**外 文 译 文**

**基于深度级联卷积网络的脸部特征点检测**

摘 要

我们提出了一种通过三级卷积神经网络估计脸部关键点位置的新方法。在每一级，多重网络的输出都是鲁棒且准确估计的。由于卷积网络的深层结构，初始化阶段，在整个面部区域提取全局高级特征，这有助于定位高精度关键点。这种方法有两个优点：首先，整个面部的纹理背景信息都被用来定位每个关键点。其次，由于网络被训练来同时预测所有关键点，关键点之间的几何约束被隐式编码。因此，该方法可以避免由于遮挡、大的位姿变化和极端光照而造成的图像样本模糊和数据损坏造成的局部最小值。下面两级的网络，被训练用于局部改进初始预测值，并且它们的输入仅限于初始预测周围的小区域。我们研究了几种对于脸部特征点检测精度和鲁棒性至关重要的网络结构。大量实验表明，我们的方法在检测准确性和可靠性方面均优于现有最先进的方法。

# 1 简介

人脸关键点检测对于人脸识别和分析至关重要，并且近年来已被广泛研究[3,4,5,5,9,11,20,21,23,25,26,27,28]。 当采用极端姿势、光照、表情和遮挡拍摄人脸图像时，这个问题具有挑战性，如图1所示。现有方法通常可分为两类：分类搜索窗口[3,4,11,20,28] ，或直接预测关键点位置（或形状参数）[5,8,9,21,25,26]。对于第一类，针对每个关键点，训练了一种称为分量检测器的分类器，它基于局部区域做出决定。由于局部特征可能是模糊的或损坏的，因此可能会发现多个候选区域都像人脸点或没有合适的候选区域。在这种情况下，通过形状约束来估计人脸部点的最佳配置[3,4,11,20,23,28]。 与分量检测器相比，直接预测关键点位置（或形状参数）更有效，因为它不需要扫描。 基于人脸特征点附近的局部斑块[9,26]或整个图像区域[5,25]，常使用回归因子作为预测因子。空间约束也可以添加到回归项中[25,26]。

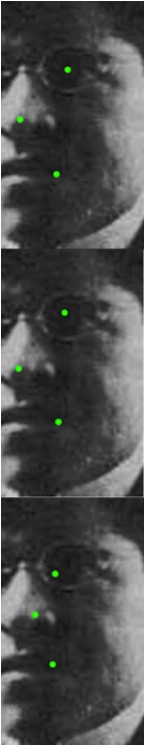


图1:人脸特征点检测示例。第一行:第一级卷积网络的初始检测。即使某些面部成分是看不到或模糊的，也能很好地利用全局背景信息进行估计。第二行:我们的第二和第三级网络的优化结果。结果的计算精度得到了提高。第三行:来自[5]的结果。从训练集中学习到的形状模板对它的限制更大，在一些不寻常的姿势和表情下不准确。

许多方法[5,8,11,20,21,23,25,26]反复地更新脸部特征点的位置，并且好的初始化是十分关键的。从训练集采样的平均形状或形状通常用作初始化，它可能远离目标位置，更新也可能以局部最小值结束。另外，许多方法都面临的问题是，所提取的视觉特征不具有辨别力或不足以预测脸部特征点，因此背景信息变得十分重要。大多数方法采用相对较弱的形状约束。因为纹理背景信息包含丰富的内容，我们希望在整个脸部区域上直接提取它们。由于视觉复杂性随着图像区域的大小呈指数增加，这需要更强大的分类器或回归器。

为了解决这些问题，我们提出了一种通过三级卷积神经网络估计脸部关键点位置的级联回归方法。不同于已有的粗略估计人脸特征点初始位置的方法，我们的卷积网络即使在非常具有挑战性的情况下，也可以在第一级进行精确的预测，如图1所示。它有效地避免了其他方法面临的局部最小问题。卷积网络以整个脸部作为输入，充分利用纹理背景信息，并在深层结构的较高层提取全局高级特征，在深层结构的较高层次提取全局高层次特征，即使在具有挑战性的图像示例中局部区域的低层次特征模糊或损坏，也可以有效地预测关键点。我们的卷积网络被训练成同时预测所有的关键点，并且对关键点的约束进行隐式编码。

其余两级卷积网络细化了关键点的初始估计。与现有的在不同级联阶段应用相同回归量的方法[5,25,26]不同，我们设计了不同的卷积网络。这两级网络结构的层次较浅，因为它们的任务级别低，并且它们的输入仅限于初始位置周围的小区域。在每一级，融合多个卷积网络，可以提高估计的准确性和可靠性。通过详细的实证调查，我们发现，在人脸特征点检测中，与网络结构相关的几个因素是实现良好性能的关键。详细的实验评估表明，我们的方法在精度和可靠性方面均优于目前最先进的方法。

# 2 相关工作

近年来，人脸关键点检测方面已经取得了重大进展。许多人使用Adaboost [20]，SVM [4,28]，或随机森林[3]分类器作为分量检测器，并且基本局部图像特征进行检测。形状约束对于改进元件检测结果的精细化具有重要的意义，这方面的研究一直备受关注。局部分量检测器给出的证据和形状约束可以通过优化设计的目标函数来平衡[4,28]。Liang等人[20]训练了一组方向分类器来指导寻找良好的形状。Amberg和Vetter [3]采用分支定界算法从组件检测器提出的大量候选中有效地找到最佳配置。

在基于回归的方法中，Dantone等人[9]和Valstar等人[26]分别用随机森林和支持向量回归器从局部斑块预测人脸特征点。为了解决预测中的不确定性，Valstar等人用马尔可夫随机场建立了人脸点的空间关系模型，Dantone等人[9]融合了来自人脸区域内密集采样的斑块的许多预测。Patrick等人[25]使用回归器更新了一个活动外观模型的参数。Cao等人[5]采用整面区域作为输入，随机蕨类植物作为回归因子。预测的形状表示为训练形状的线性组合。

卷积网络和其他深度模型已成功用于视觉任务，如人脸检测和姿态估计[24]、人脸解析[22]、图像分类[6,17]和场景解析[10]。卷积网络的研究主要集中在两个方面：网络结构和特征学习算法。Coates等人[7] 分析了具有不同滤波器步长、滤波器大小和特征图数量的单层网络的性能。Jarrett等人[14]引入卷积后的强非线性，包括绝对值校正和局部对比归一化，并比较了不同非线性组合和池策略。直到最近，当卷及网络变得很大（每层有数百个地图）和深层（最多有五个卷积阶段）时，它的潜力才真正被发现。通过使用大规模卷积网络，Ciresan等人[6]显著提高了一些标准分类数据集的技术水平，[17]引入了更大的卷积网络，极大地提高了图像分类精度。最近提出的特征学习算法包括卷积稀疏编码[16]和地形独立分量分析[18]。

# 3 级联卷积网络

在本文中，我们关注单个网络的结构设计及其组合策略。图2是我们方法的概述。有五个人脸特征点要检测：左眼中心（LE）、右眼中心（RE）、鼻尖（N）、左嘴角（LM）和右嘴角（RM）。我们级联三级卷积网络以进行粗到细的预测。在第一级，我们采用三个深度卷积网络，F1、EN1和NM1，其输入区域覆盖整个面部（F1）、眼睛和鼻子（EN1）、鼻子和嘴巴（NM1）。每个网络同时预测多个人脸特征点。对于每个人脸特征点，对多个网络的预测进行平均，以减小方差。图3显示了F1的深层结构，其中包含四个卷积层，后面是最大池，以及两个完全连接的层。 EN1和NM1采用相同的深层结构，但每层的尺寸不同，因为它们的输入区域的大小不同。第二级和第三级的网络采用以先前级别的面部点的预测位置为中心的局部补丁作为输入，并且仅允许对先前的预测进行小的改变。补丁和搜索范围的大小随着级联而减少。最后两个级别的预测受到严格限制，因为局部外观有时是模糊且不可靠的。最后两级上每个点的预测位置由两个补丁大小不同的网络的平均值给出。第一级网络的目标是鲁棒性地估计关键点位置，误差较小，而最后两级网络的目标是达到较高的精度。最后两级的所有网络都具有一个共同的较浅的结构，因为它们的任务都是低级的。

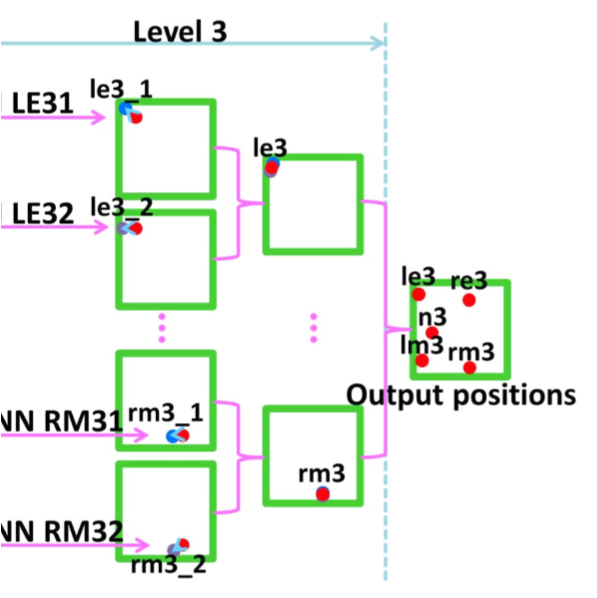
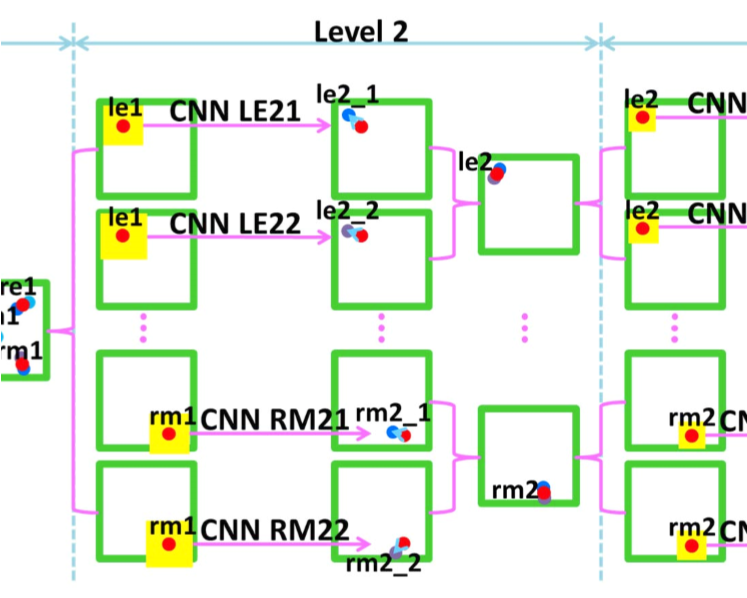
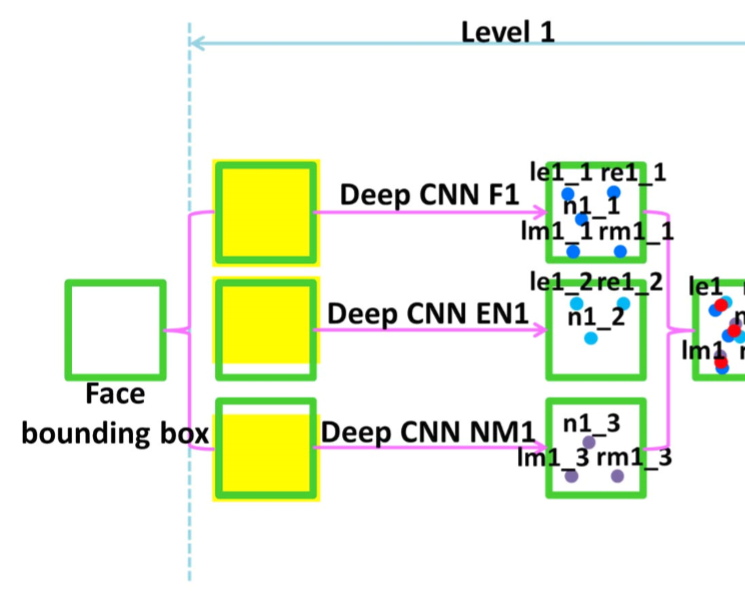


图2:三级级联卷积网络。输入是由人脸检测器返回的人脸区域。第1级的三个网络分别表示为F1、EN1和NM1。第2级的网络表示为LE21、LE22、RE21、RE22、N21、N22、LM21、LM22、RM21和RM22。LE21和LE22都能预测左眼中心，以此类推。第3级的网络表示为LE31、LE32、RE31、RE32、N31、N32、LM31、LM32、RM31和RM32。绿色方块是人脸检测器给出的人脸包围框。黄色阴影区域是网络的输入区域。红点是每一层的最终预测。其他颜色的点是由单个网络给出的预测。

### 3.1 网络结构的选择

我们分析了网络结构选择的三个重要因素。讨论仅限于最难训练的第一级网络。 首先，第一级的卷积网络应该很深。从大输入区域预测关键点是一项高级任务。更深的结构有助于形成高级特征，这些特征是全局的，而较低层次的神经元由于接受域局限，提取的特征是局部的。通过结合低层提取的空间邻近特征，高层神经元可以从较大区域提取特征。此外，高级特征是高度非线性的。增加额外的层增加了从输入到输出的非线性，使得表示输入和输出之间的关系成为可能。

其次，对于卷积层中的神经元，双曲正切激活函数后的绝对值校正（见第4节中的细节）可以有效地改善性能。在[14]中提出了对传统卷积网络的这种修改，其中观察到了对Caltech-101的改进。我们的实证研究表明它在我们的应用中也是有效的。

第三，在同一地图上本地共享神经元的权重可以改善性能。传统的卷积网络基于两个考虑因素共享同一地图上所有神经元的权重。首先，它假设图像中的任何地方都可能出现相同的特征。因此，在一个地方有用的过滤器在其他地方也应该有用。其次，权重共享有助于防止在通过多层反向传播时的梯度扩散，因为权值共享的梯度被聚合，这使得深层结构上的监督学习更容易。然而，全局权值共享在具有固定空间布局的图像(如人脸)上效果并不好。例如，虽然眼睛和嘴巴可能有共同的低级特征(例如边缘)，但它们在高层次上是非常不同的。因此，对于输入包含不同语义区域的网络，局部在高层共享权值对于学习不同的高层特征(如眼睛、鼻子和嘴巴)更有效。局部共享权的思想最初是针对卷积深度置信网提出的，用于人脸识别[12]。

### 3.2 多级回归

我们找到了几种有效的方法来组合多级卷积网络。首先是多级回归。面部边界框是第一级网络唯一的先验知识。由于姿态变化较大，人脸检测器不稳定，人脸点与边界框的相对位置变化幅度较大。因此，为了涵盖许多可能的预测，第一层网络的输入区域应该很大。但是大输入区域是不准确的主要原因，因为包含的不相关区域可能会降低网络的最终输出。第一层网络的输出为以下检测提供了较强的先验条件，人脸点的真实位置应该在第一层预测周围的小区域内。因此，第二级检测可以在一个小区域内完成，在这个小区域内，来自其他区域的干扰显著减少，并且这个过程重复。然而，在没有背景信息的情况下，局部区域的出现是模糊的，并且预测是不可靠的。为了避免移动，我们不应该级联太多级或过多地信任其他级。这些网络只能在很小的范围内调整初始预测。

为了进一步提高检测精度和可靠性，我们建议联合预测每个级别的多个网络的每个点的位置。这些网络在输入区域不同。人脸特征点的最终预测位置可以正式表示为

(1)

表示li预测在i级的n级级联网络。注意，第一级的预测是绝对位置，而下一级的预测是调整。

# 4 实施细节

输入层由I(h,w)表示，其中h和w是输入区域的高度和宽度。因为不使用颜色信息，输入表示为2D。如果使用绝对值校正，则卷积层由CR(s,n,p,q)表示，否则为C(s,n,p,q)。 s是方形卷积内核（或滤波器）的边长。n是卷积层中的映射数。p和q是权重共享参数。卷积层中的每个映射均匀地划分为p×q个区域，并且在每个区域中局部权值共享。通过设置p = q = 1，可以将传统卷积网络视为一种特殊情况。默认情况下，滤波器步长在两个方向上均为1像素。设(h,w,m)为前一层的大小，即m个映射，每个大小为h的映射为w。然后C(s,n,p,q)采取的操作是

x和y是前一层和当前层的输出。w是权重，b是偏差。前一层的m映射与 s \* s内核相关联。得到的映射与一个偏差一起累积并传递双曲正切非线性，形成卷积层中的一个n映射。对于不同的输出映射和映射中的不同区域，内核集和偏差是不同的。CR(s,n,p,q)相似，但在双曲正切后有额外的取绝对值操作。

池化层用P(s)表示。s是正方形汇聚区域的边长。使用最大池，池区域没有重叠。池化结果与增益系数(g)相乘，并由偏置(b)偏移，然后是双曲正切非线性。增益和偏置系数的共享方式与前一卷积层的权值共享方式相似。P(s)表示为

完全连接的层级由具有功能的F(n)表示

其中n和m是当前层和前一层的神经元数。

**结构** 第一级网络是深度卷积网络，包括四个卷积阶段、绝对值校正和局部共享权值。第二级和第三级的网络共享一个共同的浅层结构。由于它们旨在提取局部特征，因此不需要深层结构和局部权值共享。表格1总结了网络结构的细节。

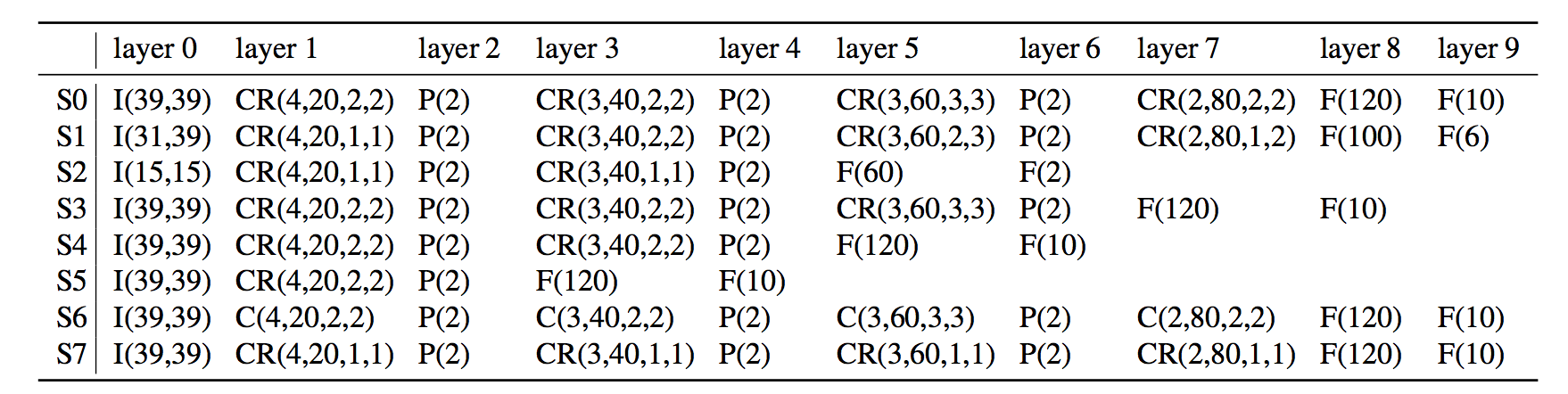


表1:网络结构概述。F1采用S0。EN1和NM1都采用S1。第二层和第三层的所有网络共享S2。为了研究网络结构的不同设计，我们还对不同结构的S3-S7进行了比较

**输入范围** F1将整个人脸作为输入并输出所有五个特征点的位置。 EN1将脸部的顶部和中部作为输入，并输出两个眼睛中心和鼻尖的位置。 NM1将面部的中部和底部作为输入，并输出鼻尖和两个嘴角的位置。所有第二级和第三级的网络都以上一及预测位置为中心的小正方形作为输入和输出增量预测。在这两级中的每个级，我们使用两个不同大小的区域来预测每个点。第三级的区域面积小于第二级的区域面积。表2列出了所有网络的精确输入范围。



表2:网络输入范围的摘要，由左、右、顶和底边界位置描述。对于水平为1 (L1)的网络，四个边界位置相对于具有边界位置(0,1,0,1)的归一化面包围框。对于水平2 (L2)和水平3 (L3)的网络，四个边界位置相对于预测的面点位置。

**训练** 在第一级，我们根据人脸边界框提取训练补丁，并通过小幅度平移和旋转对其进行增强。在接下来的层级中，我们以地面真实位置随机移动的位置为中心进行训练补丁。水平方向和垂直方向的最大偏移量在第二级为0.05，第三及为0.02，其中距离用人脸边界框归一化。第三级网络的目标是比第二级对先前的预测进行更精细的调整。可学习的网络参数包括：权值w、增益g和偏置b，这些参数由小随机数初始化，通过随机梯度下降学习。采用Levenberg-Marquardt方法[19]分别对神经元的学习速度进行估计。训练一直持续到收敛为止。

# 5 实验

我们首先用自己收集的训练集和验证集来研究网络和级联结构的不同设计。然后，在不改变训练集的情况下，对两个公共测试集上的最新方法和商业软件进行了比较。我们的训练和验证集与两个公共测试集没有重叠。

### 5.1 研究网络和级联结构

我们创建了一个包含13,466张人脸图像的数据集，其中5,590张图像来自LFW [13]，其余7,876张图像从网上下载。每张人脸都标有五个关键点的位置。 我们随机选择10,000张图像进行训练，剩余的3,466张图像进行验证。用平均检测误差和每个面部点的失败率来测量性能。 它们表明了算法的准确性和可靠性。 检测误差测量为

(2)

其中和是真实位置和检测到的位置，l是我们的人脸检测器返回的边界框的宽度。如果错误大于5％，则将其视为失败。需要注意的是，双目距离作为检测误差的归一化器更为常用，但对于姿态变化较大的人脸存在问题，因为近侧面人脸的双目距离远小于正面人脸的双目距离。它的缺点也在[28]中被提到。因此，我们使用人脸边界框的宽度代替在5.1节中验证我们的算法，并在5.2节中切换到双眼距离进行公平比较。来自我们的验证集和检测结果的一些示例图像显示在图8a中。

**网络结构** 我们以网络F1为例来研究网络深度、绝对值校正和权重共享方案如何影响性能。我们研究了表1中总结的六种不同的网络结构，并且在图4中比较了它们的性能。S0、S3-S5都具有绝对值校正和局部权值共享，但具有不同的深度。结果表明，通过包含更多级网络，可以显著提高性能。由于输入的人脸区域大小为39×39，并且随着图像的上下即网络不断向下采样，S0已经达到了可能的最大层数。S6和S7与S0具有相同的层。但是S6并没有采用绝对值校正，S7在所有卷积层中全局权重共享。结果表明，绝对值校正和局部权值共享在人脸点检测中都是有效的。我们还发现，在较高的层中局部权值共享更重要，而只有在较低的层中局部权值共享会降低性能，这与我们推测的高级特征比低级特征不太可能共享的猜想相吻合。

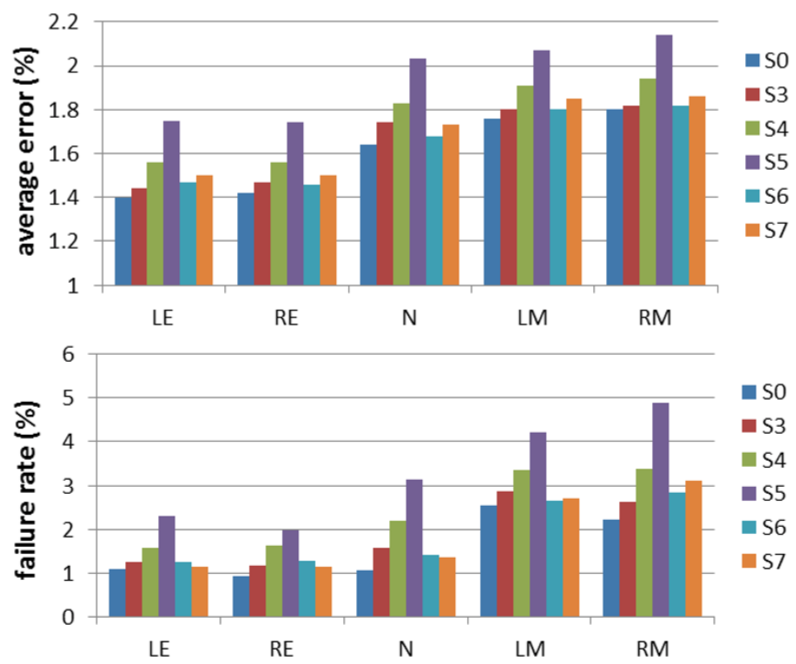


图4:不同结构卷积网络F1的平均检测误差和故障率。

**多级预测** 通过多级级联预测和多级预测在同一级别的融合，可以有效地降低检测误差。图5比较了三个网络在第一级时的性能，并对不同级别的网络进行了级联预测。在第二级大大减少了检测误差和故障。在第三级，平均检测误差略有降低，而故障率几乎保持不变。

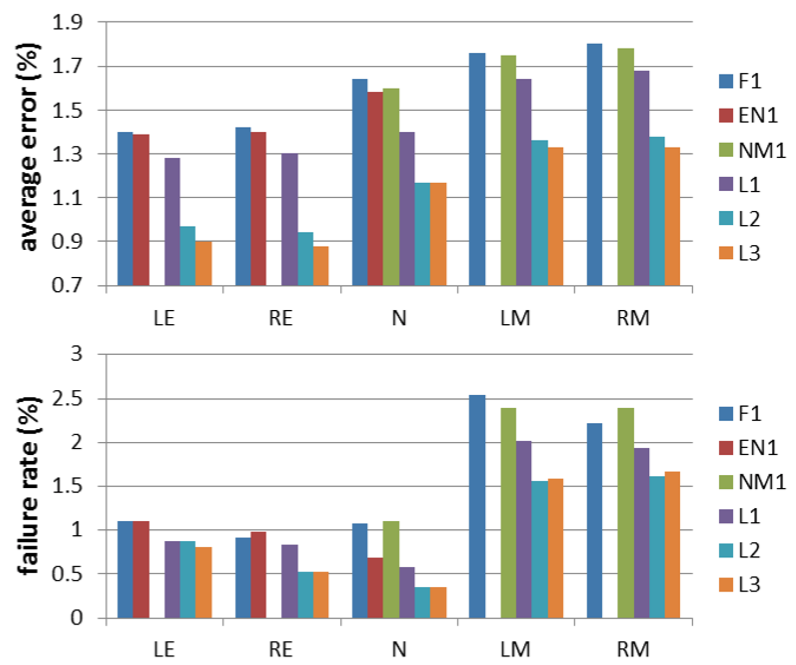


图5:F1、EN1、NM1三个网络在第1级的平均检测错误和故障率，它们的组合表示为L1、前两个级联级联水平(L2)和三个级联水平(L3)。由于NM1没有预测LE，因此错过了相应的结果。它也适用于其他关键点和网络。

### 5.2 与其他方法比较

我们将两种公共数据集BioID [15]和LFPW [4]与最先进的方法和最新的商业软件进行比较。BioID包含在实验室条件下收集的人脸图像，而LFPW包含来自网络的面部图像。在两个测试集上，我们使用在5.1节中描述的数据集上训练的模型。为了与大多数先前的工作保持一致，我们使用双眼距离来规范化检测误差，并将故障率重新定义为归一化误差大于10％的情况的比例。结果总结在图6中。

BioID有23名试验者的1,521张图像。所有的脸都是正面的，在光照和表情上有适度的变化。并与基于组件的判别搜索[20]进行了比较，利用马尔科夫网络[26]进行了增强回归，并与两款最新的商业软件Luxand Face SDK[1]和Microsoft Research Face SDK[2]进行了比较。由于Microsoft Research face SDK没有检测到眼睛中心和鼻尖，我们只比较嘴角。我们的方法显著减少了检测误差，并且我们的故障率接近于零。图8b显示了我们对BioID的一些检测结果。

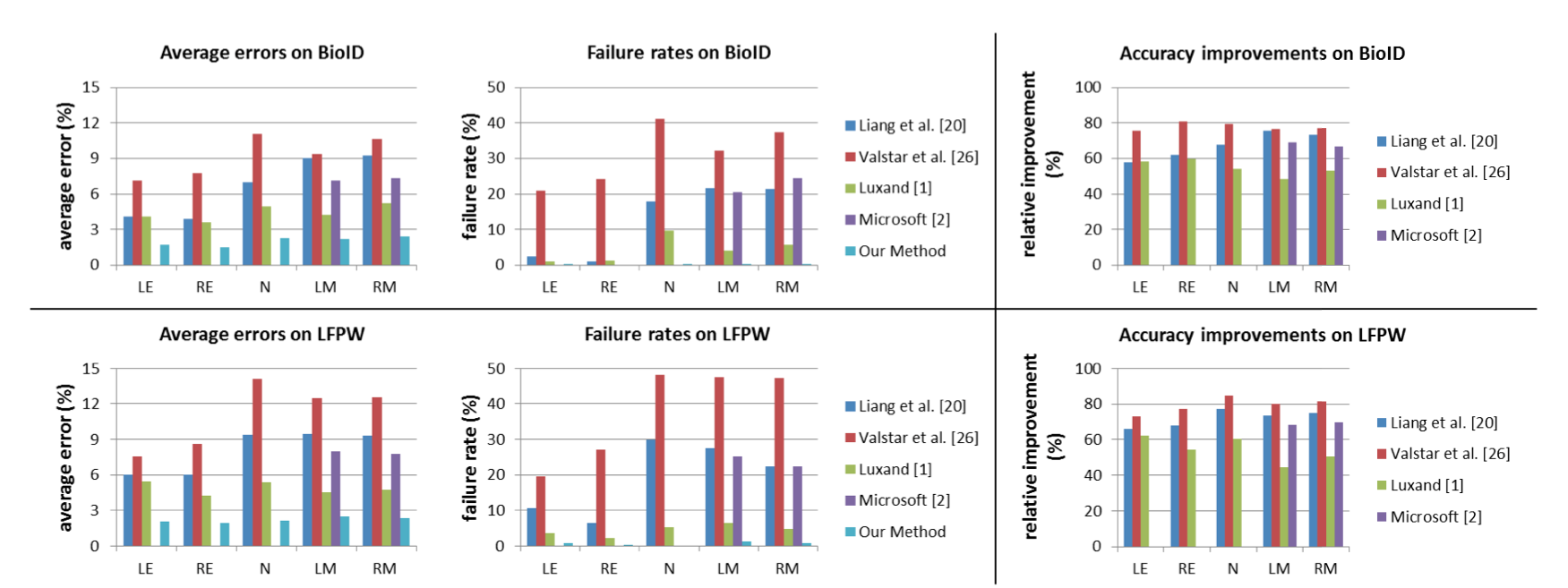


图6:BioID与LFPW比较。尽管我们基于BioID检测的失败率接近为零,它可能从图中无法看出。 。我们在这两个数据集都取得了超过50%的精度改进。

LFPW包含1,432张来自网络的脸部图像。它分为1,132张训练图像和300张测试图像。该数据集旨在测试无约束条件下的面部点检测，并且面部在姿势、光照和表达方面显示出很大的变化，并且可能包含遮挡。它仅共享图像URL，并且一些图像链接不再有效。我们只下载了781张训练图像和249张测试图像。因为我们的训练/验证数据集与LFPW之间没有重叠，我们使用LFPW训练和测试图像作为我们的测试图像，并与上述四种方法进行比较。我们的方法再次在无约束条件下表现出优异的表现。图8c显示了我们在LFPW上的一些检测结果。

Belhumeur等人[4]和Cao等人[5]报道了LFPW测试图像的结果，后者将这些图像定义为当前最新技术。[4]中的结果是所有300个LFPW测试图像。由于图像URL消失，[5]中的结果是300个LFPW测试图像中的249个。我们还在这249 张LFPW测试图像上评估我们的算法，以便与该数据集上的两种最佳方法进行比较。图7显示了两种方法的平均误差比较结果和相对精度的改进。[4,5]是非常有竞争力的方法，其表现明显优于同时代人。尽管如此，我们还是大幅改善了他们的成绩。鼻尖和两个嘴角的相对精度提高了20％以上。我们算法的C ++实现需要0.12秒来处理3.30GHz CPU上的一个图像。由于每个级别的卷积网络是独立的，因此系统可以很容易地并行化。

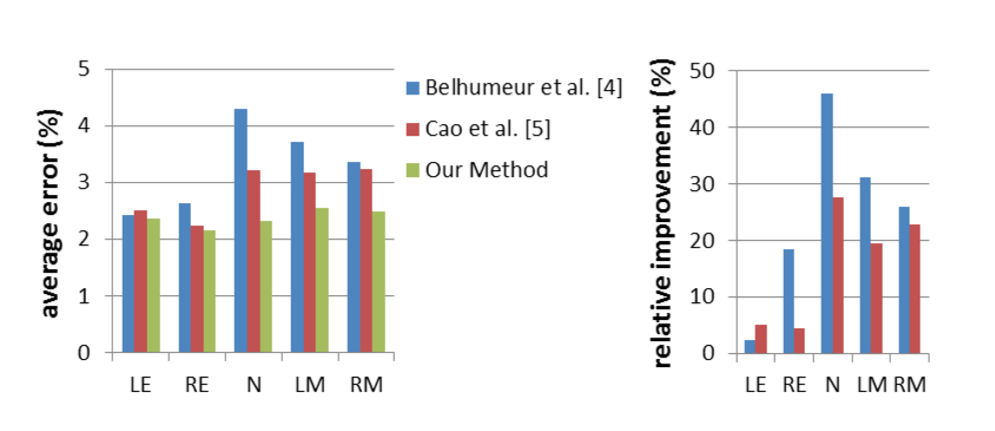


图7:LFPW测试图像与belmood等人[4]、Cao等人[5]比较。

# 6 结论

提出了一种有效的用于人脸点检测的级联卷积网络方法。第一级的深度卷积网络提供了高度鲁棒的初始估计，而第二级的较浅卷积网络则对初始预测进行微调，以达到较高的精度。通过研究网络结构的几个关键特性，我们实现了规模相对较小的高性能卷积网络。我们的方法大大提高了最先进方法和最新商业软件的预测精度。



(a)验证图像:面部在姿势(第1、2、3栏)和表情(第6栏)上差异很大，可能有遮挡(第4、5栏)、伪影(第7栏)或模糊效果(第8栏)。



(b)BioID



(c) LFPW.面部在姿势(第1、2栏)、表情(第5、6栏)和光照(第8栏)上变化很大，可能有遮挡(第3、4栏)或模糊效果(第7栏)。

图8:验证图像、BioID和LFPW上的结果。

**参考文献**

[1] http://www.luxand.com/facesdk/. 6

[2] http://research.microsoft.com/en-us/ projects/facesdk/. 6

[3] B. Amberg and T. Vetter. Optimal landmark detection using shape models and branch and bound. In Proc. ICCV, 2011. 1, 2

[4] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In Proc. CVPR, 2011. 1, 2, 6, 7

[5] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In Proc. CVPR, 2012. 1, 2, 6, 7

[6] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In Proc. CVPR, 2012. 2

[7] A. Coates, A. Y. Ng, and H. Lee. An analysis of singlelayer networks in unsupervised feature learning. Journal of Machine Learning Research, 2011. 2

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In Proc. ECCV, 1998. 1

[9] M. Dantone, J. Gall, G. Fanelli, and L. J. V. Gool. Real-time facial feature detection using conditional regression forests. In Proc. CVPR, 2012. 1, 2

[10] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. PAMI, 2013. 2

[11] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In Proc. ECCV, 2008. 1

[12] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representations for face verification with convolutional deep belief networks. In Proc. CVPR, 2012. 4

[13] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007. 5

[14] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In Proc. ICCV, 2009. 2, 3

[15] O. Jesorsky, K. J. Kirchberg, and R. Frischholz. Robust face detection using the hausdorff distance. In Proc. AVBPA, 2001. 6

[16] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. LeCun. Learning convolutional feature hierarchies for visual recognition. In Proc. NIPS, 2010. 2

[17] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In Proc. NIPS, 2012. 2

[18] Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean, and A. Ng. Building high-level features using large scale unsupervised learning. In Proc. ICML, 2012. 2

[19] Y. LeCun, L. Bottou, G. Orr, and K. Muller. Efficient backprop. In G. Orr and M. K., editors, Neural Networks: Tricks of the trade. Springer, 1998. 5

[20] L. Liang, R. Xiao, F. Wen, and J. Sun. Face alignment via component-based discriminative search. In Proc. ECCV,2008. 1, 2, 6

[21] X. Liu. Generic face alignment using boosted appearance model. In Proc. CVPR, 2007. 1

[22] P. Luo, X. Wang, and X. Tang. Hierarchical face parsing via deep learning. In Proc. CVPR, 2012. 2

[23] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In Proc. ECCV, 2008. 1

[24] M. Osadchy, Y. L. Cun, and M. L. Miller. Synergistic face detection and pose estimation with energy-based models. Journal of Machine Learning Research, 2007. 2

[25] P. Sauer, T. Cootes, and C. Taylor. Accurate regression procedures for active appearance models. In Proc. BMVC, 2011. 1, 2

[26] M. Valstar, B. Martinez, X. Binefa, and M. Pantic. Facial point detection using boosted regression and graph models. In Proc. CVPR, 2010. 1, 2, 6

[27] H. Wu, X. Liu, and G. Doretto. Face alignment via boosted ranking model. In Proc. CVPR, 2008. 1

[28] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In Proc. CVPR, 2012. 1, 2, 5