人脸检测与识别文献调研报告

一、 引言

随着计算机技术的高速发展和人工智能推向高潮，AI 技术在各个领域有着迫切的应用需求，从自动考勤、无人卡口，到智能家居、智能安防，工业自动化，再到临床诊疗，金融投资，用户群体从行业到企业再到个人。人脸识别作为计算机视觉和 AI 领域中十分重要的一个研究方向，在经过多年发展后，已逐步走向实用化，成功应用在身份验证，智能安防，图像检索，交通管制等公共及信息安全领域。

相比于目前常用的用于识别的生物特征，人脸识别凭借其便捷快速，高智能化，人机友好，安全稳定，不宜伪造等优势一跃成为了身份识别领域的热门技术。随着各个应用领域对于安全防护的要求越来越高，人脸识别技术也毫无疑问拥有巨大发展前景和市场空间。

一个人脸检测与识别系统，包括以下 4 个方面的内容：（1）人脸检测(Detection):即从各种不同的场景中检测出人脸的存在并确定其位置。(2)人脸的规范化(Normalization):校正人脸在尺度、光照和旋转等方面的变化，也称为人脸对齐。(3)人脸表征(Face Representation):采取某种方式表示检测出人脸和数据库中的已知人脸。(4)人脸识别(Reoognition):将待识别的人脸与数据库中的已知人脸比较，得出相关信息。人脸识别框架如下图所示。

二、 国内外研究现状

2.1 国外研究现状

19 世纪末 20 世纪初法国人 Galton 等人在自然杂志上发表了一篇关于使用人脸图像进行识别的文章，这是一种非自动识别人脸的方法，这是人脸识别技术的研究开端。20 世纪中期，科学家才提出了自动识别人脸的算法，到了 20 世纪末，人脸识别研究逐渐成为的研究热点。20 世纪 90 年代以后，随着计算机技术的飞速发展，人脸识别已经成为人工智能和模式识别领域的一个重要研究课题，得到了快速的发展。自 21 世纪初以来，人们对人脸识别系统的识别准确性

和适应性提出了越来越高的要求。目前，理想的人脸识别效果已达到商业应用水平。近年来，在巨大的社会需求和商业价值的情况下，在复杂的条件下的人脸识别技术也取得了重大突破。下面，根据近 50 年以来人脸识别领域的研究成果，将人脸识别技术的研究进程分为以下三个阶段：

第一阶段：20 世纪 60 年代至 20 世纪 90 年代年。这是人脸识别研究的初期阶段，这一时期的人脸识别研究研究的内容较为简单和基础，主要研究的是基于人脸几何特征的方法。

第二阶段：20 世纪 90 年代年至 21 世纪初。这是人脸识别技术发展的黄金时期，随着计算机技术的发展，人脸识别技术在这一时期取得了非常快速的发展，在此期间，研究人员提出了很多人脸识别经典的算法，包括融合线性判别分析（LDA）的 Fisher face 人脸识别方法、主成分分析（PCA）为主特征脸方法、局部特征识别法、基于统计的模式识别方法以及弹性图匹配技术等，在识别率和识别速度上有了大幅提升。

第三阶段：21 世纪初至今。在此期间，研究人员针对消除表情、光照、姿态变化等因素对人脸识别系统性能的影响等方面做了大量工作，相继提出了许多改进的算法，这些算法均在一定程度上解决了环境因素造成的识别率低等问题，并且一些算法成功用于商业系统中。目前，国外在人脸识别领域的相关研究已经取得了较大的突破。目前比较先进的人脸识别算法是 3D 模型的重建算法，大大提高了现有人脸识别系统的识别率。当前，国外比较突出的人脸识别机构主要集中在欧美，其中美国一些大学和研究机构在人脸识别领域取得了较为丰硕的成果。这些大学和研究机构包括美国国防部的人脸识别技术 FEACE，卡内基梅隆大学机器人研究和交互系统实验室，Harvard 大学、东京大学、Facebook 的 Deep Face 和 ATR 研究所，微软、IBM 等。

2.2 国内研究现状

国内在人脸识别领域的研究起步较晚，大概开始于 20 世纪 90 年代初期，但研究技术发展很快。国内对人脸识别的研究主要集中在一些大学和研究机构，虽然近年来取得了一定的成果，但与国外技术相比还存在差距。香港中文大学汤晓鸥团队提出将卷积神经网络应用到人脸识别上，采用 20 万训练数据，在 LFW 上第一次得到超过人类水平的识别精度，这是人脸识别发展历史上的一座里程碑。目前，国内一些具有代表性的人脸识别方法和技术介绍如下：

1.清华大学彭辉、张长水等深入研究了多模板的人脸检测问题，取得了不少

成果，在确保识别效果的同时有效降低了运算量。

2.清华大学苏光老师提出的人脸识别系统目前识别速度世界最快，它的设计数据运行达到每秒 2560000 幅的识别速度。

3.四川大学周激流等人研究了“稳定视点”的特征提取方法。在该系统中，使用积分投影方法来提取面部特征，该系统还具有反馈机制，正面脸部识别系统识别效果较好。

4.东南大学富清等人用基于神经网络的 PCA 方法提取人脸图像的特征，并利用提取出来的主特征向量进行人脸识别以及复原人脸图像，取得了较好的效果。

5.南京理工大学杨静宇等人对人脸识别领域应用神经网络方法进行了深入研究，并在基于 Fisher 最佳判别向量的人脸识别方面做出了很大贡献。

6.哈尔滨工业大学高文等研究了多候选加权识别方法，他们还对基于彩色人脸图像识别方法的人脸跟踪算法进行了深入的探讨和研究。

7.复旦大学和中科院自动化所等都对人脸 3D 模型重建进行了研究，并对基于 3D 模型的人脸识别做出了一定的贡献。

8.李子青带领中科院自动化所的研究人员利用近红外人脸识别技术，实现了中远程人脸识别，系统可用于中长距离，快速准确地跟踪多个人脸，具有高可靠性，已达到了国际先进水平并在公安、边检、司法、政府等多部门使用并发挥了巨大作用。以上这些人脸识别的方法和技术的研究和应用极大地推动了国内人脸识别技术的发展和进步，为后面的研究奠定了坚实的基础。

9. 中科院自动化所和京东 AI 研究院最近提出的人脸检测技术,提出了一种改进的 SRN 人脸检测器，并在广泛使用的人脸检测基准 WIDER FACE 数据集上获得了最佳性能。

10.前不久深圳大学于仕琪提出 libfacedetection,其人脸加测速度最高达 1500FPS 是目前最强的人脸检测开源库，同时其能检测的最小脸大小小至 12*12.

三、 综述分析

3.1 传统方法分析

1. 基于几何特征的人脸识别

前面有提到，Bledsoe 提出了最早的人脸识别相关的学术论文，其中用到的便是基于几何特征的方法。基于几何特征人脸识别的实现步骤：定位人脸的面部五官特征点，并测量这几个特征点间的欧氏距离，得到一组能够描述面部特征点的矢量，如特征点的位置坐标、宽度等，或者眉毛的弯曲程度、淡浓等，以及他

们之间的联系。通过计算特征点间的距离，即可找出类似的人脸。该方法的优点主要有：因为只需要存储特征矢量即可，故所需的存储空间较小；对光照的变化具有一定的鲁棒性；这也是常用的人脸识别原理，易于理解。但是想要从一个未知的人脸图像中提取具有稳定性的特征矢量仍具有一定的复杂性，因此近年来对该算法的研究在逐渐地减少。

2. 基于特征脸的人脸识别

基于特征脸方法的核心技术是主成分分析(Principal Component Analysis,PCA)。PCA 算法最早由 Sirovitch 和 Kirby 引入人脸识别领域，它是一种基于 K-L 变换(Karhunen-Loevetransform)的算法，通过对人脸图像进行统计特征提取，从而实现在子空间模式下进行人脸识别。这种方法主要是利用系数变换，因此具有简便、快捷、实用的优点。但该方法对训练样本与测试样本之间的相关性要求较高，且受光照和面部表情的变化影响较大，导致识别率较低。

3. 基于模板匹配的人脸识别

事先给定一些包含不同人脸特征尺度的样本作为模板的方法称为模板匹配方法。其主要原理是：通过比较待测人脸与给定模板的图像窗口，来判定图像窗口中是不是有待测人脸。然而该方法只是简单机械的对比未知图像和已知模板，因此无法处理姿态和尺度变化丰富的人脸，当然这种简单机械的方法比较容易实现。所以目前应用得也较少。

4. 基于神经网络的人脸识别

如今学术界讨论比较多的人脸识别算法便是基于神经网络的方法，BP(Back Propagation)神经网络学习算法是该技术的核心内容。它是受生物神经网络单元间行为特征的启发而建立的一种仿生的运算模型。其主要步骤也依赖于所设计的神经网络：将待识别人脸图像中的像素点与事先设计好的神经网络中的神经元依次对应。这种算法的特点是：较容易提取到人脸特征，通过学习就能得到人脸识别的规律，适应于不同的人脸，而且由于是并行处理信息，速度较快。但是若神经元数量太多，样本的训练时间就会增加。

5. 基于隐马尔可夫模型的人脸识别

Samaria 等人最先提出人脸马尔可夫模型理论，他将人脸划分为 5 个代表性特征区域：额头、眼睛、鼻子、嘴巴和下巴，并且这 5 个特征区域均可用一组特

征值表示。这种方法的优点即是：由于特征区域间位置的相对性，无论人脸表情如何变化，该算法的识别精度都较高。但是这种方法实现起来不仅复杂度高，运算量大，而且还需要大量的内存用于存储特征。

6. 基于弹性匹配方法的人脸识别基于动态链接结构(Dynamic Link Architecture,DLA)是弹性图匹配算法的核心思想。在 1992 年，Lades M 等人曾首次在人脸识别中应用该方法，并取得了较为满意的成果。它通过使用网状的稀疏图来表示人脸图像，并在网状稀疏图的节点处对人脸图像实现 Gabor 小波分解，从而获得特征向量标记，然后对模型图进行搜索，找到最相似于待测样本脸的图，并将图中的节点按照相似性进行依次匹配，形成一个变形图。弹性匹配方法能在光照不好或脸上有装饰的情况下使用，而且当数据库较大时，该方法可以利用神经系统中的突触将神经元按照图的结构划分为若干个组织，从而很大程度上减少了识别时间。但是正因为所有输入的人脸模型都需要计算，因此该算法不仅计算量大，而且需要的内存空间也很大。

7. 基于贝叶斯决策的人脸识别

在样本有缺失的情况下，可以利用贝叶斯决策的方法实现人脸识别。该算法的主要步骤是：首先利用主观概率对某些未知的数据预估，然后使用贝叶斯公式较正发生概率，最后根据较正后的概率作出判断。贝叶斯决策算法主要是利用最大后验概率准则来解决样本有缺失情况下的分类问题，但因总需要计算概率，所以算法较为复杂。

8. 基于支持向量机的人脸识别

由于机器学习算法的流行，支持向量机(support vector machine,SVM)也成为现如今比较热门的识别方法，最早由 Vapnik 等人提出的。它的主要思想是致力于结构风险降到最低，常应用于分类回归的相关问题，其中最重要的原则是在进行样本训练时，学习机器的数量需要对应于训练样本数量。最早在人脸检测过程中使用 SVM 算法的是 Osuna 等人，用于真实人脸和虚假人脸的分类。SVM 算法与传统算法不同的是，不需要对图像进行空间降维，而是根据实际情况，对图像进行空间升维，即将低维映射到高维，从而实现非线性问题到线性问题的转换。目前，SVM 算法是在训练样本定量的情况下，应用最多的机器学习方法。它依靠严谨的理论，能够很好地实现小样本量、非线性以及高维等实际应用中常见问

题的解决。然而在进行样本训练时会耗费大量的内存是它唯一的缺点。

3.2 网络模型方法分析

1. ASM (Active Shape Models)

ASM(Active Shape Model)[1] 是由 Cootes 于 1995 年提出的经典的人脸关键点检测算法，主动形状模型即通过形状模型对目标物体进行抽象，ASM 是一种基于点分布模型（Point Distribution Model, PDM）的算法。在 PDM 中，外形相似的物体，例如人脸、人手、心脏、肺部等的几何形状可以通过若干关键点（landmarks）的坐标依次串联形成一个形状向量来表示。ASM 算法需要通过人工标定的方法先标定训练集，经过训练获得形状模型，再通过关键点的匹配实现特定物体的匹配。

ASM 主要分为两步：第一步：训练。首先，构建形状模型：搜集 n 个训练样本（n=400）；手动标记脸部关键点；将训练集中关键点的坐标串成特征向量；对形状进行归一化和对齐（对齐采用 Procrustes 方法）；对对齐后的形状特征做 PCA 处理。接着，为每个关键点构建局部特征。目的是在每次迭代搜索过程中每个关键点可以寻找新的位置。局部特征一般用梯度特征，以防光照变化。有的方法沿着边缘的法线方向提取，有的方法在关键点附近的矩形区域提取。第二步：搜索。首先：计算眼睛（或者眼睛和嘴巴）的位置，做简单的尺度和旋转变化，对齐人脸；接着，在对齐后的各个点附近搜索，匹配每个局部关键点（常采用马氏距离），得到初步形状；再用平均人脸（形状模型）修正匹配结果；迭代直到收敛。

ASM 算法的优点在于模型简单直接，架构清晰明确，易于理解和应用，而且对轮廓形状有着较强的约束，但是其近似于穷举搜索的关键点定位方式在一定程度上限制了其运算效率。

2. AAM（Active Appearance Models）

1998 年，Cootes 对 ASM 进行改进，不仅采用形状约束，而且又加入整个脸部区域的纹理特征，提出了 AAM 算法[2]。AAM 于 ASM 一样，主要分为两个阶段，模型建立阶段和模型匹配阶段。其中模型建立阶段包括对训练样本分别建立形状模型 (Shape Model) 和纹理模型 (Texture Model)，然后将两个模型进行结合，形成 AAM 模型。

3. CPR（Cascaded pose regression）

2010 年，Dollar 提出 CPR（Cascaded Pose Regression, 级联姿势回归）[4]，CPR 通过一系列回归器将一个指定的初始预测值逐步细化，每一个回归器都依靠前一个回归器的输出来执行简单的图像操作，整个系统可自动的从训练样本中学习。

人脸关键点检测的目的是估计向量

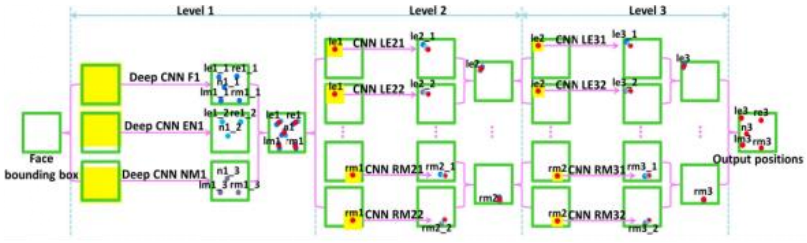
$$S=(x_1, \ldots, x_k, \ldots, x_K)\in \boldsymbol{R}^{2K}$$

，其中 K 表示关键点的个数，由于每个关键点有横纵两个坐标，所以 S 得长度为 2K。CPR 检测流程如图所示，一共有 T 个阶段，在每个阶段中首先进行特征提取，得到, 这里使用的是 shape-indexed features，也可以使用诸如 HOG、SIFT 等人工设计的特征，或者其他可学习特征（learning based features），然后通过训练得到的回归器 R 来估计增量 ΔS(update vector)，把 ΔS 加到前一个阶段的 S 上得到新的 S，这样通过不断的迭代即可以得到最终的 S(shape)。

4. DCNN

2013 年，Sun 等人 [5] 首次将 CNN 应用到人脸关键点检测，提出一种级联的 CNN（拥有三个层级）——DCNN(Deep Convolutional Network)，此种方法属于级联回归方法。作者通过精心设计拥有三个层级的级联卷积神经网络，不仅改善初始不当导致陷入局部最优的问题，而且借助于 CNN 强大的特征提取能力，获得更为精准的关键点检测。

如图所示，DCNN 由三个 Level 构成。Level-1 由 3 个 CNN 组成；Level-2 由 10 个 CNN 组成（每个关键点采用两个 CNN）；Level-3 同样由 10 个 CNN 组成。



Level-1 分 3 个 CNN，分别是 F1（Face 1）、EN1（Eye，Nose）、NM1（Nose，Mouth）；F1 输入尺寸为 39*39，输出 5 个关键点的坐标；EN1 输入尺寸为 39*31，输出是 3 个关键点的坐标；NM11 输入尺寸为 39*31，输出是 3 个关

键点。Level-1 的输出是由三个 CNN 输出取平均得到。

Level-2，由 10 个 CNN 构成，输入尺寸均为 15*15，每两个组成一对，一对 CNN 对一个关键点进行预测，预测结果同样是采取平均。

Level-3 与 Level-2 一样，由 10 个 CNN 构成，输入尺寸均为 15*15，每两个组成一对。Level-2 和 Level-3 是对 Level-1 得到的粗定位进行微调，得到精细的关键点定位。

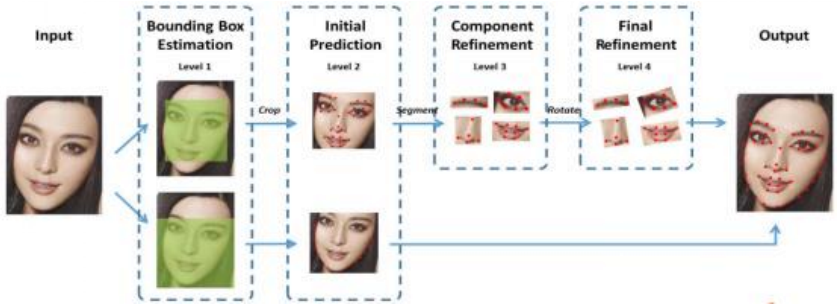
Level-1 之所以比 Level-2 和 Level-3 的输入要大，是因为作者认为，由于人脸检测器的原因，边界框的相对位置可能会在大范围内变化，再加上面部姿态的变化，最终导致输入图像的多样性，因此在 Level-1 应该需要有足够大的输入尺寸。Level-1 与 Level-2 和 Level-3 还有一点不同之处在于，Level-1 采用的是局部权值共享（Lcally Sharing Weights），作者认为传统的全局权值共享是考虑到，某一特征可能在图像中任何位置出现，所以采用全局权值共享。然而，对于类似人脸这样具有固定空间结构的图像而言，全局权值共享就不奏效了。因为眼睛就是在上面，鼻子就是在中间，嘴巴就是在下面的。所以作者借鉴文献 [28] 中的思想，采用局部权值共享，作者通过实验证明了局部权值共享给网络带来性能提升。

DCNN 采用级联回归的思想，从粗到精的逐步得到精确的关键点位置，不仅设计了三级级联的卷积神经网络，还引入局部权值共享机制，从而提升网络的定位性能。最终在数据集 BioID 和 LFPW 上均获得当时最优结果。速度方面，采用 3.3GHz 的 CPU，每 0.12 秒检测一张图片的 5 个关键点。

5. Face++版 DCNN

2013 年，Face++在 DCNN 模型上进行改进，提出从粗到精的人脸关键点检测算法[6]，实现了 68 个人脸关键点的高精度定位。该算法将人脸关键点分为内部关键点和轮廓关键点，内部关键点包含眉毛、眼睛、鼻子、嘴巴共计 51 个关键点，轮廓关键点包含 17 个关键点。

针对内部关键点和外部关键点，该算法并行的采用两个级联的 CNN 进行关键点检测，网络结构如图所示。



针对内部 51 个关键点，采用四个层级的级联网络进行检测。其中，Level-1 主要作用是获得面部器官的边界框；Level-2 的输出是 51 个关键点预测位置，这里起到一个粗定位作用，目的是为了给 Level-3 进行初始化；Level-3 会依据不同器官进行从粗到精的定位；Level-4 的输入是将 Level-3 的输出进行一定的旋转，最终将 51 个关键点的位置进行输出。针对外部 17 个关键点，仅采用两个层级的级联网络进行检测。Level-1 与内部关键点检测的作用一样，主要是获得轮廓的 bounding box；Level-2 直接预测 7 个关键点，没有从粗到精定位的过程，因为轮廓关键点的区域较大，若加上 Level-3 和 Level-4，会比较耗时间。最终面部 68 个关键点由两个级联 CNN 的输出进行叠加得到。

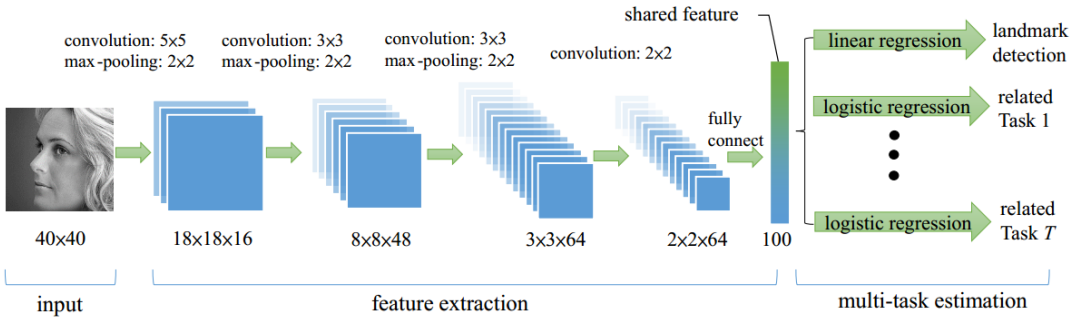
算法主要创新点由以下三点：（1）把人脸的关键点定位问题，划分为内部关键点和轮廓关键点分开预测，有效的避免了 loss 不均衡问题；（2）在内部关键点检测部分，并未像 DCNN 那样每个关键点采用两个 CNN 进行预测，而是每个器官采用一个 CNN 进行预测，从而减少计算量；（3）相比于 DCNN，没有直接采用人脸检测器返回的结果作为输入，而是增加一个边界框检测层（Level-1），可以大大提高关键点粗定位网络的精度。

Face++版 DCNN 首次利用卷积神经网络进行 68 个人脸关键点检测，针对以往人脸关键点检测受人脸检测器影响的问题，作者设计 Level-1 卷积神经网络进一步提取人脸边界框，为人脸关键点检测获得更为准确的人脸位置信息，最终在当年 300-W 挑战赛上获得领先成绩。

6. TCDCN

2014 年，Zhang 等人将 MTL（Multi-Task Learning）应用到人脸关键点检测中，提出 TCDCN（Tasks-Constrained Deep Convolutional Network）[7]。作者认为，在进行人脸关键点检测任务时，结合一些辅助信息可以帮助更好的定位关键点，这些信息如，性别、是否带眼镜、是否微笑和脸部的姿势等等。作者将人脸关键

点检测（5 个关键点）与性别、是否带眼镜、是否微笑及脸部的姿势这四个子任务结合起来构成一个多任务学习模型，模型框架如图所示。



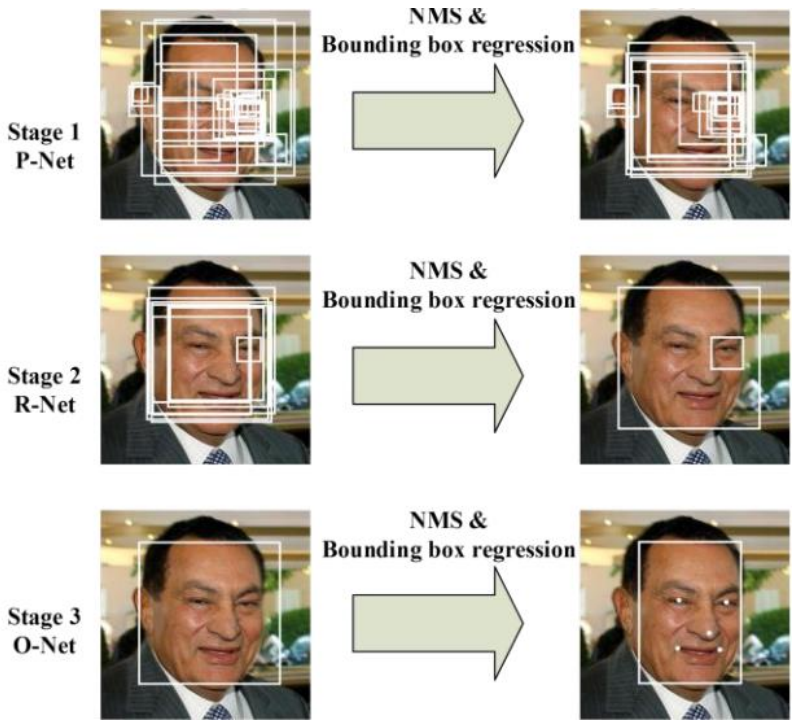
网络输出为 40*40 的灰度图，经过 CNN 最终得到 2*2*64 的特征图，再通过一层含 100 个神经元的全连接层输出最终提取得到的共享特征。该特征为所有任务共同享用，对于关键点检测问题，就采用线性回归模型；对于分类问题，就采用逻辑回归。

TCDCN 采用多任务学习方法对人脸关键点进行检测，针对多任务学习在人脸关键点检测任务中的两个主要问题——不同任务学习难易程度不同以及不同任务收敛速度不同，分别提出了新目标函数和提前停止策略加以改进，最终在 AFLW 和 AFW 数据集上获得领先的结果。同时对比于级联 CNN 方法，在 Intel Core i5 cpu 上，级联 CNN 需要 0.12s，而 TCDCN 仅需要 17ms，速度提升七倍有余。

7. MTCNN

2016 年，Zhang 等人提出一种多任务级联卷积神经网络（MTCNN, Multi-task Cascaded Convolutional Networks）[9] 用以同时处理人脸检测和人脸关键点定位问题。作者认为人脸检测和人脸关键点检测两个任务之间往往存在着潜在的联系，然而大多数方法都未将两个任务有效的结合起来，本文为了充分利用两任务之间潜在的联系，提出一种多任务级联的人脸检测框架，将人脸检测和人脸关键点检测同时进行。

MTCNN 包含三个级联的多任务卷积神经网络，分别是 Proposal Network (P-Net)、Refine Network (R-Net)、Output Network (O-Net)，每个多任务卷积神经网络均有三个学习任务，分别是人脸分类、边框回归和关键点定位。网络结构如图所示：



MTCNN 实现人脸检测和关键点定位分为三个阶段。首先由 P-Net 获得了人脸区域的候选窗口和边界框的回归向量，并用该边界框做回归，对候选窗口进行校准，然后通过非极大值抑制（NMS）来合并高度重叠的候选框。然后将 P-Net 得出的候选框作为输入，输入到 R-Net，R-Net 同样通过边界框回归和 NMS 来去掉那些 false-positive 区域，得到更为准确的候选框；最后，利用 O-Net 输出 5 个关键点的位置。

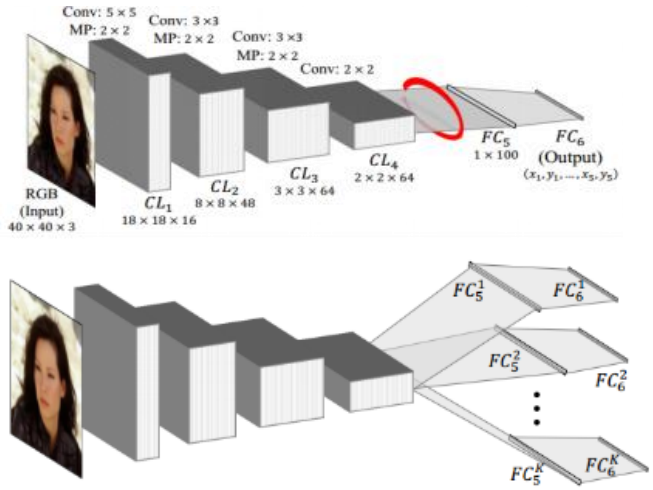
为了提升网络性能，需要挑选出困难样本（Hard Sample），传统方法是通过研究训练好的模型进行挑选，而本文提出一种能在训练过程中进行挑选困难的在线挑选方法。方法为，在 mini-batch 中，对每个样本的损失进行排序，挑选前 70% 较大的损失对应的样本作为困难样本，同时在反向传播时，忽略那 30% 的样本，因为那 30% 样本对更新作用不大。

实验结果表明，MTCNN 在人脸检测数据集 FDDB 和 WIDER FACE 以及人脸关键点定位数据集 LFPW 均获得当时最佳成绩。在运行时间方面，采用 2.60GHz 的 CPU 可以达到 16fps，采用 Nvidia Titan Black 可达 99fps。

8. TCNN（Tweaked Convolutional Neural Networks）

2016 年，Wu 等人研究了 CNN 在人脸关键点定位任务中到底学习到的是什么样的特征，在采用 GMM（Gaussian Mixture Model, 混合高斯模型）对不同

层的特征进行聚类分析，发现网络进行的是层次的，由粗到精的特征定位，越深层提取到的特征越能反应出人脸关键点的位置。针对这一发现，提出了 TCNN（Tweaked Convolutional Neural Networks）[8]，其网络结构如图所示：

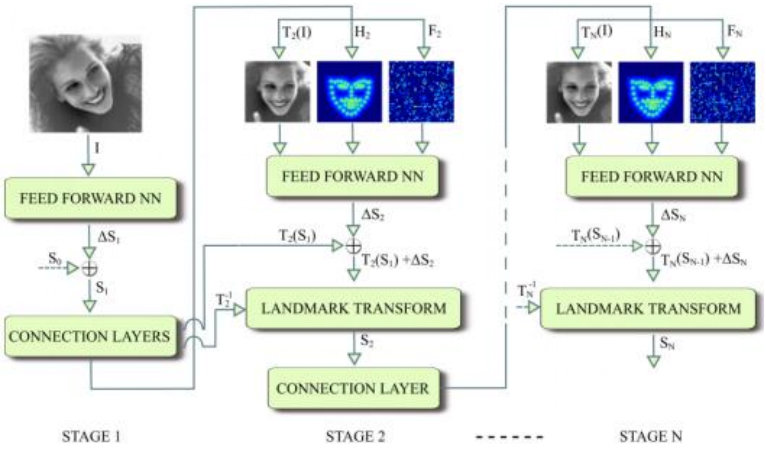


上图为 Vanilla CNN，针对 FC5 得到的特征进行 K 个类别聚类，将训练图像按照所分类别进行划分，用以训练所对应的 FC6K。测试时，图片首先经过 Vanilla CNN 提取特征，即 FC5 的输出。将 FC5 输出的特征与 K 个聚类中心进行比较，将 FC5 输出的特征划分至相应的类别中，然后选择与之相应的 FC6 进行连接，最终得到输出。

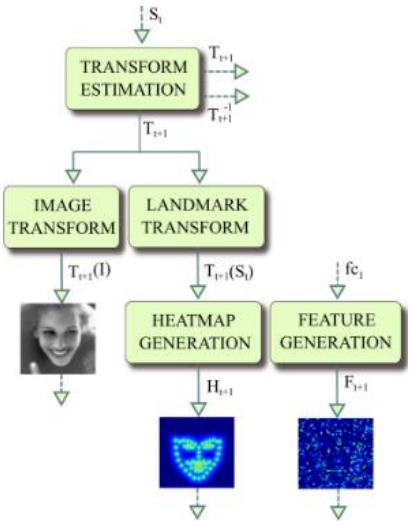
作者使用具有相似特征的图片训练对应的回归器，最终在人脸关键点检测数据集 AFLW,AFW 和 300W 上均获得当时最佳效果。

9. DAN（Deep Alignment Networks）

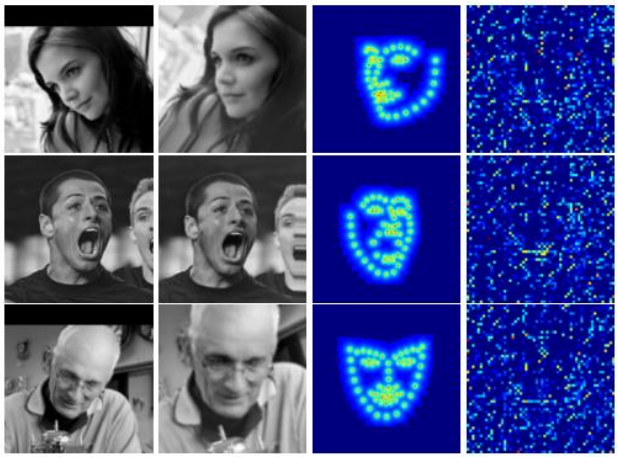
2017 年，Kowalski 等人提出一种新的级联深度神经网络——DAN（Deep Alignment Network）[10]，以往级联神经网络输入的是图像的某一部分，与以往不同，DAN 各阶段网络的输入均为整张图片。当网络均采用整张图片作为输入时，DAN 可以有效的克服头部姿态以及初始化带来的问题，从而得到更好的检测效果。之所以 DAN 能将整张图片作为输入，是因为其加入了关键点热图（Landmark Heatmaps），关键点热图的使用是本文的主要创新点。DAN 基本框架如图所示：



DAN 包含多个阶段，每一个阶段含三个输入和一个输出，输入分别是被矫正过的图片、关键点热图和由全连接层生成的特征图，输出是面部形状（Face Shape）。其中，CONNECTION LAYER 的作用是将本阶段得输出进行一系列变换，生成下一阶段所需要的三个输入，具体操作如下图所示：



第一阶段的输入仅有原始图片和 S0。面部关键点的初始化即为 S0，S0 是由所有关键点取平均得到，第一阶段输出 S1。对于第二阶段，首先，S1 经第一阶段的 CONNECTION LAYERS 进行转换，分别得到转换后图片 T2（I）、S1 所对应的热图 H2 和第一阶段 fc1 层输出，这三个正是第二阶段的输入。如此周而复始，直到最后一个阶段输出 SN。文中给出在数据集 IBUG 上，经过第一阶段后的 T2（I）、T2（S1）和特征图，如图所示：



从图中发现，DAN 要做的“变换”，就是把图片给矫正了，第一行数据尤为明显，那么 DAN 对姿态变换具有很好的适应能力，或许就得益于这个“变换”。至于 DAN 采用何种“变换”，需要到代码中具体探究。

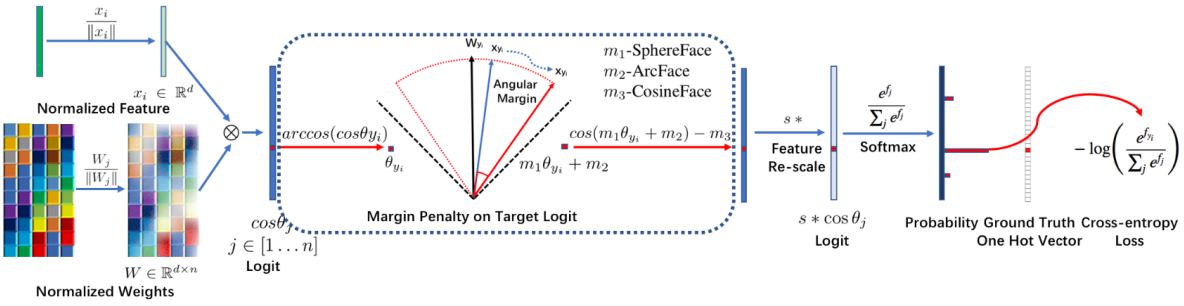
总而言之，DAN 是一个级联思想的关键点检测方法，通过引入关键点热图作为补充，DAN 可以从整张图片进行提取特征，从而获得更为精确的定位。

10. Insight Face

2018 年，Jiankang Deng, Jia Guo 提出的论文[15]ArcFace: Additive Angular Margin Loss for Deep Face Recognition。论文原名是 ArcFace，但是由于与虹软重名，后改名为 Insight Face。这篇论文可以看作是 AmSoftmax/CosFace 的一种改进版本，总体思路相对较为简单。

以往接触的分类问题有很大一部分使用了 Softmax loss 来作为网络的损失层，实验表明 Softmax loss 考虑到样本是否能正确分类，而在扩大异类样本间的类间距离和缩小同类样本间的类内距离的问题上有很大的优化空间，因而作者在 Arcface 文章中讨论了 Softmax 到 Arcface 的变化过程，同时作者还指出了数据清洗的重要性，改善了 Resnet 网络结构使其“更适合”学习人脸的特征。

下图是 ArcFace 的输入到输出流程：



假设样本类别数为 nn，输入数据 x 的维度为 dd，模型权重 ww 的维度为

d*nd*n，首先对样本 xx 和权重 ww 进行归一化，归一化之后样本经过网络最后得到 1·n1·n 维的全连接输出，输出后计算得到 Target Logit 再乘以归一化参数 ss 再经过 Softmax 计算得到 Prob。此文算法在也是当时领先的人脸识别算法

11.libfacedetection

2019 年，深圳大学于仕琪提出 libfacedetection[13]人脸检测算法库。相比于 OpenCV 自带的 CascadeClassifier 人脸检测，无论在速度上还是精度上，都有巨大的优势，是目前已知开源库中最好的一款。如下图是在 pc 和树莓派上运行速度测试结果。其最高速度已达到 1500FPS，是目前最快速的人脸检测方案。

CNN-based Face Detection on Windows

Method	Time	FPS	Time	FPS
	X64	X64	X64	X64
	Single-thread	Single-thread	Multi-thread	Multi-thread
OpenCV Haar+AdaBoost (640x480)	--	--	12.33ms	81.1
cnn (CPU, 640x480)	64.21ms	15.57	15.59ms	64.16
cnn (CPU, 320x240)	15.23ms	65.68	3.99ms	250.40
cnn (CPU, 160x120)	3.47ms	288.08	0.95ms	1052.20
cnn (CPU, 128x96)	2.35ms	425.95	0.64ms	1562.10

- OpenCV Haar+AdaBoost runs with minimal face size 48x48
- Face detection only, and no landmark detection included
- Minimal face size ~12x12
- Intel(R) Core(TM) i7-7700 CPU @ 3.6GHz

CNN-based Face Detection on ARM Linux (Raspberry Pi 3 B+)

Method	Time	FPS	Time	FPS
	Single-thread	Single-thread	Multi-thread	Multi-thread
cnn (CPU, 640x480)	512.04ms	1.95	174.89ms	5.72
cnn (CPU, 320x240)	123.47ms	8.10	42.13ms	23.74
cnn (CPU, 160x120)	27.42ms	36.47	9.75ms	102.58
cnn (CPU, 128x96)	17.78ms	56.24	6.12ms	163.50

- Face detection only, and no landmark detection included.
- Minimal face size ~12x12
- Raspberry Pi 3 B+, Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz

表 1 中列出目前部分人脸检测方案的识别率[14]。

表 1. 部分现有人脸检测识别方案与识别率

人脸检测方案	识别率
Simile classifiers	0.8472 ± 0.0041
Attribute and Simile classifiers	0.8554 ± 0.0035
Multiple LE + comp	0.8445 ± 0.0046
Associate-Predict	0.9057 ± 0.0056
Tom-vs-Pete	0.9310 ± 0.0135
Tom-vs-Pete + Attribute	0.9330 ± 0.0128
combined Joint Bayesian	0.9242 ± 0.0108
high-dim LBP	0.9517 ± 0.0113
DFD	0.8402 ± 0.0044
TL Joint Bayesian	0.9633 ± 0.0108
face.com r2011b	0.9130 ± 0.0030
Face++	0.9950 ± 0.0036
DeepFace-ensemble	0.9735 ± 0.0025
ConvNet-RBM	0.9252 ± 0.0038
POOF-gradhist	0.9313 ± 0.0040
POOF-HOG	0.9280 ± 0.0047
FR+FCN	0.9645 ± 0.0025
DeepID	0.9745 ± 0.0026
GaussianFace	0.9852 ± 0.0066
DeepID2	0.9915 ± 0.0013
TCIT	0.9333 ± 0.0124
DeepID2+	0.9947 ± 0.0012
betaface.com	0.9953 ± 0.0009
DeepID3	0.9953 ± 0.0010
insky.so	0.9551 ± 0.0013
Uni-Ubi	0.9900 ± 0.0032
FaceNet	0.9963 ± 0.0009
Baidu	0.9977 ± 0.0006
AuthenMetric	0.9977 ± 0.0009
MMDFR	0.9902 ± 0.0019
CW-DNA-1	0.9950 ± 0.0022
Faceall	0.9967 ± 0.0007
JustMeTalk	0.9887 ± 0.0016
Facevisa	0.9955 ± 0.0014
pose+shape+expression augmentation	0.9807 ± 0.0060
ColorReco	0.9940 ± 0.0022
Asaphus	0.9815 ± 0.0039
Daream	0.9968 ± 0.0009
Dahua-FaceImage	0.9978 ± 0.0007
Skytop Gaia	0.9630 ± 0.0023

CNN-3DMM estimation	0.9235 ± 0.0129
Samtech Facequest	0.9971 ± 0.0018
XYZ Robot	0.9895 ± 0.0020
THU CV-AI Lab	0.9973 ± 0.0008
dlib	0.9938 ± 0.0027
Aureus	0.9920 ± 0.0030
YouTu Lab, Tencent	0.9980 ± 0.0023
Orion Star	0.9965 ± 0.0032
Yuntu WiseSight	0.9943 ± 0.0045
PingAn AI Lab	0.9980 ± 0.0016
Turing123	0.9940 ± 0.0040
Hisign	0.9968 ± 0.0030
VisionLabs V2.0	0.9978 ± 0.0007
Deepmark	0.9923 ± 0.0016
Force Infosystems	0.9973 ± 0.0028
ReadSense	0.9982 ± 0.0007
CM-CV&AR	0.9983 ± 0.0024
sensingtech	0.9970 ± 0.0008
Glasssix	0.9983 ± 0.0018
icarevision	0.9977 ± 0.0030
Easen Electron	0.9983 ± 0.0006
yunshitu	0.9987 ± 0.0012
RemarkFace	0.9972 ± 0.0020
IntelliVision	0.9973 ± 0.0027
senscape	0.9930 ± 0.0053
Meiya Pico	0.9972 ± 0.0008
Faceter.io	0.9978 ± 0.0008
Pegatron	0.9958 ± 0.0013
CHTFace	0.9960 ± 0.0025
FRDC	0.9972 ± 0.0029
YI+AI	0.9983 ± 0.0024
Aratek	0.9972 ± 0.0021
Cylltech	0.9982 ± 0.0023
TerminAI	0.9980 ± 0.0016
ever.ai	0.9985 ± 0.0020
Camvi	0.9987 ± 0.0018
IFLYTEK-CV	0.9980 ± 0.0024

四、 总结与展望

人脸识别是一个具有发展潜质的领域，也是一个具有极大挑战力的领域。就现在的发展阶段而言，人脸识别技术还面临着多方面的困难，如被识别的人需要正脸面对图像采集器，并且被识别的人要保持合适的距离才能采集到比较准确的

数据。在正常应用时这些问题可能就是我们无法识别的难题。目前而言还没有一种能适应不同环境和干扰的识别方法。对于未来而言我们仍然需要提高识别的精准度扩大识别的范围，从而使人脸识别技术可以使用于任何复杂环境。所以未来人脸识别将向着多方面发展如：免干扰人脸特征数据采集、远距离人脸识别技术、3D 细节模型构建等等。

另外，对于人脸活体的检测方向研究偏少，普遍解决活体检测的方案就是以3D 重建实现，对二维图像如何判别活体与照片也是一个研究的难题。

参考文献

[1] T.F. Cootes, C.J. Taylor, D.H. Cooper, et al. Active Shape Models-Their Training and Application[J]. Computer Vision and Image Understanding, 1995, 61(1):38-59.

[2] G. J. Edwards, T. F. Cootes, C. J. Taylor. Face recognition using active appearance models[J]. Computer Vision—Eccv』, 1998, 1407(6):581-595.

[3] Cootes T F, Edwards G J, Taylor C J. Active appearance models[C]// European Conference on Computer Vision. Springer Berlin Heidelberg, 1998:484-498.

[4] Dollár P, Welinder P, Perona P. Cascaded pose regression[J]. IEEE, 2010, 238(6):1078-1085.

[5] Sun Y, Wang X, Tang X. Deep Convolutional Network Cascade for Facial Point Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2013:3476-3483.

[6] Zhou E, Fan H, Cao Z, et al. Extensive Facial Landmark Localization with Coarse-to-Fine Convolutional Network Cascade[C]// IEEE International Conference on Computer Vision Workshops. IEEE, 2014:386-391.

[7] Zhang Z, Luo P, Chen C L, et al. Facial Landmark Detection by Deep Multi-task Learning[C]// European Conference on Computer Vision. 2014:94-108.

[8] Wu Y, Hassner T. Facial Landmark Detection with Tweaked Convolutional Neural Networks[J]. Computer Science, 2015.

[9] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.

[10] Kowalski M, Naruniec J, Trzcinski T. Deep Alignment Network: A Convolutional Neural Network for Robust Face Alignment[J]. 2017:2034-2043.

[11] 蓝振潘.基于深度学习的人脸识别技术及其在智能小区中的应用[D]. 华南理工大学,2017.

[12] Daniel Saez Trigueros, Li Meng. Face Recognition: From Traditional to Deep Learning Methods. Computer Vision and Pattern Recognition.2018

[13] https://github.com/ShiqiYu/libfacedetection

[14] http://vis-www.cs.umass.edu/lfw/results.html

[15] Shifeng Zhang, Rui Zhu. Improved Selective Refinement Network for Face Detection. Computer Vision and Pattern Recognition.2019

[16] 黄建，李文书，高玉娟．人脸表情识别研究进展［Ｊ］．计算机科学，2016,43(S2):123-126

[17] Sun Y,Wang X,Tang X.Deep Learning Face Representation from Predicting 10,000 Classes[C].Computer Vision and Pattern Recognition. IEEE,2014:1891-1898.

[18] Deng J, Guo J, Zafeiriou S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition[J]. arXiv preprint arXiv:1801.07698, 2018.

[19] Qi X, Zhang L. Face Recognition via Centralized Coordinate Learning[J]. arXiv preprint arXiv:1801.05678, 2018

[20] 清华大学计算机系－中国工程科技知识中心.2018 人脸识别研究报告.

[21] Wang F, Liu W, Liu H, et al. Additive Margin Softmax for Face Verification[J]. arXiv preprint arXiv:1801.05599, 2018.

[22] https://github.com/TadasBaltrusaitis/OpenFace

[23] Weiyang Liu, Yandong Wen.SphereFace: Deep Hypersphere Embedding for Face Recognition. CVPR 2017

[24] Florian Schroff, Dmitry Kalenichenko.FaceNet: A Unified Embedding for Face Recognition and Clustering. Computer Vision and Pattern Recognition.2015

[25] Xiong X, Torre F D L. Supervised Descent Method and Its Applications to Face Alignment[C]// Computer Vision and Pattern Recognition. IEEE, 2013:532-539.

[26] 宋嘉程.人脸识别技术的现状和发展，电子技术与软件工程，2017，09.

[27] 肖冰，等.人脸识别综述 [J]. 计算机学报，2016，8

[28] Wang N, Gao X, Tao D, et al. Facial Feature Point Detection: A Comprehensive Survey[J]. Neurocomputing, 2017.