人脸关键点定位概述

15210240008 贺珂珂

本文为人脸领域内关键点定位的概述，将从关键点定位的作用，关键点定位的挑战，关键点定位技术，关键点定位展望这四方面进行展开。其中重点分析目前的关键点定位技术，分为传统的关键点定位和基于深度学习的关键点定位 2 部分。

一、关键点定位作用：

随着生物认证技术的发展，人脸，作为最为自然和普遍的身份特征，吸引了大量的研究。而人脸识别由于具有操作便捷、可交互性强等独特的优势，易于为用户所接受，具有广泛的应用前景。人脸识别技术极大推动了图像处理、模式识别、计算机视觉等诸多学科的发展。人脸特点点定位，不仅是人脸识别研究领域中一个关键问题，也是计算机视觉和图形学领域中一个基本问题，其目标是在人脸图像中定位出人脸形态特征，包括眼、嘴、鼻、人脸外轮廓等位置和形状。面部关键特征点的定位在人脸识别，人脸追踪，人脸属性分析，3D 人脸建模方面都起着重要的作用，其精确性直接关系着后续应用的可靠性。图 1-1 为一种特征点标注的示意图，此种格式共需要标注 29 个特征点。

图 1-1 一种特征点标注示意图

二、关键点定位的挑战：

1. 人脸姿态。由于人脸会有不同的朝向、夸张的表情，这些情况会使特征点信息的缺失或偏差，从而加大了特征点定位的难度；
2. 光照因素。如图 1-2 所示，同一张人脸不在不同光照下，视觉效果的变化是很大的。因此不同方向的光照、光照的强度等都会使特征点定位点的问题变得复杂；

图 2-1 同一个人脸在不同光照下的视觉效果

3. 人脸遮挡。人脸遮挡会直接损失人脸一部分的信息，显然会大幅度增加特征点定位的难度；

4. .关键点的数量。目前的定位点有 5 个点，68 个点，83 个点等等。随着定位点的增加，定位的难度也有一定程度的增加；

5. 关键点精确性。关键点定位，相当于在已有粗略人脸定位的基础上，再进行细致的关键区域定位，对精确性也有较高的要求。

三、 传统的关键点定位技术：

最初，人脸关键点定位主要是一个迭代优化的过程，对于测试图片，先随机给个点的分布（初始模型），然后，根据当前关键点附近的纹理特征和结构特征信息，进行一步步的迭代优化。同时为了保证关键点的整体分布的合理性，会对所有的关键点加上一个全局的约束。为了便于区分，我们将利用关键点附近的纹理信息和结构信息进行迭代优化的方法，称为传统方法。ASM 是传统方法中的一个主流算法，其最初由 T.F.Coots 等人于 1992 年提出，是一种基于统计模型的图像搜索方法[1][2]。它是建立在点分布模型的基础上，通过训练样本图像获取训练样本的特征点分布的统计信息，并且获取特征点允许存在的变化方向，从而实现在目标图像上寻找对应特征点的位置。

ASM 分为训练和搜索两部分，流程图如图 3-1[3]所示。

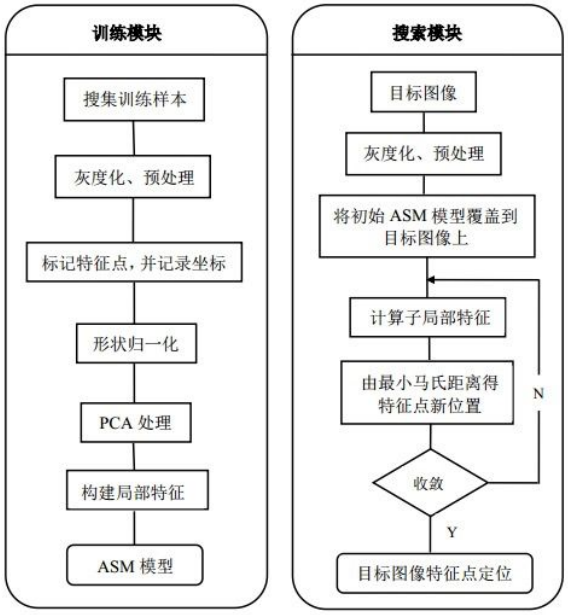


图 3-1 ASM 训练和搜索模块流程图

3.1　构建形状向量

在 ASM 训练中，我们基于点分布模型，对训练集中任意一幅人脸图像，标记 68 个关键特征点，如图 3-2 [4]所示，并记录每个关键特征点的位置坐标信息。

图 3-2　人脸图像 68 个特征点定位

这样每一幅人脸图像都可以表示为一个形状向量，该向量元素由人脸图像中标记的 68 个关键特征点坐标交替组成。

3..2　归一化处理

归一化处理就是把一系列的点分布模型通过适当的平移、旋转、缩放变换，在不改变点分布模型的基础上对齐到同一个点分布模型，即

$$M(s,\theta)\begin{bmatrix}x_{ik}\\y_{ik}\end{bmatrix}=\begin{bmatrix}s\cos\theta & s\sin\theta\\-s\sin\theta & s\cos\theta\end{bmatrix}\begin{bmatrix}x_{ik}\\y_{ik}\end{bmatrix}=\begin{bmatrix}(s\cos\theta)x_{ik}+(s\sin\theta)y_{ik}\\(s\cos\theta)y_{ik}-(s\sin\theta)x_{ik}\end{bmatrix}$$

3.3　PCA 处理

人脸训练样本集中的形状向量作归一化处理后，就可以利用 PCA 方法对对齐后的形状向量进行分析和降维，找出包含形状模型可以存在的变化方向的统计信息，得到统计形状模型。统计形状模型的参数反映了特点点形状的可变化模式。

3.4　构建灰度模型

我们通过对每个特征点附近的像素灰度信息进行求导、归一化、协方差矩阵等计算（如图 3-3），得到每个特征点的灰度特征。在 ASM 搜索过程中，我们通过计算特征点灰度特征之间的最小马氏距离来为特征点寻找新的位置。

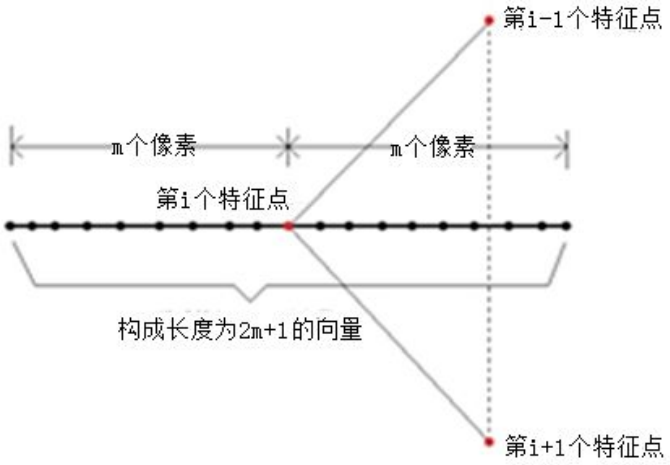


图 3-3 灰度模型示意图

3.5 ASM 搜索

ASM 搜索过程即将初始模型覆盖到目标图像上，然后通过计算特征点灰度特征之间的最小马氏距离来为特征点寻找新的位置，并利用统计形状模型对寻找到的新位置进行约束，以此达到迭代搜索，不断优化参数，最终模型匹配到新的目标图像上，即特征点定位（如图 3-4 [5]所示）。

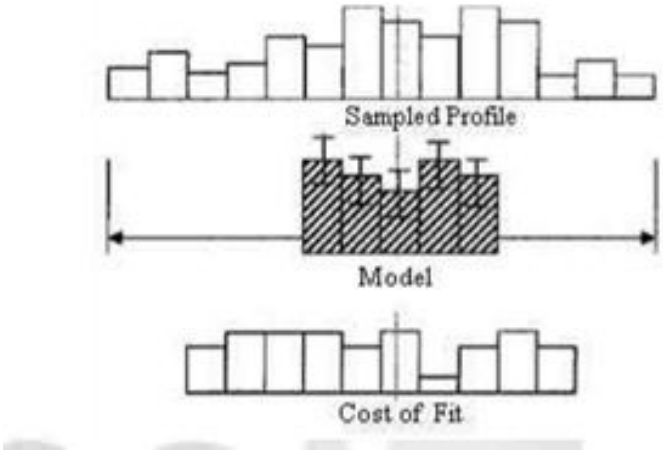


图 3-4 灰度模型匹配示意图

3.1.6 算法的优劣

优：ASM 通过训练集获得的先验知识，从而达到自上向下的人脸特征定位的方法。对于正面人脸关键点定位，由于纹理特征和结构特征信息充分明显，因此具有很高的精确性。

劣：在 ASM 搜索过程中，仅使用了特征点局部纹理特征在作为搜索信息，没有使用全局的纹理约束，因此在实际应用中，ASM 很容易陷入局部最小的缺陷。同时 ASM 只有在正面人脸特征点定位上具有较高的精确性，但对于侧面人脸，由于缺失部分特征纹理信息，定位效果很差。

四、基于深度卷积网络的关键点定位技术：

在此之前，人脸关键点定位主要是一个迭代优化的过程，对于测试图片，先随机给个点的分布，然后，根据当前关键点附近的纹理特征和结构特征，进行一步步的优化。同时为了保证关键点的整体分布是合理的，会对所有的关键点加上一个全局的约束。这种方法的主要缺点就是太依赖于初始值，容易陷入局部最小值，同时，因为是一个不断迭代的过程，速度较慢。

之后，一个有突破性进展的工作，来源于 Cao [6]。不同于之前的迭代更新模型，Cao 通过训练级联的回归树，对输入图片，能够直接回归出关键点。这个技术不仅定位精确性高，而且速度非常快。之后的工作[7]，通过训练一组局部二进制特征的回归树，直接得到关键点位置，在作者的实现中，可以达到 3000FPS。速度和精度都非常快。上面这些技术所采用的特征都比较简单。目前， 深度学习可以从大量的数据中自动学习到特征的表示，因其强大的特征表示能力，在图像分类，识别，检测等各个领域都取得了突破性的进展[8]，下面将主要对深度学习在人脸关键点定位方面的技术进行详细的展开。

现有的利用深度卷积网络进行定位的方法主要可以分为 3 类：

第一类框架基于级联回归的方法，即先得到较粗的定位结果，之后一步步优化得到精准的定位结果。

第二类框架是基于多任务辅助定位的方法。利用人脸的其他先验信息，辅助特征点定位。

最后一类是基于语义分割的方法。

4.1 由粗到细（Corse to fine）的级联回归的定位框架

4.1.1 Deep Convolutional Network Cascade for facial point detection (2013) [9]

4.1.1.1 效果图

4.1.1.2 主要思想

这篇文章可以说是利用深度学习中的 CNN(Convolutional Neural Network) 模型进行人脸关键点定位的开山之作。之前的 CNN 主要是用来进行分类，网络的监督信息为 0~N 的类别。在这里，作者直接将关键点的坐标作为网络的监督信息，让网络根据训练图像，学习到一组非线性映射，直接预测坐标。

4.1.1.3 具体方法

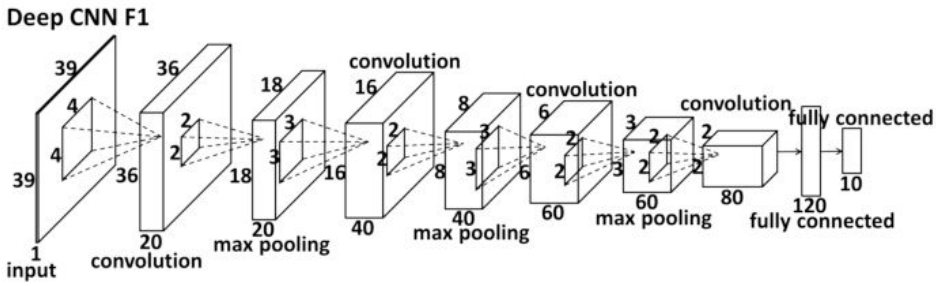


图 4-1　5 个点定位的深度卷积网络结构

图 4-1 是一个网络的示意图，输入 39*39 大小的人脸灰度图，直接通过网络预测 10 维数据。即等于 5 个点的坐标。

之后，为了获得更高的定位准确性，作者提出了一个 3 阶段级联深度网络的模型。

在第一个阶段，深度 网络在整个人脸区域上抽取高层特征，直接预测所有的关键点。

这样有 2 方面优点 ：

1）在整个人脸区域上抽取纹理特征，利用了整体信息。

2）隐含了所有点直接的几何约束。

这样利用整体特征的方法能够防止陷入局部最小值。后 2 个阶段，在前一阶段预测的点附近区域利用局部特征进行预测，寻求更合适的点，通过这后 2 阶段的级联修正，最初始预测的点将被进一步优化，最终获得更高的精确度。 图 4-2 为级联定位的网络框架。

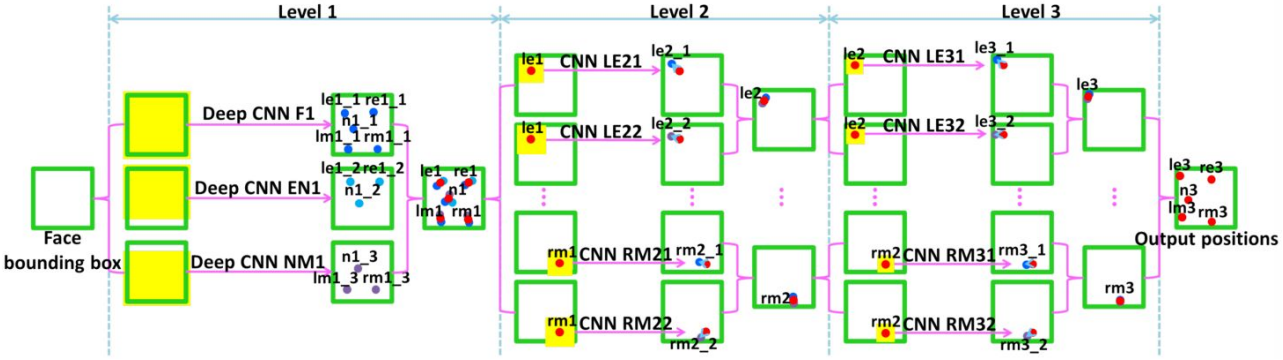


图 4-2　级联定位的网络框架

4.1.1.4 数据集

训练集：作者标注了 5 个人脸关键点数据，共 13466 张图片

测试集：共 2555 张图片

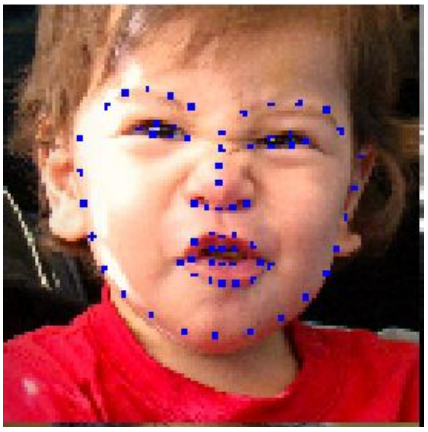
4.1.1.5 方法评估

方法的优点：利用 CNN 进行特征表示，然后进行级联，定位的点有很高精确度。

方法的缺点：当特征点数量大的时候不适用。而且网络数目较多。

4.1.2 Extensive Facial Landmark Localization with Coarse-to-fine Convolutional Network Cascade (2013) [10]

4.1.2.1 效果图



4.1.2.2 主要思想

这篇文章跟第一篇的思想是类似的，利用级联的 4 阶段网络，实现粗到细的关键点定位。

跟第一篇文章不同的是，上面一篇文章定位的点比较有限，（5 个点），而且，所需的网络个数过多（23 个）

这篇文章，共训练了 14 个不同的网络，能够实现 68 个点的定位。

4.1.2.3 具体方法

本文中一个主要的特点将 68 个关键点分而治之。

这样有 2 方面的优点：

1）定位点的过程本身有独立性的。如果为了定位眼睛的点，那么人脸的下半部分信息是不需要的。

2）专注于定位某一部分的店。作者发现轮廓的点的误差远远大于轮廓内部的点的误差。那么在训练过程中，内部点的误差将会主要被轮廓点主导。

如果将轮廓点和内部点分开，2 个方面都能更好地学习相关的细节信息。

网络的测试共分为 4 个阶段，

第一阶段：利用人脸检测后得到的人脸，分别通过 2 个网络，学习到内部的点的框，和轮廓点的框的坐标。

第二阶段：利用第一阶段的框裁剪出来的图像，分别通过 2 个网络，预测到内部的点和轮廓点。

第三阶段：前 2 个阶段，都利用了整体的信息，此阶段主要进行部件级的优化。作者把人脸细分为 6 个部分，左眉毛，右眉毛，左眼睛，右眼睛，鼻子，嘴巴。

利用第三阶段的 6 个网络，对部件进行了更细的修正。

第四阶段：利用第三阶段的检测结果，将人脸的部件摆正，然后利用摆正的图片，通过 6 个网络进行更细致的修正。

最后，将检测到的点合起来，得到了最终的定位结果。 图 4-3　为 68 个点的级联定位的网络框架。

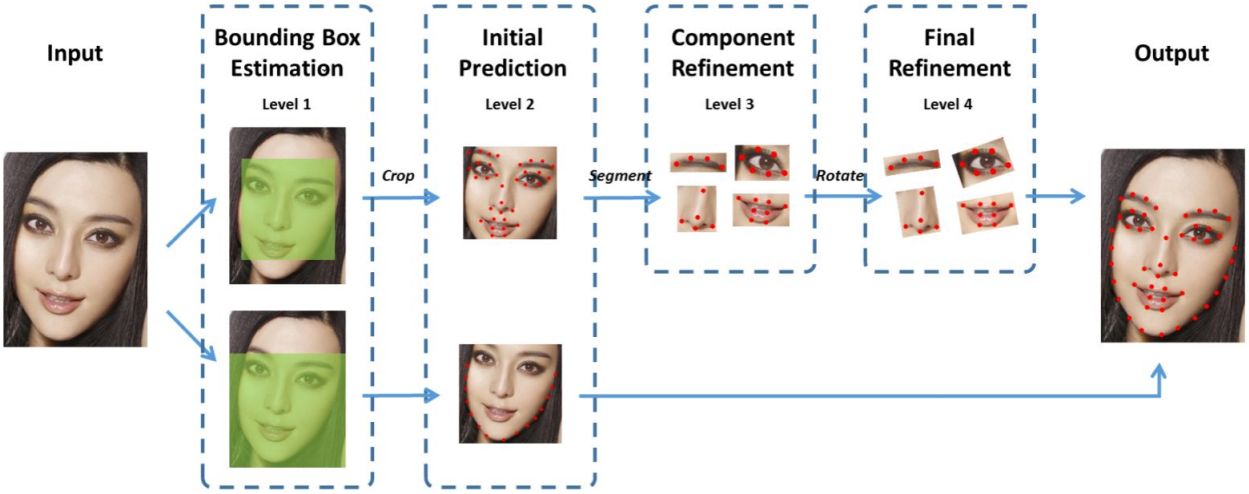


图 4-3　68 个点的级联定位的网络框架

4.1.2.4 数据集

训练集：AFW, LFPW-TRAIN, HELEN-TRAIN [11]共 3148 张图片

测试集：LFPW-TEST, HELEN-TEST 和 IBUG 共 689 张图片

4.1.2.5 方法评估

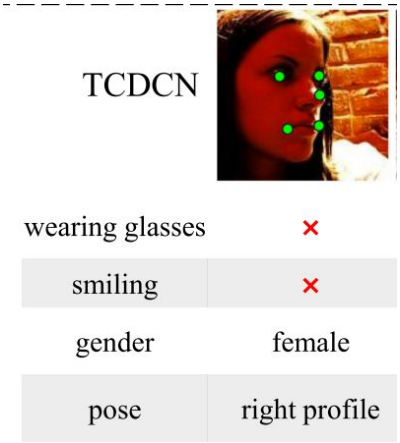
方法的优点：能够定位数量大的人脸特征点

方法的缺点：网络的数目较多

4.2 基于多任务辅助的定位框架

4.2.1 Facial Landmark Detection by Deep Multi-task Learning (2014) [12]

4.2.1.1 效果图



4.2.1.2 主要思想

这篇文章的主要贡献就是，作者发现头部姿势和遮挡等属性对关键点定位有较大的影响，之后作者通将关键点与影响关键点定位的其他任务联合进行优化，提升特征点定位的精度。通过多任务学习，利用单网络学习到了鲁棒的特征，能够超过之前 23 个级联网络的效果。

4.2.1.3 具体方法

在这篇论文中，作者选择了头部姿势，性别，是否带眼睛，是否微笑这 4 个对关键点定位有影响的任务联合进行训练。

多任务之间会相互有影响，如何借助多任务信息对关键点位置的约束，提高关键点的定位精确度是这篇论文重点讨论的。

如图 4-4，为多任务约束的网络的结构图。

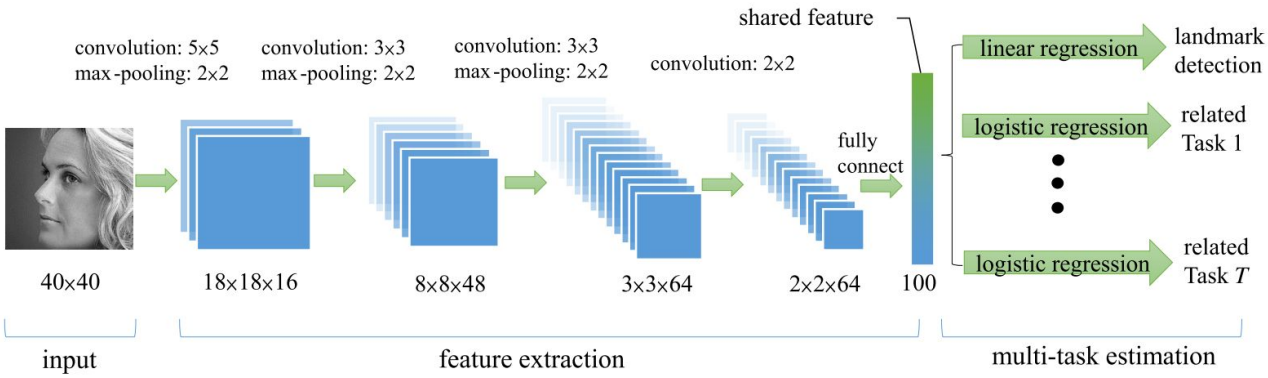


图 4-4，多任务约束的网络的网络框架

整个网络分为特征提取和多任务优化 2 个部分。

特征提取部分：

网络的输入为 40*40 的灰度图，经过 4 次交替的卷积，池化，最后，全连接到 100 维的向量。这 100 维为最后提取到的特征。

多任务优化部分：

在这里，主任务为关键点定位任务，是一个回归问题，其他的辅助任务，为分类问题。所有的任务，共同利用这 100 维特征来进行回归任务和分类任务的预测。

在网络的反向传播过程中，所有任务的误差共同回传，进行联合的优化。

考虑到不同辅助任务的复杂性是不同的，所有不同任务的收敛时间也是不同的，作者提出了一个 early stop 的策略，通过比较一个时间段内训练误差和验证误差的比例，如果超过阈值，就设置辅助任务的权重为 0，不再参与网络的优化。

通过多任务的学习，对关键点进行了更好的优化，作者的实验显示，准确度要比之前 23 个级联网络好。

4.2.1.4 数据集

训练集：在第一篇论文[9]的数据集基础上，加上了 4 个人脸属性

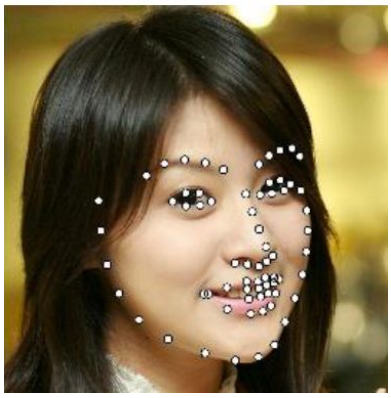
测试集：同上

4.2.1.5 方法评估

方法的优点：定位准确性高，模型的复杂度低

4.2.3 Learning Deep Representation for Face Alignment with Auxiliary Attributes（2015) [13]

4.2.3.1 效果图



4.2.3.2 主要思想

这篇文章也是一个多任务负责关键点定位的思想。是对第三篇工作的进一步扩展。

在这里，作者将辅助任务的数量从 4 个扩展到了 22 个。

同时，对辅助任务之间的关联度进行了分析，通过一个时间段内的训练误差和测试误差的比例来控制不同任务的权重，联合进行优化。

4.2.3.3 具体方法

作者先通过 5 个关键点+ 22 个辅助任务的监督数据用来学习网络。

之后进行 transfer learning ，即将 5 个点训练得到的权重进行初始化，利用 68 个关键点的数据作为进行信息，进行 fine-tuning。

通过利用 5 个点 + 22 个辅助任务的数据集，网络已经学习到了很多人脸的信息，当面对定位较为复杂的 68 个点的任务，能过更好地进行处理，效果远远好过直接利用 68 个点从头开始学习。图 4-4 为训练 5 个点的多任务网络，再 transfer learning 学习到 68 个点的网络的过程。

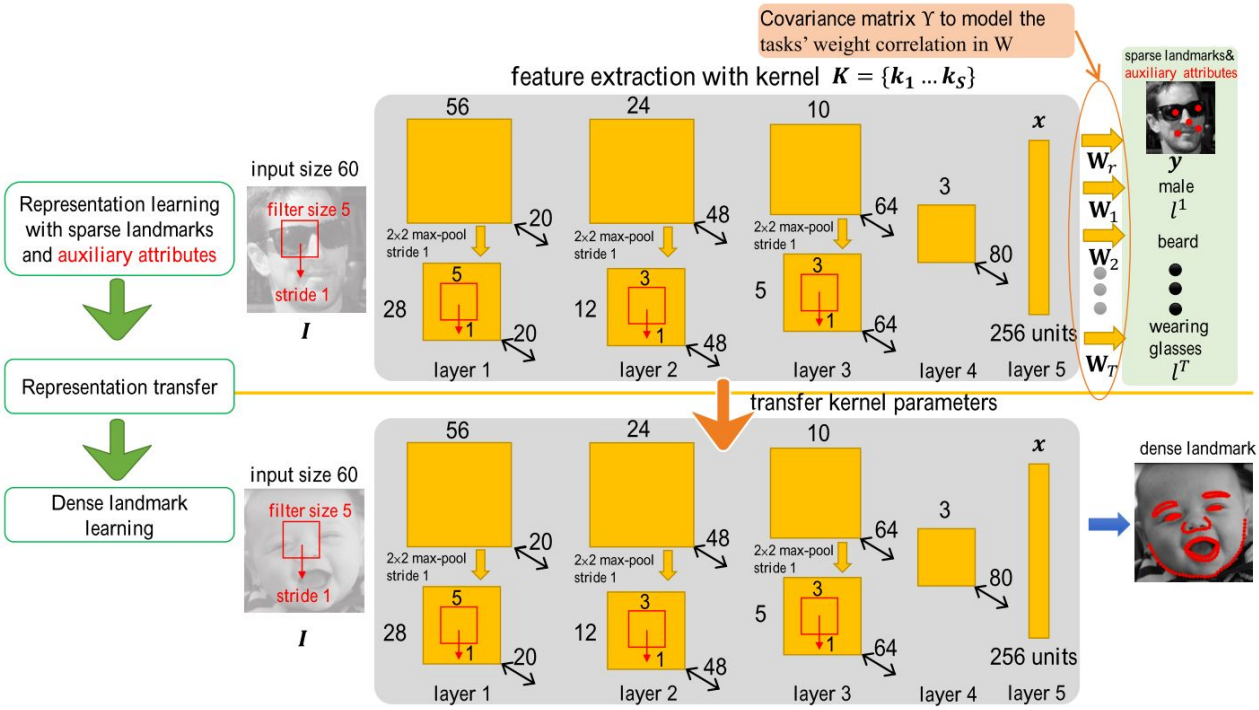


图 4-4　transfer-learning 的网络框架

4.2.3.4 数据集

训练集：20000 张 5 个点+22 个属性的图片和 AFW, LFPW-TRAIN, HELEN-TRAIN 共 3148 张图片

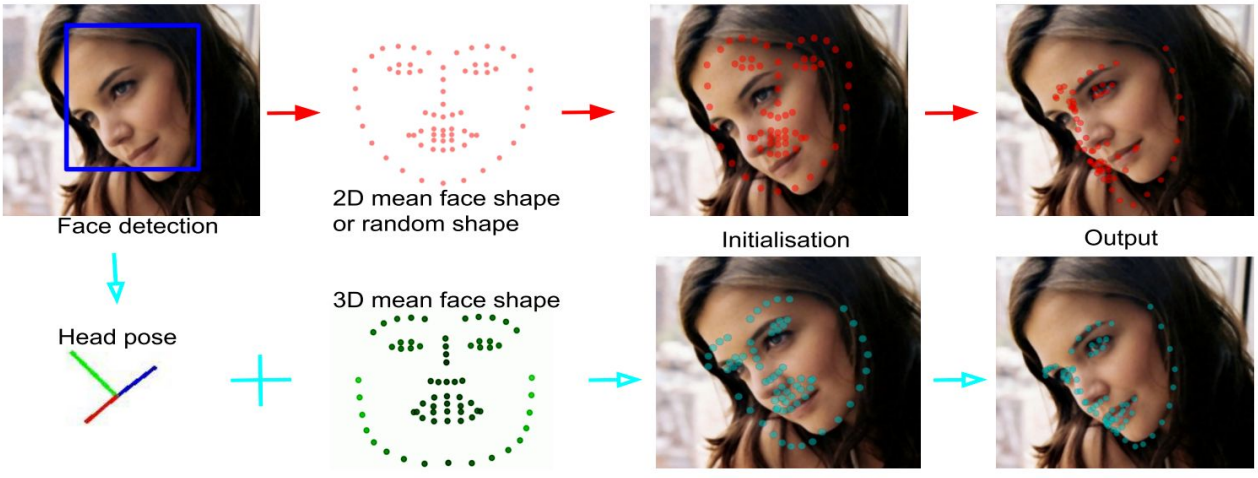
测试集：LFPW-TEST, HELEN-TEST 和 IBUG 共 689 张图片

4.2.3.5 方法评估

方法的优点：通过 transfer learning ，68 个点的精确度非常高。

4.2.3 Face Alignment Assisted by Head Pose Estimation （2015）[14]

4.2.3.1 效果图：



对于传统的回归算法，使用平均人脸关键点作为初始，和先计算头部姿势，用与样本库中头部姿势相近的人脸关键点作为初始化的效果差别。

4.2.3.2 主要思想

作者发现，现有的关键点定位技术，在测试集上误差较大的样本，往往头部姿势有较大的变化，作者希望能够

借助先验的头部姿势估计，提高定位的准确度。

4.2.3.3 具体方法

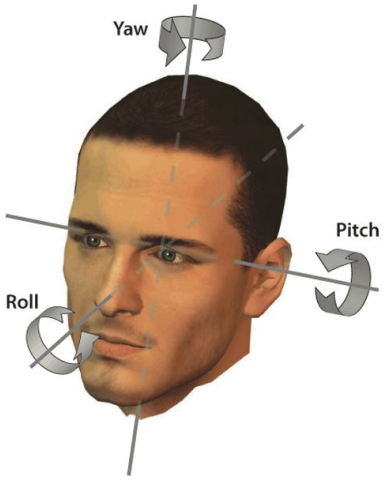


图 4-5　3D 头部姿势示意图

图 4-5 为 3D 头部姿势示意图

主要就是先用 CNN 训练面部姿势，在用这个面部姿势作为初始化，在算法如：Robust Cascaded Pose Regression 上作为初始化，取得更好的效果。

训练过程:

1. 先用 68 个点的数据计算出 头部的 3 维姿势，作为标签（label)
2. 训练 3 维头部姿势预测网络。图 4-5 为头部姿势的网络结构。

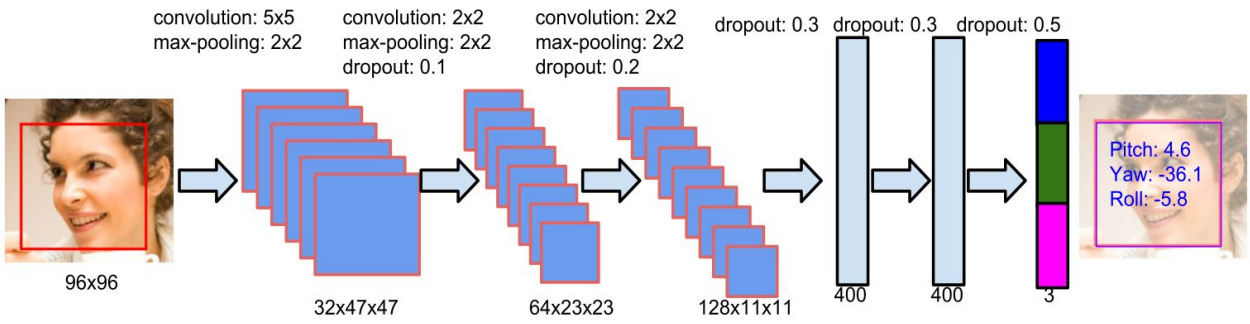


图 4-5　头部姿势的网络结构

测试过程 ：

1. 对于测试图片，先通过训练好的网络预测头部姿势
2. 在训练集中找相近头部姿势的样本
3. 用相近的头部姿势的样本的 68 个点来作为算法的关键点初始化
4. 迭代地优化关键点

4.2.3.4 数据集

训练集：AFW, LFPW-TRAIN, HELEN-TRAIN 共 3148 张图片

测试集：LFPW-TEST, HELEN-TEST 和 IBUG 共 689 张图片

4.2.3.5 方法评估

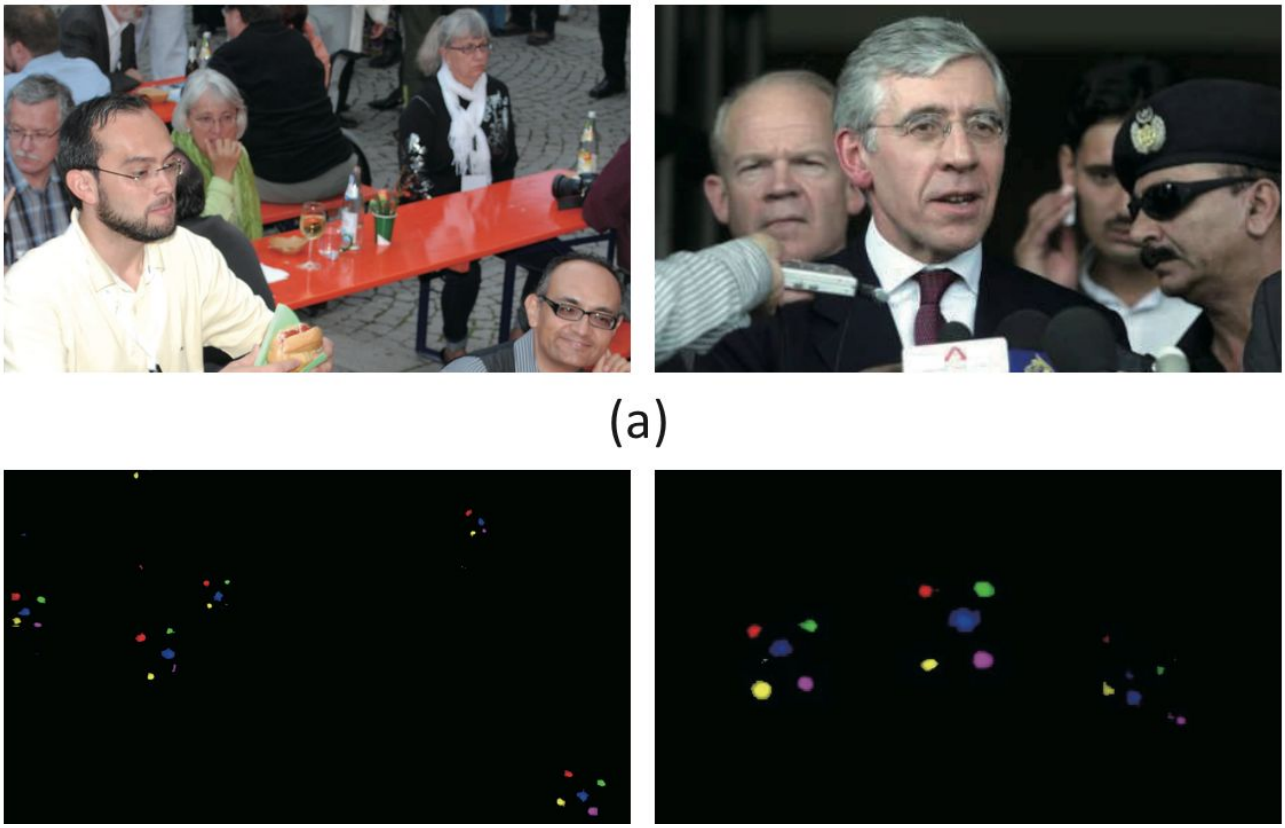
方法的优点：一种新的有效的初始化策略。

方法的缺点：头部姿势的估计是用网络估计，之后的迭代又是利用传统的方法进行优化，不够简便。

4.3 基于语义分割的方法

4.3.1 Unconstrained Facial Landmark Localization with Backbone-Branches Fully-Convolutional Networks(2015) [15]

4.3.1.1 效果图：



(a)

4.3.1.2 主要思想

对于原始的输入图片，直接产生关键点的响应图，不需要额外的人脸检测，截取过程。能够有效且精确地定位关键点。

4.3.1.3 具体方法

主要的思想是借助于全卷积网络做语义分割的思想，将点特征点的响应图作为监督信息。越亮代表这个点是特征点的可能性越大。

网络也是一个整体与部分的设计考虑。

先是一个主干网络，直接得到 5 个响应图，分别对应 5 个点。

下面是 5 个分支网络，每个网络产生 1 个响应图，分别对应 1 个点。

最后将主干网络和分支网络组合起来，得到整体的响应图。图 4-6 为主干与分支的网络结构。

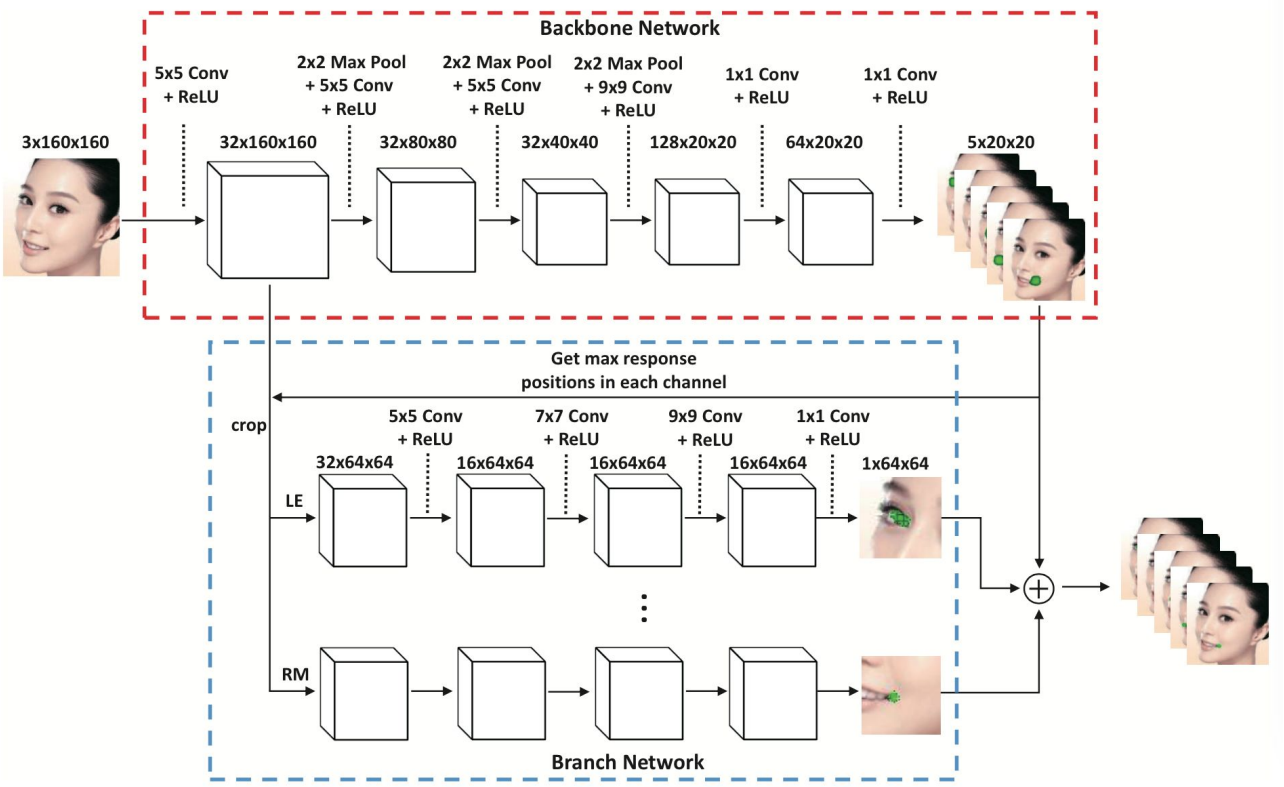


图 4-6　主干与分支的网络结构

4.3.1.4 数据集

训练集：作者标注了 15533 张 5 个点的训练数据

测试集：标注了 2650 张 5 个点的测试数据

4.3.1.5 方法评估

优先：不需要先进行人脸检测等其他预处理过程，方法很新颖，直接。

缺点：模型比较复杂，不适合检测大量点（68 个点）。

五、关键点定位展望

人脸识别中面部的关键特征点定位是一个具有相当挑战性的问题，也是计算机视觉和图形学领域的一个基本问题。人脸识别技术具有广阔的市场需求及应用前景，广泛应用于公安、安全、海关、金融、军队、机场、边防口岸、安防等多个重要行业及领域。但是目前商业系统的实用性依然不是很完善，技术需求跟不上应用需求。阻碍人脸识别技术的因素很多，主要包括：姿态问题、表情问题、光照问题等。关键点的精确定位是解决这些问题的基础。现有的许多关键点定位方法的精确度都会随光照、姿态等外界条件变化而下降，因此研究一个具有高精确性和高鲁棒性的特征点定位方法显得特别有必要。

根据现有的定位技术，关键点定位在以下这 3 个方法还有待深入研究。

1.3D 人脸关键点

目前的定点算法一般都是基于 2D 图片的，如何借助于深度信息，提升定位的精确度，同时辅助 3D 人脸建模是值得探索的。

2.头部姿势

现有人脸关键点检测的算法的误差主要来自于头部姿势变动较大的图片，这些图片一方面跟普通的图片外观差异较大，另一方面，训练数据中这类数据本身就比较少。关键点定位算法对有较大姿势的图片依旧鲁棒是一个亟待解决的问题。

3.算法性能

在这里介绍了比较多的基于深度卷积网络的定位技术，有的算法有相当多的模型，大大增加了计算代价，而传统方法一般借助于简单特征，有较好的实时性。如何在不增加模型复杂度的情况下，提升定位的精度也是值得探索的。

参考文献

[1] T.F.Coots and C.J.Talor. Active shape model search using local grey-level models: A quantitative evaluation,in Proceeding,British Machine Vision Conference,1993.

[2] T.F.Coots and C.J.Talor. Technical Report:Statistical Models of Appearance for Computer Vision[D]. The University of Manchester School of Medicine,2004.

[3] 李洪升.基于 ASM 算法的人脸特征点定位研究及应用,东南大学硕士学位论文，2009.10。

[4] “ASM（Active Shaped Model）算法介绍” [Online]Available: http://blog.csdn.net/carson2005/article/details/8194317。

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

[6] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In Proc. CVPR, 2012.

[7] Ren S, Cao X, Wei Y, et al. Face alignment at 3000 fps via regressing local binary features[C]//Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 1685-1692.

[8] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[9] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[C]. Computer Vision and Pattern Recognition (CVPR), US, 2013.

[10] Zhou E, Fan H, Cao Z, et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]//Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. IEEE, 2013: 386-391.

[11] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 faces in-the-wild challenge: The first facial landmark localization challenge[C]//Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. IEEE, 2013: 397-403.

[12] Zhang Z, Luo P, Loy C C, et al. Facial landmark detection by deep multi-task learning[M]//Computer Vision–ECCV 2014. Springer International Publishing, 2014: 94-108.

[13] Zhang Z, Luo P, Loy C C, et al. Learning and transferring multi-task deep representation for face alignment[J]. arXiv preprint arXiv:1408.3967, 2014.

[14] Yang H, Mou W, Zhang Y, et al. Face Alignment Assisted by Head Pose Estimation[J]. arXiv preprint arXiv:1507.03148, 2015.

[15] Zhujin L, Shengyong D, Liang L, et al. Unconstrained Facial Landmark Localization with Backbone-Branches Fully-Convolutional Networks[J]. arXiv preprint arXiv:1507.03409 , 2015.