]	-
P2SGrad [333]	2.35M	ResNet-50	99.82	97.25 [®]	-
AdaptiveFace [140]	5M	ResNet-50	99.62	95.02 [®]	95.61 [®]
Circle Loss [214]	3.6M	ResNet-100	99.73	98.50 [®]	-
DUL [25]	3.6M	ResNet-64	99.83	98.60 [®]	-
DB [20]	5.8M	ResNet-50	99.78	96.35 [®]	96.56 [®]
ArcFace [42]	5.8M	ResNet-100	99.82	98.35 [®]	98.48 [®]
MV-AM-softmax [253]	3.2M	Attention-56	99.78	97.14 [®]	97.57 [®]
CurricularFace [93]	5.8M	ResNet-100	99.80	98.71 [®]	98.64 [®]
GroupFace [112]	5.8M	ResNet-100	99.85	98.74 [®]	98.79 [®]

基于视频的人脸识别的评价基准。特别是，IJB-S 和 QMUL-surveillance 是从真实世界的监控视频中构建的，这比静态图像上的任务要困难和现实得多。

CASIA NIR-VIS v2.0 [126]和 CUFSS [330]专注于跨模态人脸识别，如近红外 vs. RGB 人脸验证识别。此外，DFW [118]还致力于研究伪装的人脸识别，如化妆、留胡子、蓄小胡子和戴太阳镜等。此外，MS-Celebe-1M 低镜头数据集[73]为低镜头人脸识别提供了一个基准。

5.5 挑战和未来工作

在这一节中，我们从网络结构、训练监督、特定的人脸识别任务和数据集等多个角度回顾了深度人脸表示的最新进展。自深度学习盛行以来，深度人脸表示取得了显著的进展，并成功应用于许多现实场景中。剩下的主要挑战如下：

表 9 。 深度人脸识别常用的训练和测试公共数据集。

资料组	年	#主题	#图像视频	每个主题的 ImgVid 数量	描述
训练					
CASIA-WebFace [292]	2014	10,575	494,414-	47	第一个公开的大规模人脸数据集
VGGFace [175]	2015	2,622	2.6M-	1,000	每个主题包含大量图像
CelebA [302]	2015	10,177	202,599-	20	属性和身份的丰富注释
UMDFaces [10]	2015	8,277	367K-	45	面部姿态变化丰富
MS-Celebe-1M[74]	2016	100k	10M-	100	最大的名人脸公共数据集
MegaFace [110, 170]	2016	672,057	4.7M-	七	非名人的长尾数据集
VGGFace2 [23]	2017	9,131	3.31 米-	363	高质量的数据集，变化范围很广
视频[9]	2017	3,107	-22,075	七	从 YouTube 收集的视频训练数据集
MS-Celebe-1M 低镜头[73]	2017	20k, 1k	1M, 1k-	58,1	低镜头人脸识别

IMDb-脸[236]	2018	57k	1.7M-	29	大规模噪声控制数据集
QMUL-surface[236]	2018	5,319	220,890-	41	低分辨率监控数据集
测试					
LFW [89]	2007	5,749	13,233-	2.3	无约束条件下的经典基准
YTF [264]	2011	1,595	-3,425	2.1	无约束视频中的人脸识别
CUF5F [330]	2011	1,194	2,388-	2	照片草图人脸识别
CASIA NIR-VIS v2.0 [126]	2013	725	17,580-	24.2	近红外与 RGB 人脸识别
IJB-阿[113]	2015	500	5,7122,085	11.44.2	基于集合的大变化人脸识别
CFP [194]	2016	500	7,000-	14	正面到侧面交叉姿态人脸验证
MS-Celebe-1M 低镜头[73]	2016	20k, 1k	100k, 20k-	5,20	低镜头人脸识别
MegaFace [110, 170]	2016	690,572	1M-	1.4	百万脸的大型标杆
IJB-B [262]	2017	1,845	11,7547,011	6.373.8	基于集合的全姿态变化人脸识别
CALFW [347]	2017	4,025	12,174-	3	跨年龄人脸验证
AgeDB [167]	2017	570	16,516-	29	跨年龄人脸验证
SLLFW [45]	2017	5,749	13,233-	2.3	提高 LFW 否定对的难度
《保护所有移徙工人及其家庭成员权利国际公约》[346]	2017	3,968	11,652-	2.9	交叉人脸验证
万亿对[39]	2018	1M	1.58 米-	1.6	带有大量干扰物的大型基准
IJB-C [161]	2018	3,531	31,33411,779	63	基于集合的大变化人脸识别
IJB-南[105]	2018	202	5,656552	2812	真实世界的监控视频
RFW [248]	2018	11,429	40,607-	3.6	为了减少人脸识别中的种族偏见
DFW-[118]	2018	600	7,771-	13	伪装人脸识别
QMUL-调查[236]	2018	10,254	242,617-	23.7	低分辨率监控视频

- 在有限的条件下:虽然现有方法在各种基准上实现了高精度，但是当开发和应用(即。*训练和推理*)*受限于计算成本和训练数据量*。t.
- 监控视频人脸识别:在许多现实世界的应用中，监控人脸识别是一个常见的场景，其中的挑战包括各种面部变化，如大姿态，运动模糊，低照度，低分辨率，遮挡等..
- 标签噪声:大规模人脸数据集中的标签噪声频繁出现，对训练造成危害。抗噪声方法还有很大的发展空间。
- 不平衡数据:训练数据的不平衡分布也给人脸表征学习带来了问题，如人脸身份或领域的长尾分布等。

为了应对这些挑战，未来需要探索一些有价值的研究方向。我们在下面介绍它们。

- 轻量级人脸识别:巨大的内存和计算成本通常使得在移动或嵌入式设备上使用重型网络变得不切实际。虽然很多著作[29, 55, 159, 267, 268, 272]对轻量级人脸识别进行了研究，但仍然有必要对轻量级模型进行高效率、高精度的改进。
- 对视频变化的鲁棒性:它总是需要针对变化条件的鲁棒的人脸表示模型，特别是对于监控视频中的人脸识别任务。对低图像质量和大面部姿态的鲁棒性是许多实际应用中的核心需求。
- 噪声标签学习:标签噪声是采集大规模人脸数据集时不可避免的问题。某些著作[39, 42, 236, 335]研究了如何去除噪声数据来建立一个干净的数据集，而其他一些[87, 252, 351]旨在学习噪声鲁棒的人脸表示。但是它们中的大多数容易受到初始模型的能力的影响，并且需要在现实世界场景中更加灵活。人脸识别中的噪声标签学习仍然是一个未解决的问题。
- 跨域人脸识别:人脸数据中有许多不同的域因素，如人脸年龄、姿态、种族、成像方式等。一些著作[22, 53, 227, 247, 248, 260, 321]研究了其中一小部分人的人脸识别。如何获得跨域人脸识别的通用表示是一个具有挑战性的研究课题。

- 不平衡数据下的学习:长尾数据上的表示学习是许多人脸数据集中存在的问题。由于类内变异的代表性不足，样本有限的被试在训练过程中往往被忽略。不平衡数据尺度引起的领域偏差是人脸识别中的另一个常见问题。这些问题值得统一框架处理。
- 无标签人脸学习:实际应用中有大量的无标签人脸数据。然而，当数据集不断增长时，手动注释它们的成本过高。近年来，半监督学习和人脸聚类方法越来越受到人们的关注。如何有效地利用未标记数据来提高人脸识别是一个很有前途的方向。

6 讨论和结论

6.1 讨论

近年来，深度人脸识别取得了很大的进展，但对于每个元素来说，仍然存在许多挑战性的问题。读者可以参考每个正文部分的结尾部分，我们在那里提供了对问题的详细分析。在这里，我们将深入讨论挑战和未来的工作。表 10 的上半部分阐述了人脸检测、预处理和表示之间的共同问题。我们可以发现，问题主要包括三个方面，即，。面部和图像变化，数据和标签分布，计算效率。例如，在第一方面，面部变化包括大的面部姿态、极端表情、遮挡和面部比例，而图像变化包括在视频面部识别中频繁出现的诸如运动模糊、低照度和分辨率等客观因素。另一个例子表明了训练效率的需要，包括更快的训练和快速的收敛，这两者都致力于将大型人脸表示网络(通常为数百层)的学习从数周加速到数小时；前者一般侧重于混合精度训练或大规模训练的分布式框架(超过百万个身份)，而后者侧重于改进监督、初始化、更新方式、激活、架构等.. 在这里，我们不再重复每一个细节，而是将表 10 留给读者来探索常见的挑战和进一步的改进。值得一提的是，所有元素都将从针对这些问题的解决方案中受益，因为它们是元素之间的共同问题。

尽管个别要素最近取得了进展，但仍然有必要从整体框架的角度讨论和探索未来的发展趋势，因为每个要素对整个系统都有重大影响。任何一个元素中的劣势都会成为木桶中最短的一块，损害最终的性能。例如，人脸检测是端到端人脸识别的第一步，人脸包围盒的准确性直接影响后续的预处理；不准确的人脸定位会带来不对齐的信息和来自非人脸区域的干扰，导致不适当的特征计算，从而影响性能。另一个例子是，一组不准确的面部标志会损害对齐，然后还会妨碍后续的特征计算，即使

表 10。端到端深度人脸识别的主要挑战概述。

Challenges		Description
The common issues across the elements.	Facial / image variations	<ul style="list-style-type: none">• Large pose, extreme expression, occlusion, facial scale.• Motion blur, low illumination, low resolution.
	Data / label distribution	<ul style="list-style-type: none">• Limited labeled data, label noise.• Usage of unlabeled data.• Imbalance over scale, identity, race, domain, modality.
	Computational efficiency	<ul style="list-style-type: none">• Inference on non-GPU server and edge computing.• Fast training and convergence.
The issues concerning to the entire system.	Joint modeling and optimization	<ul style="list-style-type: none">• End-to-end training and inference.• Unified learning objective.• Mutual promotion.
	Interpretability	<ul style="list-style-type: none">• Explainable learning and inference.

前面检测到的边界框是公平的。对于作为端到端人脸识别系统核心操作的人脸表示的最后一步，追求其自身的性能至关重要

在对齐的面上进行给定的裁剪。

表 10 的下半部分从整个系统的角度指出了主要的挑战。该系列包括两个主要方面。第一个方面涉及深度人脸识别中的可解释性。虽然可解释的人工智能，即所谓的 XAI，已经被研究了很长时间，但可解释的深度人脸识别还处于起步阶段[263, 294, 305, 350]。我们认为有两种方法可以获得深度人脸识别的可解释性，即，。分别为自上而下和自下而上。自顶向下的方法依靠人类的先验知识进行算法探索，因为人类在许多粗糙条件下表现出比深度模型更优越的人脸识别能力。自下而上的方式是指从人脸数据本身的角度进行探索，如在空间和尺度维度上对可解释的深层人脸识别进行建模。

第二个方面是指人脸检测、预处理和表示的联合建模和优化。理想情况下，这三个元素应该在端到端精度方面联合建模和优化。一方面，这种整合为寻找整体系统的全局最优解提供了可能性；另一方面，系统的各个元素可以从上游元素中受益。然而，这些要素对于自己的任务有不同的学习目标。例如，人脸检测的目的是回归真实人脸的正确边界框，而人脸表示学习的目的是从给定的裁剪人脸中跨越一个有区别的特征空间。因此，如何统一这些学习目标对于联合优化是一个具有挑战性和关键性的问题。人们可以找到一组作品[[41, 77, 258, 265, 319, 340, 342, 349, 352]，试图集成人脸检测和对齐，或者人脸对齐和表示，用于联合增强。但是人脸检测仍然很难与人脸表示相结合，因为它们的目标和实现机制有很大的不同。尽管如此，开发端到端可训练的深度人脸识别，并研究如何通过共同学习来进一步提高它们仍然是值得的。此外，除了本次调查的主题之外，还有一个悬而未决的问题，即我们如何开发一个单一的网络来执行端到端的人脸识别。

6.2 结论

在这篇综述中，我们系统地回顾了由人脸检测、人脸预处理和人脸表示组成的端到端深度人脸识别的最新进展。虽然有许多关于人脸识别的调查，但它们大多集中在人脸表示问题上，而没有考虑管道中其他元素的相互影响，而这次调查是第一次全面回顾端到端深度人脸识别的元素。我们从多个方面对每个元素中的许多方法进行了详细的讨论和比较。此外，我们分析了现有的挑战，并收集了一些有前途的未来研究方向。此外，我们还讨论了它们的相互影响以及整体框架的未来工作。我们希望这项调查能给人们带来有益的思考，以便更好地理解端到端人脸识别的全貌，并以系统的方式进行更深入的探索。

引用

- [1] W. Abd-Almageed, y. 吴, s. 罗尔斯, s. Harel, t. 哈斯纳我. Masi, j. 崔杰. T. Leksut, j. 金, p. Natarajan. Nevatia, 和 g. G. 梅迪纳. 2016.使用深度多姿态表示的人脸识别. 2016 年 IEEE 计算机视觉应用冬季会议(WACV) (2016), 1–9.
- [2] 爱达荷 (Idaho 的缩写). Adjabi, a. 瓦哈比, a. 苯扎维, 还有 a. 塔勒布-艾哈迈德. 2020.人脸识别的过去、现在和未来:综述. 电子学 9, 8 (2020).
- [3] T. Ahonen, a. Hadid, 和 m. Pietikinen. 2004.基于局部二值模式的人脸识别.《欧洲计算机视觉会议论文集》(ECCV). 斯普林格, 469–481.
- [4] G. Antipov, m. Baccouche 和 j. 杜杰雷. 2017.通过生成性年龄归一化促进跨年龄脸验证.在美国电气和电子工程师协会生物统计学国际联合会会议(IUCB)会议录中. 191–199.
- [5] G. Antipov, m. Baccouche 和 j. 杜杰雷. 2017.用条件生成对抗网络面对衰老.在美国电气和电子工程师协会国际图像处理会议(ICIP)记录中. 2089–2093.
- [6] H. R. 阿拉比尼亚. 2009.人脸识别技术综述. 信息处理系统杂志 5, 2 (2009), 41–68.
- [7] B. 杨, j. 颜, z. 雷, 还有 s. Z. 李. 2014.用于多视角人脸检测的集合通道特征.在电气和电子工程师协会国际生物识别联合会议上. 1–8.
- [8] Y. 白, y. 张, m. 丁, 还有 b. 加内姆. 2018.用生成性对抗网络在野外寻找小脸.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 21–30.
- [9] A. Bansal, c. D. 卡斯蒂略, r. Ranjan 和 r. Chellappa. 2017.doãžts 和 donãžts 的 CNN 人脸验证.在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 2545–2554.
- [10] A. Bansal, a. 南部里, c. D. 卡斯蒂略, r. Ranjan 和 r. Chellappa. 2017.用于训练深层网络的带注释的人脸数据集.在美国电气和电子工程师协会生物统计学国际联合会会议(IUCB)会议录中. 464–473.

- [11] 页 (page 的缩写)。名词 (noun 的缩写)。Belhumeur, d. W. 雅各布斯, d. J. 克里格曼和 n. 库马尔。2013.使用样本的一致性来定位面部的一部分。《IEEE 模式分析和机器学习学报》35, 12 (2013), 2930–2940。
- [12] 页 (page 的缩写)。名词 (noun 的缩写)。Belhumeur, p. H. Joo, 还有 d. J. 克里格曼。1997.特征脸 vs. 鱼眼空间:使用类特定线性投影的识别。《模式分析和机器智能的电气和电子工程师协会交易》19, 7 (1997), 711–720。
- [13] C. 薄伽瓦图拉, c. 朱, k. Luu, 还有 m. Savvides。2017.比实时面部对齐更快:一种不受约束的 3D 空间变换网络方法。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。4000–4009。
- [14] 动词 (verb 的缩写)。Bianz 和 t. 维特尔。2003.基于拟合三维形变模型的人脸识别。《模式分析和机器智能杂志》25, 9 (2003), 1063–1074。
- [15] K. W. 鲍耶, k. 常, 和 p. 弗林。2006.3D 和多模态 3D+2D 人脸识别方法和挑战综述。《计算机视觉和图像理解》101, 1 (2006), 1–15。
- [16] 南。C. Brubaker, j. 吴, j. 孙, m. D. 马林, 还有 j. 米 (meter 的缩写)。瑞格。2008.面向人脸检测的级联增强集成设计。《国际计算机视觉杂志》77, 1-3 (2008), 65–86。
- [17] A. Bulat 和 g. Tzimiropoulos。2016.大姿态人脸对齐的局部证据卷积融合。在英国机器视觉会议上。
- [18] A. Bulat 和 g. Tzimiropoulos。2017.我们离解决 2D & 3D 人脸对齐问题还有多远? (以及 230, 000 个 3D 面部标志的数据集)。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。1021–1030。
- [19] X. 页 (page 的缩写)。Burgos-Artizun, p. 佩罗娜和 p. 多拉。2013.遮挡下鲁棒的人脸标志点估计。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。1513–1520。
- [20] D. 曹, x. 朱, x. 黄, j. 郭, 和 z. 雷。2020.领域平衡:长尾领域的人脸识别。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。5671–5679。
- [21] J. 曹, y. 胡, h. 张, r. 他, 还有 z. 星期日。2019.走向高保真的野外正面化。《国际计算机视觉杂志》128 (2019), 1485–1504。
- [22] K. 曹, y. 荣, c. 李, x. 唐, 还有 c. C. 阿来。2018.基于深度残差等变映射的姿态鲁棒人脸识别。《美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)》(2018), 5187–5196。
- [23] q. 曹, l. 沈, w. 谢, o. 米 (meter 的缩写)。帕克希和 a. 齐泽曼。2018.VGGFace2:一个跨姿势和年龄识别人脸的数据集。在美国电气和电子工程师协会自动人脸和手势识别国际会议记录中。67–74。
- [24] F. 昌, a. T. Tran, t. 哈斯纳我。Masi. Nevatia, 和 g. 梅迪纳。2017.FacePoseNet:为无地标面对齐做准备。在美国电气和电子工程师协会国际计算机视觉研讨会会议录中。1599–1608。
- [25] J. 常, z. 兰, c. 程, 还有 y. 魏。2020.人脸识别中的数据不确定性学习。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。5710–5719。
- [26] D. 陈, g. 华, f. 文, 还有 j. 星期日。2016.高效人脸检测的监督变换网络。《欧洲计算机视觉会议记录》(ECCV), 第一卷。9909.122–138。
- [27] J. 陈, r. Ranjan, s. Sankaranarayanan, a. 库马尔, c. 陈, v. Patel, c. D. 卡斯蒂略, 和 r. Chellappa。2017.基于深度卷积神经网络的无约束静止/视频人脸验证。《国际计算机视觉杂志》126 (2017), 272–291。
- [28] 长度。陈, h. 苏, 又问。纪。2019.基于核密度深层神经网络的人脸对齐。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。6991–7001。
- [29] 南。陈, y. 刘, x. 高, 和 z. 韩。2018.Mobilefacenet:高效的 cnns, 用于在移动设备上进行的实时人脸验证。在中国生物识别会议上。斯普林格, 428–438。
- [30] Y. 程, j. 赵, z. 王, y. 徐, k. Jayashree, s. 沈, 还有 j. 冯。2017.一目了然:用于低级学习的紧凑向量表示。在美国电气和电子工程师协会计算机视觉研讨会国际会议记录中。1924–1932。
- [31] C. 池, s. 张, j. 兴, z. 雷。Z. 李, 和 x. 邹。2019.用于高性能人脸检测的选择性细化网络。《AAAI 人工智能会议录》, 第一卷。33.8231–8238。
- [32] J. 崔斯。帕克, k. 金, j. 公园, d. 金, 还有 h. 垫片。2017.使用生成对抗网络的低射学习的人脸生成。在美国电气和电子工程师协会国际计算机视觉研讨会会议录中。1940–1948。
- [33] F. Chollet。2017.例外:深度可分离卷积的深度学习。《IEEE 计算机视觉和模式识别会议记录(CVPR)》(2017), 1800–1807。
- [34] A. R. 乔杜里, t. 林, s. Maji, 和 e. G. 博学——米勒。2016.双线性 CNNs 的一对多人脸识别。2016 年 IEEE 计算机视觉应用冬季会议(WACV) (2016), 1–9。
- [35] F. 科尔, d. 贝朗格, d. 克里希南, a. 莎娜, 我。莫塞里, 还有 w. T. 弗里曼。2017.从人脸身份特征合成归一化人脸。《IEEE 计算机视觉和模式识别会议记录(CVPR)》(2017), 3386–3395。
- [36] T. F. Cootes 和 c. J. 泰勒。1992.活动形状 modelsääääsmart snakesääz. 英国机器视觉会议记录。斯普林格, 266–275。
- [37] T. F. 库特斯, k. 沃克和 c. J. 泰勒。2000.基于视图的活动外观模型。在第四届自动人脸和手势识别国际会议记录中。227–232。
- [38] A. Dapogny. Cord, 和 k. 拜伊。2019.DeCaFa:用于野外人脸对齐的深度卷积级联。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。6892–6900。
- [39] 深冈。2020.万亿对。<http://trillionpairs.deeplint.com/overview>. (2020). (2020 年 9 月 15 日访问)。
- [40] J. 邓, w. 董, r. Socher, l. 李, k. 李, 和 l. 飞飞。2009.Imagenet:大规模分层图像数据库。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。leee, 248–255。
- [41] 邓建康。郭嘉。伊万杰洛斯-韦弗拉斯。艾琳-科塔西亚和斯特凡诺斯-扎费里奥。2020.视网膜面部:野外单次多层次面部定位。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。5203–5212。
- [42] J. 邓, j. 郭, n. 薛, 还有 s. Zafeiriou。2019.弧面:深度人脸识别的附加角裕度损失。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。4690–4699。
- [43] J. 邓, 成, 薛, 杨, 周, 和 s. Zafeiriou。2018.面向姿态不变人脸识别的对抗式人脸紫外线地图完成。《IEEE 计算机视觉和模式识别会议记录(CVPR)》(2018), 7093–7102。
- [44] J. 邓, g. Trigeorgis, y. 周, 和 s. Zafeiriou。2019.野外多视角联合人脸对齐。《美国电气和电子工程师协会图像处理学报》28, 7 (2019), 3636–3648。
- [45] W. 邓, j. 胡, n. 张, b. 陈, 还有 j. 郭。2017.细粒度的人脸验证:FGLFW 数据库、基线和人类-DCMN 伙伴关系。《模式识别》66 (2017), 63–73。
- [46] Z. 邓, x. 彭, z. 李, 和 y. 乔。2019.异构人脸识别的互分量卷积神经网络。《美国电气和电子工程师协会图像处理学报》28 (2019), 3102–3114。
- [47] C. 丁和丁。陶。2015.基于多模态深度人脸表示的鲁棒人脸识别。《IEEE 多媒体交易》17 (2015), 2049–2058。
- [48] C. 丁和丁。陶。2016.姿态不变人脸识别综述。《ACM Trans. 智力. Syst. 科技》7 (2016), 37:1–37:42。
- [49] C. 丁和丁。陶。2018.用于基于视频的人脸识别的主干-分支集成卷积神经网络。《模式分析和机器智能》40 (2018), 1002–1014。
- [50] X. 董, y. 颜, w. 欧阳, 还有 y. 杨。2018.用于面部标志检测的风格聚合网络。在 IEEE 计算机视觉和模式识别会议记录中。379–388。
- [51] X. 董和 y. 杨。2019.教师指导学生如何从部分标记的图像中学习面部标志点检测。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。783–792。
- [52] X. 董, s. 余, x. 翁, s. 魏, y. 杨, 还有 y. 谢赫。2018.配准监督:一种提高面部标志检测器精度的无监督方法。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。360–368。
- [53] 长度。杜, h. 胡, 还有 y. 吴。2019.基于迁移学习和对抗学习的跨年龄人脸识别年龄因素去除网络。《IEEE 视频技术电路和系统交易》(2019), 1–1。
- [54] Y. 段, j. 鲁, 和 j. 周。2019.UniformFace:学习用于人脸识别的深度等分布表示。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。3410–3419。

- [55] C. 名词 (noun 的缩写)。Duong, G. Quach. 乐, n. 阮, 还有 k. 鲁。2019.Mobiface:一个轻量级的深度学习移动设备上的人脸识别。在第十届国际生物统计学理论、应用和系统会议(BTAS)上。IEEE, 1-6。
- [56] 米 (meter 的缩写)。Everingham, I. 动词 (verb 的缩写)。Gool, c. K. 爱达荷 (Idaho 的缩写)。Williams, j. Winn, 还有 a. 齐泽曼。2010.帕斯卡视觉对象类挑战。《国际计算机视觉杂志》88, 2 (2010), 第303-338 页。
- [57] 米 (meter 的缩写)。Everingham 和 j. 韦恩。2011.pascal 可视化对象类挑战 2012(voc2012)开发工具包。模式分析, 统计建模和计算学习, 技术。Rep 8 (2011)。
- [58] 南。南。Farfard, m. J. 萨伯里安, 还有 l. J. 李。2015.基于深度卷积神经网络的多视角人脸检测。第五届美国计算机学会多媒体检索国际会议论文集。643-650。
- [59] 页 (page 的缩写)。F. Felzenszwalb, r. B. Girshick, d. 麦卡勒斯特和 d. 拉玛南。2010.使用基于区分训练零件的模型进行目标检测。模式分析和机器智能 32, 9 (2010), 1627-1645。
- [60] Y. 冯, f. 吴, x. 邵, y. 王, 和 x. 周。2018.利用位置图回归网络联合三维人脸重建和密集对齐。《欧洲计算机视觉会议论文集》(ECCV)。534-551。
- [61] Z. 冯, j. 基特勒, m. Awais, p. 胡伯, 还有 x. 吴。2018.卷积神经网络鲁棒面部标志点定位的翼损失。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。2235-2245。
- [62] C. Galea 和 r. A. 法鲁吉亚。2017.使用基于深度学习的体系结构的法医面部照片草图识别。电气和电子工程师协会信号处理信函 24 (2017), 1586-1590。
- [63] H. K. Galoogahi 和 t. Sim。2012.多模态人脸草图识别。在美国电气和电子工程师协会国际多媒体会议暨博览会会议录。IEEE, 224-229。
- [64] 南。Ge, j. 李问。叶, 和 z. 罗。2017.用 lle-cnns 探测野外的假面。在 IEEE 计算机视觉和模式识别会议记录中。2682-2690。
- [65] G. 黄, z. 刘, 和 k. q. 温伯格。2017.密集连接的卷积网络。美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR) (2017), 2261-2269。
- [66] 南。风, y. 石, 和 a. K. 简。2019.低质量视频人脸识别:多模式聚合递归网络(MARN)。在美国电气和电子工程师协会国际计算机视觉研讨会会议录中。1027-1035。
- [67] 南。风, y. 史, n. D. 卡卡尔, 还有 a. K. 简。2019.视频人脸识别:组件式特征聚合网络。2019 生物识别国际会议(ICB) (2019), 1-8。
- [68] 爱达荷 (Idaho 的缩写)。J. 古德费勒, j. Pouget-Abadie, m. Mirza, b. 徐, d. 沃德-法利, s. Ozair, a. C. 库维尔和 y. 本吉奥。2014.生成性对抗网络。神经网络处理系统进展。2672-2680。
- [69] R. 恶心, 我。A. 马修斯, j. F. 科恩, t. 卡纳德, 和 s. 贝克。2008.Multi-PIE。在 IEEE 自动人脸和手势识别国际会议记录中。1-8。
- [70] 页 (page 的缩写)。Grother. J. 迈克尔斯和 p. J. 非利普斯。2003.人脸识别供应商测试 2002 性能指标。基于音频和视频的生物特征认证国际会议。斯普林格, 937-945。
- [71] 页 (page 的缩写)。格罗特和 m. 颜。2014.人脸识别供应商测试:人脸识别算法的性能。NIST 机构间报告 8009, 5 (2014), 14。
- [72] 南。郭, j. 徐, d. 陈, c. 张, x. 王, 还有 r. 赵。2020.面向人脸聚类的密度感知特征嵌入。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。6698-6706。
- [73] Y. 郭和 l. 张。2017.通过推广代表性不足的班级进行一次性人脸识别。arXiv 预印本 arXiv:1707.05574 (2017)。
- [74] Y. 郭, l. 张, y. 胡, x. 他, 还有 j. 高。2016.ms-celebe-1m:大规模人脸识别的数据集和基准。在欧洲计算机视觉会议记录中。斯普林格, 87-102。
- [75] C. 韩, s. 山, m. 侃, s. 吴, 和 x. 陈。2018.基于对比卷积的人脸识别。《欧洲计算机视觉会议论文集》(ECCV)。118-134。
- [76] Z. 郝, y. 刘, h. 秦, j. 颜, x. 李, 和 x. 胡。2017.尺度感知人脸检测。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。6186-6195。
- [77] 米 (meter 的缩写)。Hayat, s. H. 可汗, n. Werghi 和 r. 哥克。2017.用于无约束人脸识别的联合配准和表示学习。IEEE 计算机视觉和模式识别会议记录(CVPR) (2017), 1551-1560。
- [78] K. 他, x. 张, s. 任, 和 j. 星期日。2016.用于图像识别的深度残差学习。2016 年 IEEE 计算机视觉和模式识别会议(CVPR) (2016), 770-778。
- [79] R. 何, j. 曹, l. 宋, z. 孙, 和 t. 谭。2020.用于近红外-可见光人脸识别的对立交叉光谱人脸完成。模式分析和机器智能 42 (2020), 1025-1037。
- [80] R. 他, x. 吴, z. 孙, 和 t. 谭。2017.用于近红外-可见光人脸识别的学习不变深度表示。AAAI 会议记录人工智能。
- [81] R. 他, x. 吴, z. 孙, 和 t. 谭。2019.沃瑟斯坦 CNN:学习近红外-可见光人脸识别的不变特征。美国电气和电子工程师协会模式分析和机器智能学报 41 (2019), 1761-1773。
- [82] 南。Honari, p. Molchanov, s. 泰瑞, p. 文森特, c. J. Pal, 还有 j. 考茨。2018.用半监督学习改进地标定位。IEEE 计算机视觉和模式识别会议记录(CVPR) (2018), 1546-1555。
- [83] 南。洪, w. lm, j. B. Ryu, 还有 h. 杨。2017.SSP-丹:用于人脸识别的深度域自适应网络, 每人一个样本。在美国电气和电子工程师协会国际图像处理会议(ICIP)记录中。825-829。
- [84] A. G. 霍华德, m. 朱, b. 陈, d. Kalenichenko. 王, t. Weyand, m. 安德烈托和 h. 亚当。2017.移动网络:用于移动视觉应用的高效卷积神经网络。ArXiv abs/1704.04861 (2017)。
- [85] J. 胡, l. 沈, 还有 g. 星期日。2018.挤压激励网络。IEEE 计算机视觉和模式识别会议记录(CVPR) (2018), 7132-7141。
- [86] 页 (page 的缩写)。胡与维。拉玛南。2017.寻找小脸。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。1522-1530。
- [87] W. 胡, y. 黄, f. 张和 r. 李。2019.训练人脸识别中神经系统的噪声容忍范式。IEEE 计算机视觉和模式识别会议记录(CVPR) (2019), 11879-11888。
- [88] Y. 胡, x. 吴, b. 余, r. 他, 还有 z. 星期日。2018.姿势引导的照片真实感人脸旋转。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。8398-8406。
- [89] G. B. 黄, m. Ramesh, t. 伯格, 和 e. 博学——米勒。2007.野外标记人脸:一个研究无约束环境下人脸识别的数据库。技术报告 07-49.马萨诸塞州大学阿默斯特分校。
- [90] 长度。黄, y. 杨, y. 邓, 和 y. 于。2015.Densebox:将地标定位与端到端对象检测统一起来。arXiv 预印本 arXiv:1509.04874 (2015)。
- [91] R. 黄, s. 张, t. 李和 r. 他。2017.超越人脸旋转:全局和局部感知 GAN 用于真实感和身份保持正面视图合成。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。2458-2467。
- [92] X. 黄, w. 邓, h. 沈, x. 张和 j. 叶。2018.传播网:将点传播到曲线以学习结构信息。在 IEEE/CVF 计算机视觉和模式识别会议记录中。7265-7274。
- [93] Y. 黄, y. 王, y. 泰, x. 刘, p. 沈, s. 李, j. 李, 和 f. 黄。2020.课程空间:深度人脸识别的适应性课程学习损失。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。5901-5910.[94]王洪志。李, x. 纪, 和 y. 王。2017.美国有线电视新闻网。arXiv 预印本 arXiv:1706.01061 (2017)。
- [95] F. 名词 (noun 的缩写)。Iandola. W. k. 莫斯科维奇。Ashraf, s. 韩, w. J. 磨磨蹭蹭, 还有 k. 库特泽。2017.挤压网:AlexNet 级精度, 参数减少 50 倍, 模型尺寸 < 1MB。ArXiv abs/1602.07360 (2017)。
- [96] J. 郑海龙, b. Yoo, c. 彩, d. 朴, 还有 j. 金。2015.用多任务深度神经网络旋转你的脸。在 2015 年 IEEE 计算机视觉和模式识别会议(CVPR)上。676-684。

- [97] 米 (meter 的缩写)). Jaderberg, k. 西蒙扬, a. 齐泽曼和 k. Kavukcuoglu. 2015. 空间变压器网络. 神经信息处理系统进展. 2017–2025.
- [98] H. 江和 e. 博学—米勒. 2017. 用更快的 r-cnn 进行人脸检测. 在 2017 年美国电气和电子工程师协会自动人脸和手势识别国际会议记录中. 650–657.
- [99] H. 金, s. 张, x. 朱, y. 唐, z. 雷, 还有 s. Z. 李. 2019. 用知识蒸馏学习轻量级人脸检测器. 2019 年生物识别国际会议(ICB). 1–7.
- [100] X. 金和 x. 谭. 2017. 野外人脸对齐: 一项调查. *计算机视觉和图像理解* 162 (2017), 1–22.
- [101] Y. 金, j. 鲁, 又问. 阮. 2015. 跨模态人脸识别的大边缘耦合特征学习. 《国际生物统计学会会议论文集》(ICB). 286–292.
- [102] Y. 靖, 问. 刘, 和 k. 张. 2017. 用于鲁棒面部标志定位的堆叠沙漏网络. 在 IEEE 计算机视觉模式识别研讨会会议上.
- [103] A. Jourabloo 和 x. 刘. 2016. 基于 CNN 的密集 3D 模型拟合的大姿态人脸对齐. 在 2016 年 IEEE 计算机视觉和模式识别会议(CVPR)上. 4188–4196.
- [104] A. Jourabloo, m. 叶, x. 刘, 还有 l. 任. 2017. 用单个 CNN 进行姿态不变的人脸对准. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 3219–3228.
- [105] 名词 (noun 的缩写). D. 卡尔卡, b. 梅兹 j. A. 邓肯, k. OrConnor, s. 埃利奥特, k. Hebert, j. 布莱恩和 a. K. 简. 2018. IJB: IARPA 骏利监控视频基准. *第九届生物识别理论、应用和系统国际会议(BTAS) (2018)*, 1–9.
- [106] 米 (meter 的缩写)). 侃, s. 山, h. 昌, 和 x. 陈. 2014. 用于跨姿态人脸识别的堆叠式渐进自动编码器. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 1883–1890.
- [107] 米 (meter 的缩写)). 侃, s. 山, 和 x. 陈. 2016. 用于跨视图分类的多视图深度网络. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 4847–4855.
- [108] B. 康, y. 金, b. 君, 而 d. 金. 2019. 用于精确人脸识别的注意特征对关系网络. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 5471–5480.
- [109] B. 康, y. 金, 还有 d. 金. 2018. 用于人脸识别的成对关系网络. 《欧洲计算机视觉会议论文集》(ECCV). 628–645.
- [110] 爱达荷 (Idaho 的缩写). Kemelmacher-Shlizerman, s. 米 (meter 的缩写)). Seitz, d. 米勒, 还有 e. 布罗萨德. 2016. megaface 基准: 100 万张人脸用于大规模识别. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 4873–4882.
- [111] 爱达荷 (Idaho 的缩写). Kemelmacher-Shlizerman, s. Suwajanakorn 和 s. 赛茨. 2014. 照明感知年龄进展. *2014 年 IEEE 计算机视觉和模式识别会议(2014)*, 3334–3341.
- [112] Y. 金, w. 朴植惠, 卢武铉和 j. Shin. 2020. 组脸: 学习潜在的组, 并为人脸识别构建基于组的表示. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5621–5630.
- [113] B. Klare, b. 克莱因, e. 塔博尔斯基, a. j. 布兰顿. 切尼, k. E. 艾伦, p. 格罗特, a. Mah, m. 伯格, 还有 a. K. 简. 2015. 推进无约束人脸检测与识别前沿: IARPA Janus Benchmark A. *2015 年 IEEE 计算机视觉和模式识别会议(CVPR) (2015)*, 1931–1939.
- [114] 米 (meter 的缩写)). Kostinger, p. Wohlfahrt, p. 米 (meter 的缩写)). 罗斯, 还有 h. 比朔夫. 2011. 野外带注释的面部标志: 用于面部标志定位的大规模真实世界数据库. 在美国电气和电子工程师协会计算机视觉研讨会国际会议记录(ICCV 研讨会)中. 2144–2151.
- [115] A. 克里哲夫斯基, 我. Sutskever 和 g. E. 韩丁. 2012. 用深度卷积神经网络进行图像网络分类. 神经信息处理系统进展. 1097–1105.
- [116] A. 库马尔和 r. Chellappa. 2018. 在树枝状 CNN 中解开 3D 姿态, 用于无约束 2D 人脸对齐. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 430–439.
- [117] A. n. Kumar, T. K. Marks, w. 牟, y. 王, m. 琼斯, a. Cherian, t. Koike-Akino, x. 刘, 还有 c. 冯. 2020. LUVU 面对齐: 估计地标的位置、不确定性和可见性可能性. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 8236–8246.
- [118] 动词 (verb 的缩写). Kushwaha, m. 辛格, r. 辛格, m. Vatsa, n. K. Ratha 和 r. Chellappa. 2018. 《荒野中的伪装面孔》. *美国电气和电子工程师协会计算机视觉和模式识别研讨会会议记录(2018)*, 1–18.
- [119] 南. 赖, z. 柴, s. 李. h. 孟, m. 阳, 和 x. 魏. 2019. 用于鲁棒面部标志检测的增强的归一化平均误差损失. 英国机器视觉会议记录. 111.
- [120] H. 法学和 j. 郑. 2018. 角网: 检测物体成对的关键点. 《欧洲计算机视觉会议论文集》(ECCV). 734–750.
- [121] 动词 (verb 的缩写). 乐, j. 勃兰特, z. 长度. 林, l. D. 布尔德夫和 t. 南. 黄. 2012. 交互式面部特征定位. 《欧洲计算机视觉会议论文集》(ECCV). 斯普林格, 679–692.
- [122] J. 萨马, 问. 秋, 和克. 萨皮罗. 2017. 不怕黑: 通过跨光谱幻觉和低阶嵌入的近红外-可见光人脸识别. *IEEE 计算机视觉和模式识别会议记录(CVPR) (2017)*, 6807–6816.
- [123] H. 李, z. 林, x. 沈, j. 勃兰特, 还有 g. 华. 2015. 用于人脸检测的卷积神经网络级联. 在 2015 年 IEEE 计算机视觉和模式识别会议(CVPR)上. 5325–5334.
- [124] J. 李, y. 王, c. 王, y. 泰, j. 钱, j. 杨, c. 王, j. 李, 和 f. 黄. 2019. DSFD: 双镜头人脸检测器. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5060–5069.
- [125] 南. 李. h. 李, j. 崔, 还有 h. 咋. 2019. 基于 CNN 和 3DMM 的姿态感知人脸对齐. 在英国机器视觉会议上. 106.
- [126] 南. Z. 李, d. 哦, z. 雷, 还有 s. 廖. 2013. CASIA 近红外-可见光 2.0 人脸数据库. *2013 年 IEEE 计算机视觉和模式识别研讨会会议(2013)*, 348–353.
- [127] 南. Z. 李. l. 朱, z. 张, a. 布雷克, h. 张和 h. 肥肥. 2002. 多视角人脸检测的统计学习. 在欧洲计算机视觉会议记录中. 斯普林格, 67–81.
- [128] Y. 李, b. 孙, t. 吴, 还有 y. 王. 2016. 一个 convNet 和一个 3d 模型端到端集成的人脸检测. 《欧洲计算机视觉会议记录》(ECCV), 第一卷. 9907.420–436.
- [129] Z. 李, d. 巩, 问. 李, d. 道, 而 x. 李. 2016. 异构人脸识别的互分量分析. *ACM Trans. 智力. Syst. 科技.* 7 (2016), 28:1–28:23.
- [130] Z. 李, x. 唐, j. 韩, j. 刘, 还有 r. 他. 2019. PyramidBox++: 寻找微小人脸的高性能检测器. *arXiv 预印本 arXiv:1904.00386 (2019)*.
- [131] 南. 廖, z. 雷, d. 哦, 还有 s. Z. 李. 2014. 大规模无约束人脸识别的基准研究. *IEEE 生物识别国际联席会议(2014)*, 1–8.
- [132] T. 林, p. r. dollaär. Girshick. 他, b. Hariharan 和 s. 贝隆吉. 2017. 用于目标检测的特征金字塔网络. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 2117–2125.
- [133] T. 林, p. Goyal. Girshick. 他, 还有 p. 多拉. 2017. 密集目标探测的焦点损失. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 2980–2988.
- [134] T. 林, a. RoyChowdhury 和 s. 马吉. 2015. 用于细粒度视觉识别的双线性 CNN 模型. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 1449–1457.
- [135] W. 林, j. 陈, c. D. 卡斯蒂略, 和 r. Chellappa. 2018. 无约束人脸的深度密度聚类. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 8128–8137.
- [136] B. 刘, w. 邓, y. 钟, m. 王, j. 胡, x. 陶, 和 y. 黄. 2019. 公平损失: 深度人脸识别的边缘感知强化学习. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 10051–10060.
- [137] C. 刘和 h. 韦克斯勒. 2002. 用于人脸识别的基于 Gabor 特征的增强型 fisher 线性判别模型分类. *图像处理的电气和电子工程师协会交易: 电气和电子工程师协会信号处理学会* 11 4 (2002) 的出版物, 467–76.
- [138] H. 刘, j. 鲁, j. 冯, 还有 j. 周. 2018. 基于视频的人脸对齐的双流变换网络. *《美国电气和电子工程师协会模式分析和机器智能学报》* 40 (2018), 2546–2554.

- [139] H. 刘, j. 卢, m. 郭, s. 吴, 还有 j. 周. 2020.用于鲁棒人脸对齐的学习推理决策网络. *美国电气和电子工程师协会模式分析和机器学习学报* 42 (2020), 679–693.
- [140] H. 刘, x. 朱, z. 雷, 还有 s. Z. 李. 2019.自适应人脸:用于人脸识别的自适应边缘和采样.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 11947–11956.
- [141] J. 刘, y. 邓, t. 白, z. 魏, 和 c. 黄. 2015.目标终级准确性:通过深度嵌入的人脸识别. (2015). arXiv:cs.CV/1506.07310
- [142] W. 刘, d. Angelov, d. 尔汉、克里斯蒂安·塞格迪、斯科特·里德、傅成阳和亚历山大·c. 伯格. 2016.固态硬盘:单次多盒检测器.《欧洲计算机视觉会议论文集》(ECCV). 21ãÅ37.
- [143] W. 刘, y. 文, z. 余, m. 李, b. 拉吉和我. 宋. 2017.球面:用于人脸识别的深度超球面嵌入.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 212–220.
- [144] W. 刘, y. 文, z. 余, 和 m. 杨. 2016.卷积神经网络的大余量软最大损失.. 在 *ICML*, 卷. 2.7.
- [145] X. 刘, b. 动词 (verb 的缩写). K. 动词 (verb 的缩写). 库马尔, c. 杨, 问. 唐, 还有 j. 你. 2018.基于图像集的无约束人脸识别的依赖感知注意力控制.《欧洲计算机视觉会议论文集》(ECCV). 548–565.
- [146] X. 刘, l. 宋, x. 吴, 和 t. 谭. 2016.近红外-可见光异构人脸识别的深度表示转换. 2016 *生物识别国际会议(ICB)(2016)*, 1–8.
- [147] Y. 刘, a. Jourabloo, w. 任, 与 x. 刘. 2017.密集面对齐.在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 1619–1628.
- [148] Y. 刘, h. 李, 和 x. 王. 2017.对大规模识别的特征识别和聚合的再思考. *ArXiv abs/1710.00870* (2017).
- [149] Y. 刘, h. 李, j. 颜, f. 魏, x. 王, 和 x. 唐. 2017.CNN 中目标检测的递归尺度近似.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 571–579.
- [150] Y. 刘, h. 沈, y. Si, x. 王, x. 朱, h. 石, z. 洪 h. 郭, z. 郭, y. 陈, b. 李, t. Xi. 于, h. 谢, g. 谢, m. 李问. 鲁, z. 王, s. 赖, z. 柴, 和 x. 魏. 2019.106 点面部地标定位大挑战.在美国电气和电子工程师协会国际多媒体会议暨博览会研讨会会议录中. 613–616.
- [151] Y. 刘, x. 唐, j. 韩, j. 刘, d. 芮, 还有 x. 吴. 2020.HAMBox:深入挖掘人脸检测的优质锚点.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 13043–13051.
- [152] Z. 刘, h. H. CloudMinds, j. 白, s. 李, 和 s. 长度. CloudMinds. 2019.视频人脸识别的特征聚合网络.在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 990–998.
- [153] Z. 刘, x. 朱, g. 胡, h. 郭, m. 唐, z. 雷, n. 米 (meter 的缩写)). 罗伯逊和 j. 王. 2019.语义对齐:为面部标志点检测找到语义一致的基本事实.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 3462–3471.
- [154] J. 吕, x. 邵, j. 兴, c. 程, 还有 x. 周. 2017.用于高性能面部标志检测的具有两阶段重新初始化的深度回归体系结构.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 3691–3700.
- [155] 米 (meter 的缩写)). t. Pham 和 T.J. Cham. 2007.在基于增强的人脸检测中使用统计快速训练和选择哈尔特征. 在 *ICCV* 召开的 *IEEE 国际计算机视觉会议* 会议记录中. 1–7.
- [156] 名词 (noun 的缩写). 马, x. 张, h. 郑, 还有 j. 星期日. 2018.Shufflenet v2:高效 cnn 架构设计的实用指南.《欧洲计算机视觉会议论文集》(ECCV). 116–131.
- [157] R. 曼玛塔, C.n. 吴, a. 斯摩拉, 还有 p. krhenbüHL. 2017.深度嵌入学习中的采样问题.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 2859–2867.
- [158] B. 马丁内斯, m. F. 威士达, x. 比奈法, 和 m. 潘提克. 2013.用于基于回归的面部点检测的局部证据聚集. *美国电气和电子工程师协会模式分析和机器学习学报* 35 (2013), 1149–1163.
- [159] Y. 马丁内斯-迪亚斯, l. 南. 卢瓦诺, h. 米 (meter 的缩写)). 瓦兹奎. 尼古拉斯-迪亚斯, l. 常, 和 m. 冈萨雷斯-门多萨. 2019.ShuffleFaceNet: 一个轻量级的人脸架构, 用于高效和高精度的人脸识别.在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 2721–2728.
- [160] 爱达荷 (Idaho 的缩写). Masl, s. 罗尔斯, g. G. Medioni, 还有 p. 纳塔拉詹. 2016.野外姿势感知人脸识别. *IEEE 计算机视觉和模式识别会议记录(CVPR)(2016)*, 4838–4846.
- [161] B. 梅兹 j. C. 亚当斯, 法学硕士. 邓肯, n. D. 卡尔卡, t. 米勒, c. 奥托, a. K. Jain. T. 黑鬼, j. j. 安德森. 切尼和帕特里克·格罗特. 2018.IARPA Janus Benchmark - C:人脸数据集和协议. 2018 *国际生物识别大会(ICB)(2018)*, 158–165.
- [162] D. Merget, m. 摇滚, 还有 g. 里戈尔. 2018.通过全卷积局部-全局上下文网络的鲁棒面部标志点检测.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 781–790.
- [163] X. 苗, x. 甄, x. 刘, c. 邓, v. 阿提索斯和 h. 黄. 2018.用于端到端对齐的直接形状回归网络.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5040–5049.
- [164] X. 明, f. 魏, t. 张, d. 陈, 和 f. 文. 2019.用于尺度不变人脸检测的组采样.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 3446–3456.
- [165] T. 塔米. Kaneko, 还有 o. 堀. [d. j. 用于人脸检测的联合类哈尔特征.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中.
- [166] 页 (page 的缩写). 米塔尔, m. Vatsa, 和 r. 辛格. 2015.基于深度网络的合成素描识别--一种迁移学习方法. 2015 *国际生物识别大会(ICB)(2015)*, 251–256.
- [167] 南. Moschoglou, a. 帕帕约安努特. Sagonas, j. 邓我. Kotsia, 和 s. Zafeiriou. 2017.年龄数据库:第一个人工收集的野生年龄数据库. *美国电气和电子工程师协会计算机视觉和模式识别研讨会会议录(2017)*, 1997–2005.
- [168] 米 (meter 的缩写)). Najibi, p. Samangouei, r. Chellappa, 和 l. 南. 戴维斯. 2017.SSH:单级无头人脸检测器.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 4885–4894.
- [169] 米 (meter 的缩写)). 纳吉比, b. 辛格, 还有 l. 南. 戴维斯. 2019.人脸检测的浮动区域方案.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 7723–7732.
- [170] A. Nech 和我. Kemelmacher-Shlizerman. 2017.百万级人脸识别的公平竞争环境. *美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)(2017)*, 3406–3415.
- [171] A. Newell. 杨, 还有 j. 邓. 2016.用于人体姿态估计的堆叠沙漏网络.《欧洲计算机视觉会议论文集》(ECCV). 483–499.
- [172] T. Ojala, m. Pietikainen, 和 t. 马恩帕. 2002.具有局部二值模式的多分辨率灰度和旋转不变纹理分类. *模式分析和机器学习* 24, 7 (2002), 971–987.
- [173] 迈克尔·奥皮茨、乔治·瓦尔纳特、乔治·奥普里尔、霍斯特·波辛格和霍斯特·比朔夫. [n. d.].
- [174] 米 (meter 的缩写)). 帕尔恰米, s. Bashbaghi, e. 格兰杰, 还有 s. 赛义德. 2017.使用深度自动编码器学习鲁棒的域不变表示用于静止视频人脸识别.在美国电气和电子工程师协会高级视频和信号监控国际会议记录(AVSS)中. 1–6.
- [175] O. 米 (meter 的缩写)). 帕克希, a. 维达尔迪, 还有 a. 齐泽曼. 2015.深度人脸识别.英国机器视觉会议记录.
- [176] X. 彭, r. 南. Feris, x. 王, 和丁. 名词 (noun 的缩写). 梅塔克萨斯. 2016.一种用于顺序人脸对齐的递归编解码网络.《欧洲计算机视觉会议论文集》(ECCV). 斯普林格, 38–56 岁.
- [177] Y. 钱, w. 邓和 j. 胡. 2019.具有极端姿势和野外表情的无监督人脸归一化. *IEEE 计算机视觉和模式识别会议记录(CVPR)(2019)*, 9843–9850.
- [178] H. 秦, j. 颜, l. 修, 和 x. 胡. 2016.级联 CNN 人脸检测的联合训练.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 3456–3465.

- [179] R. Ranjan, D. 卡斯蒂略, 和 r. Chellappa. 2017.用于鉴别人脸验证的 L2 约束软最大损失. *arXiv 预印本 arXiv:1703.09507* (2017).
- [180] R. Ranjan, v. 米 (meter 的缩写)). 帕特尔和 r. Chellappa. 2019.超空间:用于人脸检测、地标定位、姿态估计和性别识别的深度多任务学习框架. *美国电气和电子工程师协会模式分析和机器智能学报* 41, 1 (2019), 121–135.
- [181] R. Ranjan, s. Sankaranarayanan, a. Bansal, Bodla, j. 陈, v. 米 (meter 的缩写)). Patel, c. D. 卡斯蒂略, 和 r. Chellappa. 2018.理解面部的深度学习:机器可能和人类一样好, 甚至更好. *电气和电子工程师协会信号处理杂志* 35, 1 (2018), 66–83.
- [182] Y. 饶, j. 林, j. 鲁, 和 j. 周. 2017.用于基于视频的人脸识别的学习判别聚合网络. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 3801–3810.
- [183] Y. 饶, j. 鲁, 和 j. 周. 2017.面向视频人脸识别的注意感知深度强化学习. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 3951–3960.
- [184] C. 北里勒. 米 (meter 的缩写)). Nasrabadi, h. 权, 还有 r. Chellappa. 2016.从树中看到森林:近红外异质人脸识别的整体方法. *2016 年 IEEE 计算视觉和模式识别研讨会会议 (CVPRW)* (2016), 320–328.
- [185] 南. 任, k. 他, r. 吉尔希克和 j. 星期日. 2015.更快的 r-cnn:通过区域建议网络实现实时目标检测. *神经信息处理系统进展*, 91–99.
- [186] J. 页 (page 的缩写). y. 罗宾逊. 李, n. 张, y. 傅和图利亚科夫. 2019.拉普拉斯地标定位. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 10102–10111.
- [187] A. RoyChowdhury, x. 于, k. Sohn, e. 博学——米勒和 m. 钱德瑞克. 2020.通过在野外聚类未标记的人脸来改进人脸识别.《欧洲计算机视觉会议论文集》(ECCV).
- [188] C. Sagonas, e. 安东尼科斯, g. Tzimiropoulos, s. Zafeiriou 和 m. 潘提克. 2016.《300 张狂野的脸:数据库和结果》. *图像 Vis. 电脑*. 47 (2016), 3–18.
- [189] C. Sagonas, g. Tzimiropoulos, s. Zafeiriou 和 m. 潘提克. 2013.野外 300 张脸挑战:第一次面部标志定位挑战. 在美国电气和电子工程师协会计算机视觉研讨会国际会议记录中. 397–403.
- [190] 米 (meter 的缩写)). 桑德勒, a. G. 霍华德, 朱, 硕士. Zhmoginov, 和 l. 陈. 2018.移动网络 2:反向残差和线性瓶颈. *IEEE 计算机视觉和模式识别会议记录 (CVPR)* (2018), 4510–4520.
- [191] 南. Saxena 和 j. 维尔比克. 2016.基于神经网络的异构人脸识别.《欧洲计算机视觉会议 (ECCV) 研讨会论文集》.
- [192] F. 施罗夫, d. Kalenichenko 和 j. 菲尔宾. 2015.FaceNet:用于人脸识别和聚类的统一嵌入. 在 2015 年 IEEE 计算机视觉和模式识别会议 (CVPR) 上. 815–823.
- [193] 米 (meter 的缩写)). 舒尔茨和 t. 约阿希姆. 2004.从相对比较中学习距离度量. *神经信息处理系统进展*. 41–48.
- [194] 南. 森古普塔, j. 陈, c. 卡斯蒂略, v. 米 (meter 的缩写)). Patel, Chellappa, 和 d. W. 雅各布斯. 2016.野外正面对侧面验证. 2016 年 IEEE 计算机视觉应用冬季会议 (WACV). IEEE, 1–9.
- [195] J. 沈, s. Zafeiriou, g. G. Chrysos, j. Kossai, g. Tzimiropoulos, 和 m. 潘提克. 2015.第一次面部标志追踪野外挑战:基准和结果. 在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 1003–1011.
- [196] H. 石, x. 王, d. 映, z. 雷, x. 朱, 和 s. Z. 李. 2017.基于异构联合贝叶斯的跨模态人脸识别. *电气和电子工程师协会信号处理通讯* 24 (2017), 81–85.
- [197] X. 史, s. 山, m. 侃, s. 吴, 和 x. 陈. 2018.使用渐进校准网络的实时旋转不变人脸检测. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 2295–2303.
- [198] Y. 石和 a. K. 简. 2020.用未标记数据概括人脸表示. *ArXiv abs/2003.07936* (2020).
- [199] Y. 石, a. K. 普那教, 还有 n. D. 卡尔卡. 2019.概率人脸嵌入. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 6901–6910.
- [200] Y. 史, c. 奥托, 还有 a. K. 简. 2018.人脸聚类:表示和成对约束. *美国电气和电子工程师协会信息取证和安全交易* 13, 7 (2018), 1626–1640.
- [201] Y. 石, x. 于, k. Sohn, m. 钱德勒和 a. K. 简. 2020.面向深度人脸识别的通用表示学习. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 6817–6826.
- [202] A. Shrivastava, a. 古普塔和 r. 吉尔希克. 2016.使用在线硬示例挖掘训练基于区域的对象检测器. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 761–769.
- [203] K. Simonyan 和 a. 齐泽曼. 2015.用于大规模图像识别的超深卷积网络. *arXiv abs/1409.1556* (2015).
- [204] E. 斯米尔诺夫, a. Melnikov, s. Novoselov, e. Luckyanets, 和 g. Lavrentyeva. 2017.用于人脸表示学习的身重幽灵挖掘. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 1916–1923.
- [205] K. Sohn. 2016.多类 n 对损失目标的改进深度度量学习. *神经信息处理系统进展*. 1857–1865.
- [206] 南. Soltanpour, B. Boufama 和 q. 米 (meter 的缩写)). J. 吴. 2017.三维人脸识别的局部特征方法综述. *模式识别*. 72 (2017), 391–406.
- [207] G. 宋, y. 刘, m. 姜, y. 王, j. 颜, 还有 b. 冷. 2018.超越权衡:提高基于 FCN 的人脸检测器的精确度. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 7756–7764.
- [208] H. 哦宋 y. 项, s. 杰格尔卡和 s. 萨瓦斯. 2016.基于提升结构化特征嵌入的深度度量学习. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 4004–4012.
- [209] J. 宋, j. 张, l. 高, x. 刘和沈. 2018.针对面部衰老和年轻化的双重条件性干眼症. 在 ICAI. 899–905.
- [210] 长度. 宋, m. 张, x. 吴, 还有 r. 他. 2018.敌对的区别异构人脸识别. *AAAI 人工智能会议录 智力*.
- [211] K. 孙, w. 吴, t. 刘, s. 杨, 问. 王, 问. 周, z. 叶, 还有 c. 钱. 2019.一个健壮的运动模糊视频面部标志检测框架. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 5461–5470.
- [212] X. 孙, p. 吴, 还有 s. C. H. 海. 2018.基于深度学习的人脸检测:一种改进的快速 rcnn 方法. *神经计算* 299 (2018), 42–50.
- [213] Y. 孙, y. 陈, x. 王, 和 x. 唐. 2014.基于联合识别验证的深度学习人脸表示. *神经信息处理系统进展*. 1988–1996.
- [214] Y. 孙, c. 程, y. 张, c. 张, l. 郑, z. 王, 和 y. 魏. 2020.圈视:配对相似性优化的统一视角. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 6398–6407.
- [215] Y. 孙, d. 梁, x. 王, 和 x. 唐. 2015.DeepID3:使用超深度神经网络的人脸识别. *ArXiv abs/1502.00873* (2015).
- [216] Y. 孙, x. 王, 和 x. 唐. 2013.用于面部点检测的深度卷积网络级联. 在 IEEE 计算机视觉和模式识别会议记录中. 3476–3483.
- [217] Y. 孙, x. 王, 和 x. 唐. 2013.用于人脸验证的混合深度学习. 在美国电气和电子工程师协会国际计算机视觉会议 (ICCV) 会议录中. 1489–1496.
- [218] Y. 孙, x. 王, 和 x. 唐. 2014.从预测 10, 000 个类中深度学习人脸表示. 2014 年 IEEE 计算机视觉与模式识别会议. 1891–1898.
- [219] Y. 孙, x. 王, 和 x. 唐. 2015.深度学习的人脸表示是稀疏的、选择性的和鲁棒的. *2015 年 IEEE 计算机视觉和模式识别会议 (CVPR)* (2015), 2892–2900.
- [220] Y. 孙, x. 王, 和 x. 唐. 2016.用于人脸识别的稀疏神经网络连接. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录 (CVPR) 中. 4856–4864.
- [221] C. 塞格迪, w. 刘, y. 贾, p. 塞马奈, s. 里德, d. Anguelov, d. 二寒, v. 范豪克和 a. 拉宾诺维奇. 2015.盘旋着越陷越深. *2015 年 IEEE 计算机视觉和模式识别会议 (CVPR)* (2015), 1–9.

- [222] Y. 泰, y. 梁, x. 刘, l. 段, j. 李, c. 王, f. 黄, 和 y. 陈. 2019.面向高分辨率视频的高精度稳定人脸对齐.《AAAI 人工智能会议录》, 第一卷. 33.8893–8900.
- [223] Y. 台格曼, m. 杨, m. 兰扎托, 还有 l. 狼. 2014.Deepface:缩小在人脸验证方面与人类水平的差距.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 1701–1708.
- [224] X. 唐, d. K. 杜, z. 他, 还有 j. 刘. 2018.一个上下文辅助的单镜头人脸检测器.《欧洲计算机视觉会议论文集》(ECCV). 797–813.
- [225] W. 田, z. 王, h. 沈, w. 邓, y. 孟, b. 陈, x. 张, y. 赵, 和 x. 黄. 2018.利用特征融合和分割监督学习更好的人脸检测特征.在 arXiv 预印本 arXiv:1811.08557 中.
- [226] Z. 田, c. 沈, h. 陈和 t. 他. 2019.FCOS:全卷积一级目标检测.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 9626–9635.
- [227] 长度. Tran, x. 阴, 和 x. 刘. 2017.用于姿态不变人脸识别的解纠缠表示学习神经网络. *IEEE 计算机视觉和模式识别会议记录(CVPR) (2017)*, 1283–1292.
- [228] G. Trigeorgis, p. 斯内普, m. A. Nicolaou, e. 安东尼科斯, 和 s. Zafeiriou. 2016.助记下降法:一种应用于端到端对齐的循环过程.在 2016 年 IEEE 计算机视觉和模式识别会议(CVPR)上. 4177–4187.
- [229] 米 (meter 的缩写)). 特克和 a. 彭特兰. 1991.用于识别的特征脸. *认知神经科学杂志* 3, 1 (1991), 71–86.
- [230] T. 邓, w. 邓和 j. 胡. 2017.用于年龄不变人脸识别的年龄估计引导卷积神经网络.《计算机视觉和模式识别研讨会 IEEE 会议记录》(CVPRW) (2017), 503–511.
- [231] 曹那教 V 和 e. 博学——米勒. 2010.FDDB:无约束环境下人脸检测的基准.技术报告 UM-CS-2010-009.马萨诸塞州大学阿默斯特分校.
- [232] 页 (page 的缩写). Viola 和 m. 琼斯. 2001.使用简单特征的增强级联快速目标检测.在 IEEE 计算机视觉和模式识别会议记录中. (CVPR), 第二卷. 1.IEEE, 1–I.
- [233] 赵 w, r. Chellappa, p. J. 菲利普斯, 还有 a. 罗森菲尔德. 2003.人脸识别:文献综述. *美国计算机学会计算调查(CSUR)* 35, 4 (2003), 399–458.
- [234] 南. 万, z. 陈, 张 t, b. 张和 k. 黄. 2016.用硬负面例子引导人脸检测. *arXiv 预印本 arXiv:1608.02236 (2016)*.
- [235] W. 万, y. 高, 还有 h. J. 李. 2019.面向人脸草图识别的迁移深度特征学习. *神经计算与应用(2019)*, 1–10.
- [236] F. 王, l. 陈, c. 李, s. 黄, y. 陈, c. 钱, 还有 c. C. 阿来. 2018.人脸识别的魔鬼在噪音里.《欧洲计算机视觉会议论文集》(ECCV).
- [237] F. 王, j. 程, w. 刘, 和 h. 刘. 2018.用于面部验证的附加余量软件最大值. *电气和电子工程师协会信号处理信函* 25, 7 (2018), 926–930.
- [238] F. 王, m. 姜, c. 钱, s. 杨, c. 李 h. 张, x. 王, 和 x. 唐. 2017.用于图像分类的剩余注意力网络. *IEEE 计算机视觉和模式识别会议记录(CVPR) (2017)*, 6450–6458.
- [239] F. 王, x. 项, j. 程, 还有 a. 长度. 尤伊. 2017.Normface:用于人脸验证的 l2 超球面嵌入.在第 25 届美国计算机学会国际多媒体会议记录中. ACM, 1041–1049.
- [240] H. 王, d. 巩, z. 李, 和 w. 刘. 2019.用于年龄不变人脸识别的去相关敌对学习. *美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR) (2019)*, 3522–3531.
- [241] H. 王, y. 王, z. 周, x. 纪, 丁. 龚, j. 周, z. 李, 和 w. 刘. 2018.Cosface:深度人脸识别的大边缘余弦损失.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5265–5274.
- [242] J. 王. 孙, t. 程, b. 姜, c. 邓, y. 赵, d. 刘, y. 穆, m. 谭, x. 王, w. 刘, 和 b. 小. 2020.用于视觉识别的深度高分辨率表示学习. *模式分析和机器智能的 IEEE 交易(2020)*.
- [243] J. 王, y. 元, b. 李, g. 余, 还有 s. 简. 2018.Sface:用于大规模变化中的人脸检测的有效网络. *arXiv 预印本 arXiv:1804.06559 (2018)*.
- [244] J. 王, y. 元, 和 g. 于. 2017.一种有效的遮挡人脸检测器. *arXiv 预印本 arXiv:1711.07246 (2017)*.
- [245] 长度. 王, v. 辛达吉, 和 v. 米 (meter 的缩写)). 帕特尔. 2018.使用多对抗网络的高质量面部照片草图合成.在美国电气和电子工程师协会自动人脸和手势识别国际会议记录(2018 年). 83–90.
- [246] 米 (meter 的缩写)). 王与 w. 邓. 2018.深度人脸识别:综述. *神经计算* 312 (2018), 135–153.
- [247] 米 (meter 的缩写)). 王与 w. 邓. 2020.使用倾斜感知强化学习减轻人脸识别中的偏差.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 9322–9331.
- [248] 米 (meter 的缩写)). 王, w. 邓, j. 胡, x. 陶, 和 y. 黄. 2019.荒野中的种族面孔:通过信息最大化适应网络减少种族偏见.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 692–702.
- [249] 名词 (noun 的缩写). 王, x. 高, d. 陶, h. 阳, 和 x. 李. 2018.面部特征点检测:综合调查. *神经计算* 275 (2018), 50–65.
- [250] W. 王, z. 崔, y. 颜, j. 冯, s. 颜, x. 舒, 和 n. 塞贝. 2016.复发性面部衰老. 2016 年 IEEE 计算机视觉和模式识别会议(CVPR) (2016), 2378–2386.
- [251] X. 王, l. 博, 还有 l. 阜新. 2019.通过热图回归进行鲁棒面对准的自适应翼损失.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 6970–6980.
- [252] X. 王, s. 王, j. 王, h. 石, 和 t. 梅. 2019.联合挖掘:带噪声标签的深度人脸识别.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 9358–9367.
- [253] X. 王, s. 张, s. 王, t. 傅, h. 石, 和 t. 梅. 2020.用于人脸识别的误分类向量引导的最大损失.在 AAAI 人工智能会议记录中.
- [254] Y. 王, d. 巩, z. 周, x. 纪, h. 王, z. 李, w. 刘和 t. 张. 2018.用于年龄不变人脸识别的正交深度特征分解.《欧洲计算机视觉会议论文集》(ECCV). 738–753.
- [255] Y. 王, x. 纪, 周, 何. 王, 和 z. 李. 2017.使用基于区域的完全卷积网络检测人脸. *arXiv 预印本 arXiv:1709.05256 (2017)*.
- [256] Z. 王, x. 唐, w. 罗, 还有 s. 高. 2018.身份保持条件生成对抗网络下的人脸衰老. *IEEE 计算机视觉和模式识别会议记录(CVPR) (2018)*, 7939–7947.
- [257] Z. 王, l. 郑, y. 李, 和 s. 王. 2019.基于链接的图形卷积网络人脸聚类.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 1117–1125.
- [258] H. 魏, p. 鲁, 和 y. 魏. 2020.人脸识别的平衡对齐:一种联合学习方法. *ArXiv abs/2003.10168 (2020)*.
- [259] K. q. 温伯格和 l. K. 索尔. [n. d.]. 大边缘最近邻分类的距离度量学习.在《神经信息进展》中, 页码 = 1473–1480, 年份=2006.
- [260] Y. 文, z. 李, 和 y. 乔. 2016.用于年龄不变人脸识别的潜在因子引导卷积神经网络. 2016 年 IEEE 计算机视觉和模式识别会议(CVPR) (2016), 4893–4901.
- [261] Y. 文, k. 张, z. 李, 和 y. 乔. 2016.一种用于深度人脸识别的判别特征学习方法.在欧洲计算机视觉会议上. 斯普林格, 499–515.
- [262] C. Whitelam. 塔博尔斯基, a. 布兰顿, b. 梅兹 j. C. 亚当斯, t. 米勒, n. D. 卡尔卡, a. K. Jain, j. A. 邓肯, k. 艾伦, j. 切尼和 p. 格罗瑟. 2017.IARPA Janus 基准-B 人脸数据集. *美国电气和电子工程师协会计算机视觉和模式识别研讨会会议录(2017)*, 592–600.
- [263] 乔纳森·威利福德、布兰登·梅和杰弗里·伯恩. 2020.可解释的人脸识别.《欧洲计算机视觉会议论文集》(ECCV).
- [264] 长度. 狼, t. 哈斯纳和我. Maoz. 2011.背景相似度匹配的无约束视频中的人脸识别.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 529–534.
- [265] W. 吴, m. 侃, x. 刘, y. 杨, s. 山, 和 x. 陈. 2017.用于无对齐人脸识别的递归空间转换器(ReSt).在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 3792–3800.

- [266] W. 吴, c. 钱, s. 杨, 问. 王, y. 蔡, 又问. 周. 2018.看边界:一种边界感知的人脸对齐算法.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 2129–2138.
- [267] X. 吴, r. 他, 还有 z. 星期日. 2015.一个轻松的 CNN 深度人脸表示. *ArXiv abs/1511.02683 (2015)*.
- [268] X. 吴, r. 他, z. 孙, 和 t. 谭. 2018.一个用于带有噪声标签的深度人脸表示的轻型 CNN. *美国电气和电子工程师协会信息取证和安全交易 13 (2018), 2884–2896*.
- [269] X. 吴, h. 黄, v. 米 (meter 的缩写)). Patel. 他, 还有 z. 星期日. 2019.异构人脸识别的解纠缠变分表示.《AAAI 人工智能会议录》, 第一卷. 33.9005–9012.
- [270] Y. 吴, t. 哈斯纳, k. 金, g. Medioni, 还有 p. 纳塔拉詹. 2018.使用扭曲卷积神经网络的面部标志检测. *美国电气和电子工程师协会模式分析和机器智能学报 40, 12 (2018), 3067–3074*.
- [271] Y. 吴, h. 刘, 还有 y. 傅. 2017.基于混合分类器的低镜头人脸识别.在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 1933–1939.
- [272] Y. 吴, y. 吴, r. 巩, y. 吕, k. 陈, d. 梁, x. 胡, x. 刘, 还有 j. 颜. 2020.旋转一致的边缘损失用于高效的低比特人脸识别.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 6866–6876.
- [273] X. 范, r. 刘, k. 呼延, y. 冯, 和 z. 罗. 2018. 用于面对齐的自增强级联回归. *AAAI 会议记录 人工智能*.
- [274] 南. 小, j. 冯, l. 刘, x. 聂, w. 王, s. 颜, 和 a. 卡西姆. 2017.用于大姿态面部标志检测的递归 3D-2D 双重学习.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 1642–1651.
- [275] 南. 小, j. 冯, j. 兴, h. 赖, s. 颜, 和 a. A. 卡西姆. 2016.基于递归关注细化网络的鲁棒人脸标志点检测.《欧洲计算机视觉会议论文集》(ECCV). 斯普林格, 57–72 岁.
- [276] W. 谢, l. 沈, 还有 a. 齐泽曼. 2018.比较者网络.《欧洲计算机视觉会议论文集》(ECCV). 782–797.
- [277] E. 页 (page 的缩写). 兴, a. Y. Ng, m. 爱达荷 (Idaho 的缩写). 乔丹, 还有 s. J. 拉塞尔. 2002.距离度量学习及其在边信息聚类中的应用.在 NIPS.
- [278] C. 徐问. 刘, 和 m. 叶. 2017.耦合自动编码网络的年龄不变人脸识别和检索. *神经计算 222 (2017), 62–71*.
- [279] X. 我和许. A. 卡卡迪亚里斯. 2017.使用全局和局部 cnn 特征的联合头部姿态估计和人脸对准框架.在美国电气和电子工程师协会自动人脸手势识别国际会议记录(2017). 642–649.
- [280] Y. 许, w. 颜, h. 孙, g. 杨, 还有 j. 罗. 2019.中心面:使用面作为点的联合面检测和对齐.在 arXiv 预印本 arXiv:1911.03599. arXiv:cs.CV/1911.03599
- [281] J. 颜, x. 张, z. 雷, 还有 s. Z. 李. 2014.基于结构模型的人脸检测. *图像和视觉计算 32, 10 (2014), 790–799*.
- [282] B. 杨, j. 颜, z. 雷, 还有 s. Z. 李. 2015.野外人脸检测的细粒度评价.在自动人脸和手势识别(FG), 第 11 届 IEEE 国际会议. IEEE.
- [283] J. 杨, a. Bulat, 和 g. Tzimiropoulos. 2020.FAN-Face:对深度人脸识别的简单正交改进.在 AAAI 人工智能会议记录中.
- [284] J. 杨, s. E. 里德, m. 杨, 还有 h. 李. 2015.三维视图合成的弱监督递归变换解纠缠.在神经信息处理系统年度会议记录中. 1099–1107.
- [285] J. 杨, p. 任, d. 张, d. 陈, f. 文, h. 李, 和 g. 华. 2017.视频人脸识别的神经聚合网络. *美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR) (2017), 5216–5225*.
- [286] 长度. 杨, d. 陈, x. 詹, r. 赵, c. C. 阿来, 还有 d. 林. 2020.通过置信度和连通性估计学习人脸聚类.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 13369–13378.
- [287] 长度. 杨, x. 詹, d. 陈, j. 颜, c. C. 阿来, 还有 d. 林. 2019.学习在亲和图上聚类人脸.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 2298–2306.
- [288] 南. 杨, p. 罗, c. C. 阿来, 还有 x. 唐. 2015.从面部反应到人脸检测:一种深度学习方法.在美国电气和电子工程师协会国际计算机视觉会议 (ICCV)会议录中. 3676–3684.
- [289] 南. 杨, p. 罗, c. C. 阿来, 还有 x. 唐. 2015.更宽的脸. http://shuoyang1213.me/WIDERFACE/WiderFace_Results.html. (2015). (2020 年 9 月 15 日访问).
- [290] 南. 杨, p. 罗, c. C. 阿来, 还有 x. 唐. 2016.更宽的脸:一个人脸检测基准.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5525–5533.
- [291] D. 噢, z. 雷, 还有 s. Z. 李. 2015.异构人脸识别的共享表示学习.《自动人脸和手势识别国际会议和研讨会论文集》, 第一卷. 1.1–7.
- [292] D. 噢, z. 雷, 辽, 与 s. Z. 李. 2014.从头开始学习人脸表示. *ArXiv abs/1411.7923 (2014)*.
- [293] Z. 易, h. 张, p. 谭, 和 m. 巩. 2017.图像到图像翻译的无监督双重学习.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 2868–2876.
- [294] 尹邦杰, 栾李, 李, 沈晓辉. 2019.走向可解释的人脸识别.在美国电气和电子工程师协会国际计算机视觉会议记录中. 9348–9357.
- [295] X. 阴, x. 于, k. Sohn, x. 刘, 和 m. 钱德瑞克. 2017.走向野外大姿态正面化.在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中. 4010–4019.
- [296] X. 阴, x. 于, k. Sohn, x. 刘, 和 m. 钱德瑞克. 2019.特征转移学习在欠代表数据人脸识别中的应用.在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5704–5713.
- [297] H. 于, y. 范, k. 陈, h. 颜, x. 鲁, j. 刘, 和马超. 谢. 2019.未知身份拒绝损失:利用未标记数据进行人脸识别.在美国电气和电子工程师协会国际计算机视觉研讨会会议录中. 2662–2669.
- [298] J. 于, y. 江, z. 王, z. 曹, 和 t. 黄. 2016.Unitbox:一个先进的物体探测网络.在第 24 届 ACM 国际多媒体会议记录中. 516–520.
- [299] T. 余, p. Xi. 长, s. 张和马超. 名词 (noun 的缩写). 梅塔克萨斯. 2018.CR-GAN:学习多视图生成的完整表示.在国际人工智能联合会议上.
- [300] Y. 于, g. 宋, m. 张, j. 刘, y. 周, 还有 j. 颜. 2019.面向触发器约束的人脸识别.在美国电气和电子工程师协会计算机视觉研讨会国际会议记录中.
- [301] 长度. 岳, x. 苗, p. 王, b. 张, x. 贞, 和 x. 曹. 2018.注意力校准网络.英国机器视觉会议录, 第一卷. 2.6–13.
- [302] Z. 刘, p. 罗, x. 王, 和 x. 唐. 2015.深度学习野外的人脸属性.《国际计算机视觉会议论文集》(ICCV).
- [303] 南. Zafeiriou, g. Trigeorgis, g. Chrysos, j. 邓, 与纪. 沈. 2017.门面部标志定位挑战:迈向解决方案的一步.在 IEEE 计算机视觉和模式识别研讨会会议记录中. 2116–2125.
- [304] 南. Zafeiriou, c. 张和 z. 张. 2015.野外人脸检测综述:过去、现在和未来.《计算机视觉与图像理解》(2015).
- [305] 蒂莫西·齐(Timothy Zee)、吉塔·加利(Geeta Gali)和伊费欧马·恩沃古(Ifeoma Nwogu). 2019.用可解释的神经网络增强人脸识别.在美国电气和电子工程师协会计算机视觉研讨会国际会议记录中. 0–0.
- [306] D. 曾, h. 刘, f. 赵, s. 葛, w. 沈, 和 z. 张. 2019.用于快速人脸检测的建议金字塔网络. *信息科学 495 (2019), 136–149*.
- [307] X. 詹, z. 刘, j. 颜, d. 林, 还有 c. C. 阿来. 2018.面向人脸识别的一致驱动的海量无标签数据传播. (2018), 568–583.
- [308] B. 张, j. 李, y. 王, y. 泰, c. 王, j. 李, f. 黄, y. 夏, w. 裴, 还有 r. 纪. 2020.ASFD:自动和可伸缩的人脸检测器, 载于 arXiv 预印本 arXiv:2003.11228.ArXiv.
- [309] C. 张, x. 许, 还有马超. 屠. 2018.基于改进的快速 rcnn 的人脸检测. *arXiv 预印本 arXiv:1802.02142 (2018)*.
- [310] C. 张和 z. 张. 2014.用多任务深度卷积神经网络改进多视角人脸检测.在 IEEE 计算机视觉应用冬季会议上. 1036–1041.

- [311] D. 张, l. 林, t. 陈, x. 吴, w. 谭, 和 e. 伊斯基耶多. 2017.通过分解表示学习的内容自适应草图肖像生成. *美国电气和电子工程师协会图像处理学报* 26 (2017), 328–339.
- [312] F. 张. 2006.从一个人的单个图像进行人脸识别:一项调查. *模式识别* 39, 9 (2006), 1725–1745.
- [313] F. 张, x. 范, g. Ai, j. 宋, y. 秦, 还有 j. 吴. 2019. 高性能的精确人脸检测. *arXiv 预印本 arXiv:1905.01585* (2019). arXiv:1905.01585
- [314] G. 张, h. 韩, s. 山, x. 宋, 还有 x. 陈. 2018.基于机器翻译的三维形状重建的大姿态人脸对齐. 在 *IEEE 自动人脸和手势识别国际会议记录中*. 210–217.
- [315] H. 张, b. 南. Riggan, s. 胡, n. J. 短, 还有 v. 米 (meter 的缩写)). 帕特尔. 2019.使用生成对抗网络从偏振热面合成高质量可视面. *国际计算机视觉杂志* 127 (2019), 845–862.
- [316] J. 张和 h. 胡. 2019.面向人脸对齐的显著区域注意力细化叠加沙漏网络连接. 在美国电气和电子工程师协会自动人脸和手势识别国际会议记录中. 1–7.
- [317] J. 张, s. 山, m. 侃, 和 x. 陈. 2014.用于实时人脸对齐的粗到细自动编码器网络(cfan).《欧洲计算机视觉会议论文集》(ECCV). 斯普林格, 1–16.
- [318] J. 张, x. 吴, s. C. 海, 和 j. 朱. 2020.用于单阶段人脸检测的特征聚集网络. *神经计算* 380 (2020), 180–189.
- [319] K. 张, z. 张, z. 李, 和 y. 乔. 2016.基于多任务级联卷积网络的联合人脸检测和对齐. *电气和电子工程师协会信号处理信函* 23, 10 (2016), 1499–1503.
- [320] 长度. 张, l. 林, x. 吴, s. 丁, 和 l. 张. 2015.通过完全卷积表示学习的端到端照片草图生成. (2015), 627aÅ634.
- [321] 米 (meter 的缩写)). 张, y. 李, n. 王, y. Chi, 还有 x. 高. 2020.不同照明下的级联面草图合成. *美国电气和电子工程师协会图像处理学报* 29 (2020), 1507–1521.
- [322] 米 (meter 的缩写)). 张, r. 王, x. 高, j. 李, 和 丁. 陶. 2019.双转移面草图照片合成. *美国电气和电子工程师协会图像处理学报* 28 (2019), 642–657.
- [323] 南. 张, c. 志, z. 雷, 还有 s. Z. 李. 2020.细化:用于高性能人脸检测的细化神经网络. *模式分析和机器学习*(2020).
- [324] 南. 张. 问. 苗, m. 名词 (noun 的缩写)). 黄, x. 朱, y. 陈, z. 雷, 还有 j. 王. 2019.用于真实感人脸正面化的姿态加权甘. 在美国电气和电子工程师协会国际图像处理会议记录中. 2384–2388.
- [325] 南. 张, l. 文, x. 扁, z. 雷, 还有 s. Z. 李. 2018.用于目标检测的单次细化神经网络. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 4203–4212.
- [326] 南. 张, r. 朱, x. 王, h. 史, f. 傅, s. 王, t. 梅, 还有 s. Z. 李. 2019.用于人脸检测的改进的选择性细化网络. *arXiv:计算机视觉和模式识别(2019)*.
- [327] 南. 张, x. 朱, z. 雷, h. 石, x. 王, 和 s. Z. 李. 2017.FaceBoxes:高精度 CPU 实时人脸检测器. 在 *IEEE 生物识别国际联席会议(UCB)*上. 1–9.
- [328] 南. 张, x. 朱, z. 雷, h. 石, x. 王, 和 s. Z. 李. 2017.S3FD:单镜头比例不变人脸检测器. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 192–201.
- [329] 南. 张, x. 朱,、、李. 2018.用密集连通的人脸提议网络检测人脸. *神经计算* 284 (2018), 119–127.
- [330] W. 张, x. 王, 和 x. 唐. 2011.人脸照片-草图识别的耦合信息论编码. *CVPR 2011 (2011)*, 513–520.
- [331] X. 张, z. 方, y. 文, z. 李, 和 y. 乔. 2017.长尾训练数据深度人脸识别的距离损失. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 5419–5428.
- [332] X. 张, r. 赵, y. 乔, x. 王, 和 h. 李. 2019.自适应缩放余弦逻辑以有效学习深度人脸表示. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 10823–10832.
- [333] X. 张, r. 赵, j. 颜, m. 高, y. 乔, x. 王, 和 h. 李. 2019.P2SGrad:优化深层人脸模型的细化梯度. *IEEE 计算机视觉和模式识别会议记录(CVPR) (2019)*, 9898–9906.
- [334] X. 张, x. 周, m. 林, 还有 j. 星期日. 2018.洗牌网:一个非常有效的移动设备卷积神经网络. *IEEE 计算机视觉和模式识别会议记录(CVPR) (2018)*, 6848–6856.
- [335] Y. 张, w. 邓, m. 王, j. 胡, x. 李, d. 赵, 和 维. 文. 2020.全球-本地 GCN:用于人脸识别的大规模标签噪声净化. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中.
- [336] Y. 张, m. 邵, e. K. Wong, 和 y. 傅. 2013.面向姿态不变人脸识别的随机人脸引导稀疏多对一编码器. 在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议记录中. 2416–2423.
- [337] Y. 张, x. 徐, 和 x. 刘. 2019.强大的高性能人脸检测器. arXiv:1901.02350
- [338] Z. 张, x. 陈, b. 王, g. 胡, w. 左, 和 e. R. 汉考克. 2019.使用基于外观流的卷积神经网络的人脸正面化. *美国电气和电子工程师协会图像处理学报* 28 (2019), 2187–2199.
- [339] Z. 张, l. 平, c. 长度. 陈, 和 x. 唐. 2014.基于深度多任务学习的人脸标志点检测. 在欧洲计算机视觉会议上. [340] H. 赵, x. 应, y. 石, x. 童, j. 文, 和 h. 咋. 2020.RDCFace:人脸识别的径向失真校正. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 7721–7730.
- [341] J. 赵, y. 程, y. 页 (page 的缩写)). 程, y. 杨, h. 兰, f. 赵, l. 熊, y. 徐, j. 李, s. 普拉那塔, s. 沈, j. 兴, h. 刘, s. 颜, 还有 j. 冯. 2019.跨越过去看:解开表示学习和照片真实感跨年龄人脸合成的年龄不变的人脸识别. 在 *AAAI 人工智能会议记录中*.
- [342] J. 赵, y. 程, y. 徐, l. 熊, j. 李, f. 赵, k. Jayashree, s. 普拉那塔, s. 沈, j. 兴, s. 颜, 还有 j. 冯. 2018.面向野外姿态不变的人脸识别. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 2207–2216.
- [343] J. 赵, l. 熊, y. 程, y. 程, j. 李, l. 周, y. 徐, j. Karlekar, s. 普拉那塔, s. 沈等. 2018.三维辅助深度姿态不变人脸识别.. 在国际人工智能联合会议, 第. 2.11.
- [344] J. 赵, l. 熊, j. Karlekar, j. 李, f. 赵, z. 王, s. 普拉那塔, s. 沈, s. 颜, 还有 j. 冯. 2017.用于真实感和身份保持的轮廓人脸合成的双代理GANs. 在 *NIPS*.
- [345] K. 赵, k. 许, 和 m. 程. 2019.正则化人脸:通过排他性正则化进行深度人脸识别. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 1136–1144.
- [346] T. 郑和 w. 邓. 2018.交叉姿势 Ifw:一个研究无约束环境下交叉姿势人脸识别的数据库. *北京邮电大学, 理工. 代表(2018)*, 18–01.
- [347] T. 郑, w. 邓和 j. 胡. 2017.跨年龄 Ifw:一个研究无约束环境下跨年龄人脸识别的数据库. *arXiv 预印本 arXiv:1708.08197* (2017).
- [348] Y. 郑, d. K. Pal, 还有 m. Savvides. 2018.环丢失:用于人脸识别的凸特征归一化. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中. 5089–5097.
- [349] Y. 钟, j. 陈, 和 b. 黄. 2017.通过对齐学习实现端到端人脸识别. *电气和电子工程师协会信号处理信函* 24, 8 (2017), 1213–1217.
- [350] 钟、邓. 2018.相似人脸识别中的深度差异分析.《国际模式识别会议论文集》(ICPR). IEEE, 3353–3358.
- [351] Y. 钟, w. 邓, m. 王, j. 胡, j. 彭, x. 陶, 和 y. 黄. 2019.长尾噪声数据深度人脸识别的不等训练. 在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中.
- [352] E. 周, z. 曹, 和 j. 星期日. 2018.Gridface:通过学习局部单元变换进行人脸矫正.《欧洲计算机视觉会议论文集》(ECCV). 3–19.

- [353] E. 周, h. 范, z. 曹, y. 姜和问. 尹. 2013.用粗到细的卷积网络级联进行广泛的面部标志点定位。在美国电气和电子工程师协会计算机视觉研讨会国际会议记录中。386–391.
- [354] 南. K. 周和马超. 柯曼秋. 2007.形状回归机。在两年一次的医学影像信息处理国际会议上。斯普林格, 13–25 岁。
- [355] C. 朱, y. 他, 还有 m. Savvides. 2019.用于单次目标检测的特征选择性无锚模块。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。840–849.
- [356] C. 朱, r. 陶, k. Luu, 还有 m. Savvides. 2018.从健壮主播的角度看小脸。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。5127–5136.
- [357] C. 朱, y. 郑, k. Luu, 还有 m. Savvides. 2017.Cms-rcnn:用于无约束人脸检测的基于上下文多尺度区域的 cnn。生物统计学深度学习。斯普林格, 57–79 岁。
- [358] J. 朱, t. 公园, p. 伊索拉, 还有 a. A. Efros. 2017.使用循环一致对抗网络的不成对图像到图像翻译。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。2242–2251.
- [359] 米 (meter 的缩写)). 朱, f. 史, m. 郑, 和 m. 萨迪克. 2019.基于遮挡自适应深度网络的鲁棒人脸标志检测。在 IEEE 计算机视觉和模式识别会议记录中。3486–3496.
- [360] X. 朱, z. 雷, x. 刘, h. 石, 和 s. Z. 李. 2016.跨大姿态的面对齐:3D 解决方案。在 2016 年 IEEE 计算机视觉和模式识别会议(CVPR)上。146–155.
- [361] X. 朱, z. 雷, j. 颜, 董毅, s. Z. 李. 2015.野外人脸识别的高保真姿态和表情归一化。2015 年 IEEE 计算机视觉和模式识别会议(CVPR) (2015), 787–796。
- [362] X. 朱, h. 刘, z. 雷, h. 史, f. 杨, d. 映, g. 齐, 和 s. Z. 李. 2019.身份识别与现场人脸识别的大规模双样本学习。国际计算机视觉杂志 127, 6–7 (2019), 684–700。
- [363] X. 朱与维. 拉玛南. 2012.野外人脸检测、姿态估计和地标定位。在美国电气和电子工程师协会计算机视觉和模式识别会议记录(CVPR)中。
- [364] Z. 朱, p. 罗, x. 王, 和 x. 唐. 2013.深度学习身份保持人脸空间。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。113–120.
- [365] Z. 朱, p. 罗, x. 王, 和 x. 唐. 2014.多视角感知器:一种学习人脸身份和视角表征的深度模型。在神经信息处理系统年度会议记录中。
- [366] Z. 朱, p. 罗, x. 王, 和 x. 唐. 2014.用深度神经网络恢复野生的典型视图人脸。(2014). arXiv:1404.3543
- [367] C. 庄, s. 张, x. 朱, z. 雷, j. 王, 和 s. Z. 李. 2019.一个中央处理器实时联合人脸和地标检测器。2019 年生物识别国际会议(ICB)。1–8.
- [368] C. 长度. 兹尼克和 p. 美元. 2014.边缘框:从边缘定位对象建议。《欧洲计算机视觉会议论文集》(ECCV)。斯普林格, 391–405。
- [369] X. 邹, j. 凯特勒和 k. 刀子乐队. 2007.光照不变人脸识别综述。在美国电气和电子工程师协会生物统计学国际会议论文集:理论、应用和系统中。1–8.
- [370] X. 邹, s. 钟, l. 颜, x. 赵, j. 周, 和 y. 吴. 2019.通过分层结构集成学习鲁棒的面部标志点检测。在美国电气和电子工程师协会国际计算机视觉会议(ICCV)会议录中。141–150.
- [371] O. eliktutan, S.Ulukaya 和 b. 桑库尔. 2013.人脸标记技术的比较研究。欧洲图像与视频处理杂志 2013 (2013), 1–27。