

摘要:

我们提出了一种方法来解决无约束人脸检测中的挑战,例如任意姿势变化和遮挡。首先,提出了称为归一化像素差(NPD)的新图像特征。NPD特征被计算为两个像素值之间差与和的比值,受实验心理学中的韦伯分数的启发。新特征是尺度不变,有界,并且能够重建原始图像。其次,我们通过回归树学习NPD特征的最佳子集及其组合,从而可以通过学习规则划分复杂的面流形。这样,只需要一个级联分类器来处理无约束的人脸检测。此外,我们展示了可以从查找表中有效地获得NPD特征,并且可以容易地缩放检测模板,使得所提出的面部检测器非常快(对于640x480分辨率视频大约178 FPS和对于1920x1080分辨率视频大约30 FPS在桌面计算机上),比OpenCV快6倍左右。在三个公共人脸数据集(FDDB, GENKI和CMU-MIT)上的实验结果表明,所提出的方法在检测具有任意姿势变化和杂乱场景中的遮挡的无约束面部方面优于最先进的方法。

关键词 — 无约束人脸检测, 归一化像素差, 回归树, AdaBoost, 级联分类器

1. 介绍

人脸检测的目的是在图像中查找和定位人脸。这是自动人脸识别应用的第一步。对于正面和近额面,人脸检测已经有了很好的研究。Viola和Jones的人脸检测器[1]是最著名的人脸检测算法,它基于Haar-like特征和级联AdaBoost[2]分类器。然而,在诸如人群中的面部等不受约束的场景中,由于大的姿势变化,照明变化,遮挡,表情变化,离焦模糊和低图像分辨率,最先进的面部检测器不能很好地执行。例如,由于上述困难,Viola-Jones面部检测器未能检测到面部检测数据集和基准(FDDB)数据库[3](图1中所示的示例)中的大多数面部图像。在本文中,我们将具有任意面部变化的面部检测称为无约束面部检测问题。我们对无限制场景中的人脸检测感兴趣,例如视频监控或手持设备捕获的图像。

在Viola和Jones的工作[1]之后,已经开发了许多面部检测方法,主要集中在提取不同类型的特征和开发不同的级联结构。已提出各种复杂特征[4], [5], [6], [7], [8], [9], [10], [11], [12], [13]来取代[1]中使用的Haar特征。虽然这些方法可以在一定程度上提高人脸检测性能,但是它们会产生非常大量(数十万)的特征,并且由此产生的系统需要花费太多时间来训练。人脸检测的另一个发展是学习用于多视图人脸检测的不同级联结构,例如并行级联[14],金字塔结构[15]和宽度优先搜索(WFS)树[16]。所有这些方法都需要为每个特定的面部视图(或视图范围)学习一个级联分类器。然而,在不受约束的情况下,定义所有可能的面部视图并不容易,并且计算成本随着复杂级联结构中的分类器数量的增加而增加。而且,这些方法需要在每个训练图像中手动标记面部姿势。

虽然一些可用的方法[14], [15], [16]可以处理多视图面,但它们无法同时考虑其他挑战,如遮挡。事实上,由于这些方法需要将多视图数据分成已知姿势,因此以这种方式处理遮挡并不容易。另一方面,虽然有几项研究涉及遮挡下的人脸检测[17], [18], [19], [20], [21],但他们仅限于检测遮挡下的正面。如[22]中所讨论的,强健的人脸检测算法应该在姿势和遮挡的任意变化下有效,这仍然是一个尚未解决的挑战性任务。

在本文中,我们感兴趣的是开发有效特征和鲁棒分类器,用于任意面部变化的无约束人脸检测。首先,我们提出一个简单的像素级特征,称为归一化像素差(NPD)。NPD计算为任意两个像素强度值之间的差值与它们的值之和的比率,其形式与实验心理学中的Weber分数相同[23]。NPD特征具有若干期望的属性,例如尺度不变性,有界性和重建原始图像的能力。我们进一步表明,可以从查找表中获得NPD特征,并且可以容易地缩放得到的面部检测模板以用于多尺度面部检测。

其次,我们开发了一种方法来构建单个级联AdaBoost分类器,该分类器可以有效地处理复杂的面部流形并处理任意的姿势和遮挡条件。虽然单个NPD特征可能具有“弱”判别能力,但我们的工作表明NPD特征的子集可以被最佳地学习和组合以在回归树中构建更多的判别特征。这样,不同类型的面部可以自动划分为回归树的不同叶片,并且可以在学习过程中划分高维空间中的复杂面部流形。这是一种“分而治之”的策略,用于在单个分类器中处理无约束的面部检测,而无需在面部图像的训练集中预先标记视图。得到的面部检测器对于姿势,遮挡和照明的变化以及模糊和低图像分辨率是鲁棒的。

这项工作的新颖性总结如下:

1.提出了一种称为NPD的新型特征,其有效地计算并具有若干期望的属性,包括尺度不变性,有界性以及能够重建原始图像。

2.NPD功能的子集将自动学习并组合在回归树中,以提高其可辨别性。通过这种方式,只需要一个级联的AdaBoost分类器来处理具有遮挡和任意的无约束面

该方法的优点包括:

NPD特征评估非常快,需要使用查找表进行单个内存访问。

通过应用预先缩放的检测模板，可以轻松实现多尺度人脸检测

无约束人脸检测器不依赖于姿态特定的级联结构设计；也不需要训练阶段进行姿势标记或聚类。面部检测器能够在不受约束的场景中处理照明变化，姿势变化，遮挡，离焦模糊和低分辨率面部图像。本文的其余部分安排如下。在第2节中，我们回顾了相关工作。在第3节中，我们介绍了NPD特征空间。提出的基于NPD的人脸检测方法在第4节中介绍。实验结果在第5节中提供。最后，我们总结了第6节中的贡献。

2. 相关工作

如在面部检测方法[24]的调查中所指出的，最流行的面部检测方法是基于外观的，其使用局部特征表示和分类器学习。Viola和Jones的人脸探测器[1]是第一个在级联AdaBoost分类器中应用矩形Haar类特征进行实时人脸检测的人。已经提出了许多围绕Viola-Jones检测器的方法以推进面部检测中的现有技术。Lienhart和Maydt [4]提出了一组扩展的Haar特征，其中引入了45°旋转的矩形特征。李等人。[5]提出了哈尔特征的另一种扩展，其中矩形可以在空间上以灵活的距离分开。Jones和Viola [6]也提出了类似的特征，称为对角线滤波器。已经引入了各种其他局部纹理特征用于人脸检测，例如修改后的人口普查变换[7]，局部二值模式（LBP）[8]，MB-LBP [11]，LBP直方图[10]和局部组合二进制特征[12]。已经证明这些特征对于照明变化是稳健的。米塔等人。[9]提出了类似Haar的联合特征来捕捉有效类Haar特征的共生。黄等人。[16]在粒状空间中提出了稀疏特征集，其中粒子由矩形表示，并且每个单独的稀疏特征被学习为粒子的组合。[9]和[16]中方法的一个问题是关节特征空间非常大，使得最佳组合成为一项艰巨的任务。

虽然更复杂的特征可以为面部检测任务提供比类Haar特征更好的辨别能力，但是它们通常会增加计算成本。相比之下，图像区域之间的序数关系是简单而有效的图像特征[25]，[26]，[27]，[28]，[29]，[30]。Sinha [25]在人脸图像中研究了几种强大的序数关系，并相应地开发了一种人脸检测方法。廖等人。[28]进一步表明，AdaBoost分类器可以有效地学习序数特征用于人脸识别。萨德尔等人。[26]表明像素序数特征（任意两个像素之间的序数关系）可以忠实地编码图像结构。Baluja等。[27]表明，简单的像素序数特征足以区分五个面部方向，这是一个比面部检测相对简单的任务。王等人。[30]将随机森林分类器与像素序数特征一起应用于面部标志定位。Abramson和Steux [29]提出了一种基于像素控制点的人脸检测功能，其中每个特征与两组像素位置（控制点）相关联。然而，由于控制点组合的数量巨大，因此学习基于控制点的特征并不容易。

除了不同的特征表示，一些研究人员还尝试了不同的AdaBoost算法和弱分类器。对于用于增强的弱分类器，Lienhart等人。[31]和布鲁贝克等人。[32]已经表明，分类和回归树（CART）[33]比简单的决策树桩更有效。在本文中，我们表明通过在回归树中集成所提出的NPD特征，可以学习最优的特征及其组合。以这种方式，可以将无约束的面部变化自动划分为学习的回归树的不同叶片。

鉴于最初的Viola-Jones人脸检测器在多视图人脸检测方面存在局限性[24]，已经提出了各种级联结构来解决多视图人脸检测问题[6]，[14]，[15]，[16]。Jones和Viola [6]通过为每个特定姿势训练一个面部探测器来扩展他们的面部探测器。为了避免评估每个扫描子窗口上的所有面部探测器，他们在面部检测之前开发了姿势估计步骤（类似于Rowley等人[34]），然后仅应用了在该估计姿势上训练的面部探测器。在该两阶段检测结构中，如果姿势估计不可靠，则在第二阶段中不可能检测到面部。吴等人。[14]提出了一种用于多视图人脸检测的并行级联结构，其中必须针对每个扫描窗口评估调谐到不同视图的所有面部探测器；他们确实使用了所有面部探测器的前几个级联层来估计加速的姿势。Li和Zhang [15]提出了一种用于多视图人脸检测的粗到细金字塔结构，其中整个面部姿势范围被分成越来越小的子范围，从而产生更有效的检测结构。黄等人。提出了一种基于WFS树的多视图人脸检测方法，该方法也以粗到细的方式工作。他们提出了用于多类学习的Vector Boost算法，它非常适合多视图姿态估计。然而，所有这些方法都需要为面部的每个特定视图（或视图范围）学习级联分类器，这要求输入面部图像通过检测结构的不同分支。因此，它们的计算成本通常随着复杂级联结构中的分类器的数量而增加。而且，这些方法需要在每个训练图像中手动标记面部姿势。

Lin和Liu [19]提出将多视图人脸检测器学习为单个级联分类器，而不是设计检测结构。他们通过在不同类之间共享特征来推导出一种多类增强算法，称为MBHBoost。与设计复杂的级联结构相比，这是一种更简单的多视图人脸检测方法。然而，它仍然需要手动标记姿势。然而，在不受控制的环境中，通过离散姿势空间来定义面部的特定视图并不容易，因为面部可以在偏航（平面外），滚动（平面内）中同时处于任意姿态，并且俯仰（上下）角度。为了避免手动标记，Seemann等人。[35]建议自动学习视点聚类用于对象检测。然而，对于人类面孔，Kim和Cipolla [36]表明，通过K-Means等传统技术进行聚类不会导致分类姿势。因此，他们提出了一种多分类器增强（MCBoost），用于对象图像的人类感知聚类，这显示出聚类面部姿势的希望。然而，群集并不总是与姿势变化相关；除了不同的姿势聚类之外，他们还获得了具有各种照明变化的聚类。

存在遮挡的面部检测也是无约束面部检测中的重要问题，但与多视图面部检测相比，它受到的关注较少。这可能是因为与姿势变化相比，将任意遮挡分类为预定义类更加困难。Hotta [17]提出了一种基于局部核的SVM人脸检测方法，该方法在检测被遮挡的正面方面优于基于全局核的SVM。林等人。[18]除了标准的面部检测器之外，还通过训练8个额外的级联分类器来考虑8种手动定义的面部遮挡。Lin和Liu [19]进一步提出了MBHBoost算法来处理具有12个面内旋转之一或8种类型的遮挡之一的面，每种旋转和遮挡被视为不同的类。陈等人。[20]提出了一种改进的Viola-Jones人脸检测器，其中训练的检测器被分成与几个预定义的局部片相关的子分类器，并且子分类器的输出被融合。Goldmann等人。[21]提出了一种基于组件的人脸检测方法，其中分别检测两只眼睛，鼻子和嘴巴，并进一步连接在拓扑图中。然而，在不受约束的情况下，上述方法都没有考虑同时具有遮挡和姿势变化的面部检测。如[22]中所讨论的，强健的人脸检测器应该在姿势和遮挡的任意变化下有效，这尚未解决。

最近，无约束的人脸检测受到了关注。Jain和Learned-Miller [3]开发了FDDB数据库和基准，用于开发无约束人脸检测算法。此数据库包含从Internet收集的图像，并提供面部检测的具有挑战性的方案。Subburaman和Marcel [37]提出了一种用于人脸检测的快速边界框估计技术，其中边界框是通过基于小补丁的局部搜索来预测的。Jain和Learned-Miller [38]提出了一种在线域适应方法，以提高Viola-Jones人脸检测器在FDDB数据库上的性能。李等人。[13]提出在AdaBoost级联中使用SURF特征[39]，并使用曲线下面积（AUC）标准来加速面部检测器训练。朱和拉马南[40]提出联合检测面部，估计姿势，并定位野外的面部地标。沉等人。[41]提出了一种基于样本的人脸检测方法，该方法从大的带注释的人脸数据集中检索图像，从注释中推断出面部地标位置。李等人。[42]提出了概率弹性部分（PEP）模型，以使任何预先训练的面部检测器适应特定的图像集合，如FDDB。该方法提取由一般面部检测器检测的每个候选面部的PEP表示，并训练具有顶部正面和负面样本的分类器。尽管这些方法可用于无约束人脸检测，但检测精度仍然不令人满意，尤其是当要求检测器具有低误报警时。

3. 标准像素差异特征空间

图像中两个像素之间的归一化像素差（NPD）特征定义为

$$f(x, y) = (x - y) / (x + y) \quad (1)$$

其中 $x, y \geq 0$ 是两个像素1的强度值，并且当 $x = y = 0$ 时 $f(0,0)$ 被定义为0。NPD特征测量两个像素值之间的相对差异。 $f(x, y)$ 的符号表示两个像素 x 和 y 之间的序数关系， $f(x, y)$ 的大小测量 x 和 x 之间的相对差异（作为关节强度 $x + y$ 的百分比）。年。注意，定义 $f(0,0) = 0$ 是合理的，因为在这种情况下，两个像素 x 和 y 之间没有差异。与绝对差 $|x - y|$ 相比，NPD对于像素强度的比例变化是不变的。

韦伯是实验心理学的先驱，他指出，刺激幅度变化的明显差异与刺激的幅度成正比，而不是其绝对值[23]。这被称为韦伯定律。换句话说，人类对刺激差异的感知通常被测量为原始刺激的一部分，即，形式为 $\Delta I / I$ ，其被称为韦伯分数。陈等人。[43]提出了一种局部图像描述符，称为用于人脸识别的韦伯定律描述符，它是根据 3×3 窗口中像素的韦伯分数计算出来的。方程式中提出的特征。（1）也被用于其他领域，如遥感，其中归一化差异植被指数（NDVI）[44]被定义为可见红光和近红外光谱之和的差异，以估算绿色植被覆盖度。

NPD功能具有许多理想的属性。首先，NPD特征是反对称的，因此 $f(x, y)$ 或 $f(y, x)$ 足以用于特征表示，从而导致特征空间减小。因此，在 $s \times s$ 图像块（矢量化为 $p \times 1$ ，其中 $p = s \cdot s$ ）中，计算像素对 $1 \leq i < j \leq p$ 的NPD特征 $f(x_i, x_j)$ ，得到 $d = p(p - 1) / 2$ 功能。例如，在 20×20 面部模板中，总共存在 $(20 \times 20) \times (20 \times 20 - 1) / 2 = 79,800$ 个NPD特征。我们将得到的特征空间称为NPD特征空间，表示为 $\Omega_{npd} (\in \mathbb{R}^d)$ 。

其次， $f(x, y)$ 的符号表示 x 和 y 之间的序数关系。序数关系已被证明是物体检测和识别的有效编码[25]，[26]，[28]因为序数关系编码物体图像的内在结构，并且在各种光照变化下它是不变的[25]。然而，当 x 和 y 具有相似的值时，简单地使用符号来编码序数关系可能对噪声敏感。在下一节中，我们将展示如何学习与NPD特征的强大序数关系。

第三，NPD特征是尺度不变的，其预期对于光照变化是稳健的。这对于图像表示很重要，因为对于物体检测和识别，光照变化总是一个麻烦的问题。

第四，如附录A所示，NPD特征 $f(x, y)$ 以 $[-1, 1]$ 为界。有界属性使得NPD特征适合于基于树的分类器中的直方图分级或阈值学习[1]。图2显示 $f(x, y)$ 是有界函数，它定义了一个非线性曲面。

定理1（重构）：给定NPD特征向量 $f = (f(x_1, x_2), f(x_1, x_3), \dots, f(x_p - 1, x_p)) \in \Omega_{npd}$ ，可以重建原始图像强度值 $I = (x_1, x_2, \dots, x_p) \cdot T$ 直到比例因子。

定理 1 的证明在附录 B 中示出，其还给出了线性时间方法以将原始图像重建到比例因子。定理 1 指出特征空间 Ω_{npd} 中的每个点对应于原始像素强度空间中的一组强度缩放图像。相反，尺度不变性属性表示所有强度缩放图像被“压缩”到有界特征空间 Ω_{npd} 中的点。因此， Ω_{npd} 是一个特征空间，它对于尺度变化是不变的，但它携带来自原始空间的所有必要信息。

4. NPD 用于脸部检测

序数关系[25]是一个众所周知的简单和基本概念：它比较任何两个图像区域的亮度，并相应地将结果编码为 1（更亮）或 0（更暗）。Sinha [25]表明，序数特征可以代表物体的内在结构，如人脸，它们对光照变化不敏感。在本文中，我们不是通过编码两个图像区域之间的顺序关系，而是通过 NPD 特征学习像素对之间的稳健序数关系。对于结构良好的面部图案，自动学习的序数特征组合可以比手动配置更好地表现面部。因此，我们建议通过增强回归树来学习简单序数特征的组合[33]。通过提供面部和非面部图像的训练集，通过回归树学习弱分类器。在每个节点处，树检查最佳序数特征值，然后相应地将输入数据传递到下一个分支。参见图 3。回归树也非常适合具有任意姿势变化的面部检测，因为类似的视图可以聚集在树的相同叶节点中。序数关系始终可以通过默认阈值 0 生成，但它对噪声敏感，尤其是当要比较的两个像素具有相似值时。在本

文中，我们通过学习具有 NPD 特征的回归树来学习强大的序数关系及其组合。通过这种方式，回归树不仅可以学习每个分支节点的最优 NPD 特征，还可以学习分裂的最佳阈值。通常，在分支节点上为每个 NPD 功能倾斜以下两种情况之一：

$$f(x, y) = (x - y) / (x + y) < \theta_1 < 0 \quad (2)$$

$$f(x, y) = (x - y) / (x + y) \geq \theta_2 > 0 \quad (3)$$

其中 θ_1 和 θ_2 是阈值。式。(2) 适用于物体像素 x 明显比像素 y 暗的情况，而 Eq。(3) 覆盖了像素 x 明显比像素 y 亮的情况。学习的阈值允许学习的回归树中的序数编码表示内在的对象结构。要学习这样的回归树，我们使用具有 NPD 特征的 CART 算法[33]。

鉴于提出的 NPD 特征包含冗余信息，我们还应用 AdaBoost 算法来选择最具辨别力的特征并构建强分类器[1]。我们采用 Gentle AdaBoost 算法[2]来学习基于 NPD 特征的回归树。

如[1]中所述，进一步学习级联分类器以进行快速面部检测。我们只学习一个单一的级联分类器，用于对遮挡和姿势变化具有鲁棒性的无约束人脸检测。该实施方式的优点在于，不需要在训练检测器之前手动标记每个面部图像的姿势或者对姿势进行聚类。在学习过程中，算法通过回归树自动将整个面流形分为几个子流形。

下面概述了所提出的方法如何处理无约束的人脸检测问题。

姿势。 通过学习增强回归树中的 NPD 特征来处理姿势变化，其中不同视图可以自动划分为回归树的不同叶。

遮挡。 与由于大支持而对遮挡敏感类哈尔特征相比[18]，NPD 特征仅由两个像素值计算，使其对遮挡具有鲁棒性。

照明。 由于 NPD 特征是尺度不变的，因此它们对光照变化很稳健。

模糊或低图像分辨率。 由于 NPD 功能仅涉及两个像素值，因此它们不需要在面上使用丰富的纹理信息。这使得 NPD 功能可有效处理模糊或低分辨率的人脸图像。

为了进一步加快提出的 NPD 人脸检测器，我们开发了以下两种技术。首先，对于 8 位灰度图像，我们构建了一个 256×256 查找表来存储预先计算的 NPD 特征。这样，计算公式中的 $f(x, y)$ 只需要从查找表中访问一个内存。

5. 实验

我们评估了 NPD 人脸检测器在三个公共领域数据库 FDDB [3]，GENKI [45]和 CMU-MIT [34]上的性能。我们还提供了对所提方法的分析，报告了面部检测速度，并分别报告了在光照变化，姿势变化，遮挡和模糊下的无约束面部检测性能。

[13]中的训练数据 2 的子集用于训练我们的探测器，包括 12,102 个面部图像和 12,315 个非面部图像（一些私人面部图像和 Corel5k 非面部图像不可用，因此无法使用它们）。图 4 示出了来自该训练数据集的一些示例面部和非面部图像。检测模板为 20×20 像素。探测器级联包含 15 个阶段，对于每个阶段，目标误接受率为 0.5，检测率为 0.999。对于回归树的深度，我们设置一个约束，每个叶节点必须包含训练样本总数的至少 $(1/16)$ 。在此约束下，树深度最多为 5，在测试阶段，每个回归树最多需要计算 4 个 NPD 要素。我们探测器的前五个阶段分别包括 3 个，4 个，6 个，7 个，9 个弱分类器。图 5 显示了在第一阶段的三个回

归树中学习的 NPD 特征。可以观察到，大多数学习的特征是眼睛，眉毛和鼻子周围。此外，三个回归树中的特征分布在面部区域的不同部分中。这是因为，在增强方案中，当学习弱分类器时，所有样本被重新加权，使得下一个弱分类器可以关注在当前步骤中不能被正确分类的训练样本。图 5 中所示的面部是正面，但应该记住，面部可以具有任意的姿势变化，并且一些学习的特征可能仅对特定姿势有效。

在测试阶段，为多尺度检测设置比例因子 1.2。实现了类似于 OpenCV 人脸检测模块的后处理方法，该方法通过不相交集算法合并附近的检测。对于每个检测到的面部，我们将级联的所有阶段中的 AdaBoost 分类器的得分总结为最终得分；该分数用于生成接收器操作特性（ROC）。我们使用了三个公共人脸数据库，FDDB [3]，GENKI [45]和 CMU-MIT [34]来评估我们的人脸检测算法。

FDDB 数据集[3]涵盖了面部检测的挑战性场景。FDDB 中的图像来自 Faces in the Wild 数据集 [46]，这是从雅虎新闻收集的大量互联网图像。它包含 2,845 张图像，共有 5,171 张面孔，具有各种具有挑战性的场景，包括任意姿势，遮挡，不同的光线，表情，低分辨率和失焦面部。数据库中的所有面都已使用椭圆区域进行注释。图 1 显示了 FDDB 数据库中带注释的面的一些示例。

对于基准评估，Jain 和 Learned-Miller [3]提供了一个评估代码，用于比较不同的人脸检测算法。基于 ROC 的性能评估有两个度量标准：离散得分度量和连续得分度量，其分别对应于检测和基础事实之间的粗略匹配（类似于面部检测文献中的先前评估）和精确匹配。在[3]中提出了两种实验装置。第一个实验

（EXP-1）需要 10 倍交叉验证，而第二个实验（EXP-2）允许不受限制的训练，这意味着 FDDB 外部的图像可用于面部检测器训练。

我们遵循两个实验方案。对于 EXP-1，我们使用第 5.1 节中描述的设置训练了 10 个面部检测器，并使用 10 倍交叉验证单独测试它们。平均而言，我们在 FDDB 中使用大约 4,500 个面部图像来训练单个面部检测器。图 6 显示了从 FDDB 数据库中裁剪的一些人脸图像，用于训练人脸检测器。由于 FDDB 不提供一组非面部图像，我们用 FDDB 图像中的黑色斑点替换所有带注释的面部区域，然后使用生成的图像来引导非面部样本。图 7 示出了这种修改的图像。

对于 EXP-2，我们使用经过 FDDB 外部数据训练的探测器，如前一小节所述。为了评估，该检测器分别应用于 FDDB 数据库的每个子集，并报告平均性能。

我们将我们的方法与 FDDB 网站 3 上报告的最新结果进行了比较。各种算法的 ROC 曲线在图 8 中针对离散分数度量而在图 9 中针对连续分数度量来描绘。请注意，所有基线结果都是针对 EXP-2 的，因为我们在 EXP-1 协议之后没有找到任何结果。在两个图中。在图 8 和 9 中，图例中的曲线标签按照零误报（FP = 0）的检测率的降序排序。还要注意，平均而言，FP = 285 通常意味着 FDDB 实验的每个图像的一次错误检测。因此，有用的 FP 在[0,500]范围内；我们以对数标度显示 X 轴，以强调低 FP 的性能。在基线方法中，

“Olaworks Inc.”和“Illuxtech Inc.”是两种商业检测器。他们的方法以及“深圳大学”的方法尚未发表。

“SURF Cascade”是 Li 等人提出的基于 SURF 描述符的级联方法。在[13]中，这是迄今为止低误报的最佳公布结果。Zhu-Ramanan [40]的方法由 FDDB 团队进行评估，在 FDDB 网站上报告的结果现在是已发表方法中的最新技术。对于所提出的 NPD 人脸检测器，除了以最近邻方式缩放检测模板外，我们还尝试通过 MATLAB 中的默认 imresize 函数构建图像金字塔表示，并应用 20×20 检测模板。由于此函数使用带有抗锯齿的双三次插值方法，因此我们将生成的检测器称为“平滑 NPD”。

根据图 8 所示的离散得分度量结果，可以观察到所提出的方法优于除 Olaworks 公司之外的所有基线方法。但是，当 FP < 10 时，所提出的 NPD 检测器比 Olaworks 的检测器好得多。事实上，当 FP = 0 时（如图例所示），建议的 NPD 探测器在粗略意义上检测到 45%的带注释的 FDDB 面（50%与地面实况重叠），而所有基线探测器的探测率均低于 30%。请注意，对于之前用于 SURF Cascade [13]的子训练集，EXP-2 的 NPD 显示出比 SURF Cascade 更好的性能。此外，Smooth NPD 略好于 NPD，但需要额外的平滑计算成本。还观察到针对 EXP-1 和 EXP-2 训练的 NPD 检测器的结果是可比较的，尽管 EXP-2 的训练数据大小是 EXP-1 的训练数据大小的几倍。该结果表明 FDDB 包含用于无约束面部检测的代表性图像。然而，在训练单个探测器时处理所有这些数据并不容易（回想一下图 6 中面部外观的大变化）。请注意，在回归树中学习通用 NPD 特征以划分和征服复杂的面流形。

类似的观察可以在图 9 中找到连续得分度量，除了当 FP > 5 时，Zhu-Ramanan 略好于所提出的方法，并且“Smooth NPD, EXP-1”优于 Olaworks Inc.。表 1 显示了对比在 FP = 0, 10 和 100 的 FDDB 数据库中 EXP-2 的检测率是有希望的。在低误报率下，所提出的方法要么比基线方法好得多，要么与最佳表现者相当。

图 10 示出了通过所提出的 NPD 方法在 Fddb 数据库中检测到的面部的一些示例。通过所提出的方法可以成功地检测旋转、遮挡和离焦面部，如图 10 所示。一些遮挡面（例如图 10 中的第四行和第二列）和模糊面（例如右上图像）在图 10 中，仍然可以通过所提出的方法检测未在地面实况中注释的。然而，所提出的方法无法检测到许多面部，尤其是在非常拥挤的场景中（参见第 1 行中的第 1 图像和第 3 图像，以及图 10 中的右下图像）。

GENKI 数据库[45]由加州大学圣地亚哥分校的机器感知实验室收集。我们在其 SZSL 子集上评估了 GENKI 数据库 GENKI-R2009a 的当前版本，该子集包含从 Internet 收集的 3,500 张图像。这些图像包括广泛的背景、照明条件、地理位置、个人身份和种族。来自 GENKI 数据库的面部图像的一些示例在图 12 中示出，其中通过所提出的 NPD 方法进行标记检测。GENKI 数据集中的大多数图像仅包含一个面。从这个意义上讲，GENKI 数据集并不像 Fddb 数据集那样具有挑战性。GENKI-SZSL 数据集中的某些图像包含未标记的面，因此它们不适用于面部检测评估任务。在去除这些未标记的图像之后，我们留下了 3,270 个图像用于面部检测评估。对于性能评估，应用 5.1 节中描述的学习检测器是不公平的，因为用于该检测器的训练数据包含来自 GENKI 数据库 4 的面部图像。因此，我们使用在 Fddb 10 倍交叉验证的第一次折叠训练的 NPD 面部检测器来评估 GENKI 数据库。我们还评估了在 OpenCV 2.4 中实现的 Viola-Jones 人脸检测器和商用人脸检测器 PittPatt [50]。我们再次使用[3]中的基准评估代码进行性能评估，但略微修改了允许地面实况注释为矩形的代码。对于离散和连续得分度量，三种方法的 ROC 曲线如图 11 所示。结果表明，所提出的 NPD 人脸检测器的性能远优于 Viola-Jones 和 PittPatt 人脸检测器。

CMU-MIT 人脸数据集[34]是人脸检测的早期基准之一。CMU-MIT 正面数据集包含 130 个灰度图像，总共有 511 个面，其中大部分未被遮挡。我们在该数据库中应用了第 5.1 小节中描述的相同 NPD 检测器。我们还使用[3]中修改后的基准评估代码和离散分数度量进行绩效评估。图 13 显示了所提出的 NPD 人脸检测器的 ROC 曲线，软级联方法[51]，SURF 级联方法[13]和 Viola-Jones 检测器[1]。结果表明，与 Viola-Jones 正面检测器相比，NPD 检测器在假阳性数量 $FP < 50$ 时表现更好，而在较高 FP 时比 Viola-Jones 略差。与 SURF 级联检测器相比，当 $FP < 3$ 时，NPD 检测器更好，但 SURF 级联方法在更高 FP 时优于 NPD。请注意，SURF 级联方法使用大小为 40×40 像素的面部模板，这是我们的面部检测模板（ 20×20 像素）的四倍。通常，较大的面部模板包含用于面部描述的更多特征，但是在计算上更昂贵并且可能在检测模糊面部方面具有限制。此外，提出的 NPD 方法不如 CMU-MIT 数据集上最先进的软级联方法。尽管如此，所提出的 NPD 方法可以检测大约 80% 的正面而没有任何误报，这是有希望的，因为我们没有训练正面检测器。通过所提出的 NPD 方法在 CMU-MIT 数据集中检测到的一些面部在图 14 中示出。

由于所提出的人脸检测器是回归树和 NPD 特征的组合，因此确定这两个组件中的每一个的贡献是有益的。在下文中，我们在第 5.1 节中描述的相同训练集和级联训练设置上训练了所有比较的面部检测器。

首先，我们基于 NPD 特征训练了一个人脸检测器，但是使用了树桩分类器[1]，这是一个只有一个分裂节点的基本树分类器。如表 2 所示，基于树桩分类器的检测器包含 1,597 个弱分类器。相比之下，基于回归树的检测器包含 176 个弱分类器，表明在回归树中组合 NPD 特征在构建用于 AdaBoost 学习的弱分类器方面更有效。此外，在级联处理中，每个扫描子窗口需要平均评估基于树桩分类器的 36.5 NPD 特征。另一方面，对于基于回归树的检测器，平均只需要评估 34.4 个 NPD 特征，这意味着使用回归树不会增加平均计算成本。在 Fddb 数据库上测试基于树桩分类器和回归树的面部检测器。对于离散得分度量和连续得分度量，这两个检测器的 ROC 曲线如图 15 所示。如图所示，使用回归树而不是树桩分类器，对于离散度量，面部检测性能提高约 2%-10%，对于连续度量，使用 1%-7%。较小的误报率会有所改善。

接下来，我们修复了基于弱学习器的回归树，但尝试了其他三个局部特征，即 Haar 样特征[1]，LBP [52]和像素序数特征（POF）[30]。由于 LBP 是离散标签，我们将其视为回归树学习中的分类变量，即，对于每个树节点处的分支，算法找到将离散 LBP 代码分成两组的最佳标准。使用与 5.1 节相同的训练集，我们分别使用 Haar、LBP 和 POF 训练了三个探测器。这些探测器的模型复杂性总结在表 2 中。可以观察到，NPD 模型似乎比 POF 模型更有效，尽管它需要比 Haar 和 LBP 模型稍微更多的特征评估。然而，应该注意的是，Haar 类特征的计算需要计算积分图像，而对于 LBP，每个特征需要比较 8 对像素并将得到的二进制串转换为相应的十进制数。相反，使用前面提到的查找表，计算 NPD 功能只需要一次内存访问。

在 Fddb 数据库上测试具有不同局部特征的四个检测器，并且对于离散和连续得分度量，相应的 ROC 曲线在图 16 中示出。NPD 探测器的性能优于 Haar、LBP 和 POF 探测器，具有相同的回归树弱学习器。对于离散度量，由于 NPD 特征优于 Haar、LBP 和 POF 特征的性能改善分别约为 6%、10%和 6%，对于连续度量，性能分别约为 4%、6%和 4%，在 $FP = 1$ 时。NPD 优于 POF，因为利用 NPD 特征，回归树

学习最优阈值以形成更健壮的序数规则。NPD 比 Haar 和 LBP 表现更好，特别是在低误报时，表明在回归树中组合最佳像素级特征可以更好地区分人脸和非人脸。另一方面，人们还可以观察到，除了低误报率外，NPD 的性能与 Haar 样特征和 LBP 大致相同或稍好一些。

我们还尝试了 NPD 的变体，定义为。这表示为 NPD2。同

$$f(x, y) = (x - y) / \sqrt{x^2 + y^2}$$

在与 NPD 相同的设置下，我们根据 NPD2 训练了另一个探测器。FDDB 的测试结果也显示在图 16 中；NPD 和 NPD2 的性能大致相同，NPD2 略好。然而，考虑到 NPD 比 NPD2 更简单，我们仍然建议使用 Eq. (1)。

在下文中，我们评估所提出的 NPD 面部检测器如何在照明变化，姿势变化，遮挡和模糊（或低分辨率）下执行。请注意，图像中经常同时遇到这四个挑战。在我们选择的四个子集中，每个特定挑战一个，我们专注于每个图像的主要变化来源。对于每个挑战，我们从 FDDB 数据库中选择了 100 个图像[3]（示例如图 17 所示），并分别在每个子集上运行第 5.1 小节中描述的 NPD 检测器。图 18 示出 NPD 面部检测器在照明子集上执行最佳。这并不奇怪，因为所提出的 NPD 特征对于照明变化是稳健的。此外，NPD 方法对于具有姿势变化的面部图像比对遮挡或模糊的表现更好。这些结果表明，遮挡和模糊是无约束面部检测的两个主要挑战，这在文献中尚未得到很好的解决。

还将 NPD 人脸检测器与 OpenCV 2.4 中实现的 Viola-Jones 人脸检测器和上面讨论的 FDDB 的四个子集上的商用人脸检测器 PittPatt 进行了比较。得到的具有离散得分度量的 ROC 曲线如图 19 所示。这些图表示所提出的 NPD 面部检测器在所有四个子集上都优于 Viola-Jones 和 PittPatt 面部检测器。在 4.2 小节中讨论了在照明变化，姿势变化，遮挡和模糊下所提出的方法的优越性能的原因。

对于诸如移动电话的手持设备，用于计算和存储器的可用资源相当有限。因此，面部检测器的复杂性和检测速度对于嵌入式系统非常重要。在本小节中，我们报告了所提出的 NPD 人脸检测器的检测速度，与 OpenCV 2.4 中的 Viola-Jones 人脸检测器进行了比较，已知这种检测器针对速度进行了优化。提出的 NPD 人脸检测器是用 C++ 实现的；第 5.1 节训练的模型大小为 41KB。此评估选择了两个平台：(i) 具有 Intel Core i5-2400 @ 3.1GHz CPU（4 核，4 个线程）的普通台式 PC，以及 (ii) 采用 Intel Atom N450 @ 1.6GHz 处理器的上网本（1 核心，2 个线程），模拟低端设备。对于面部检测评估，使用电影“作业”的视频剪辑。此视频剪辑显示繁忙的校园，每个帧包含一到几个面。视频片段的长度约为 2 分钟，总共包含 3,950 帧。原始分辨率为 1280×720。为了测试各种分辨率下的检测速度，原始视频剪辑被裁剪并调整为 1920×1080, 800×600 和 640×480。在该评估中，要检测的最小面部大小设置为 40×40 像素，并且缩放因子为 1.2。在 NPD 和 OpenCV 检测器中都启用了多线程技术以进行并行计算。

测试结果（以每秒帧数，FPS 测量）如表 3 所示。注意，我们只计算了面部检测时间，而不管视频解码时间如何。表 3 中还比较了快速人脸检测算法 SURF 级联[13]的检测速度。SURF 级联算法的检测速度直接取自[13]，因为我们无法访问代码。[13]中的检测参数与我们的算法相同，只是[13]中的作者使用 Intel Core-i7 CPU 作为台式计算机。从表 3 可以看出，NPD 检测器比 OpenCV 和 SURF 级联检测器快得多。在 Atom N450 处理器上，NPD 探测器的探测速度比 OpenCV 探测器的探测速度快约 9 倍；在 i5 处理器上，NPD 探测器的速度大约是 OpenCV 探测器速度的 7 倍。

表 3 显示 NPD 探测器可以在 i5 台式 PC 上实时运行（29.6 FPS），用于处理 1920×1080 高清视频。对于标准 VGA（640×480）视频，i5 处理器上的 NPD 探测器可以更快的速度（177.6 FPS）检测人脸。在低端 Atom 平台上，NPD 探测器仍可以近乎实时（19.4 FPS）运行以处理 VGA 视频。NPD 处理速度快的原因有两个。首先，NPD 功能很简单，只涉及两个像素。此外，使用查找表技术，每个 NPD 功能的评估仅需要一次存储器访问。其次，NPD 功能可以轻松扩展到各种尺寸的检测模板。因此，预先计算和存储多尺度模板可以加速检测，因为避免了重新缩放输入图像。

6. 总结与将来工作

我们已经提出了一种快速准确的杂乱场景中的人脸检测方法。该方法基于归一化像素差（NPD）特征以及增强回归树。对 NPD 特征的分析表明，它具有尺度不变性，有界性和重建能力的特性。我们开发了一种学习 NPD 特征及其组合的最佳集合的方法。因此，单个级联 AdaBoost 分类器能够实现具有大的姿势变化和遮挡的面部检测的有希望的结果。对三个公共领域数据库（即 FDDB，GENKI 和 CMU-MIT）的评估表明，所提出的方法优于无约束人脸检测的最新方法。拟议的 NPD 人脸检测器可以实时处理 1920×1080 视频帧，比 OpenCV 2.4 中实现的 Viola-Jones 人脸检测器快 6 倍左右。报告的结果还表明，遮挡和模糊是人脸检测的两大挑战。我们未来的工作将使用 NPD 特征和分类器学习方法用于其他应用，例如面部属性分类（例如姿势估计，年龄估计和性别分类）和行人检测。