

与回归树集合进行的一毫秒面对齐

瓦希德·卡泽米和约瑟芬·沙利文
KTH公司，英国皇家理工学院
计算机视觉与主动感知实验室
瑞典14号，斯德哥尔摩，瑞典
{vahidk, sullivan}@csc.kth.se

摘要

本文解决了单幅图像的人脸对齐问题。我们展示了如何使用回归树的集合直接从像素强度的稀疏子集来估计人脸的地标位置，以高质量的预测实现超实时性能。我们提出了一个基于梯度增强的通用框架来学习回归树的集合，它优化了误差损失的平方和，并自然地处理缺失或部分标记的数据。我们展示了如何使用适当的先验利用图像数据的结构有助于有效的特征选择。研究了不同的正则化策略及其对抗过拟合的重要性。此外，我们还分析了训练数据的数量对预测的准确性的影响，并利用综合数据探索了数据增强的影响。

1. 介绍

在本文中，我们提出了一种新的算法，在毫秒内执行人脸对齐，并实现优于标准数据集上最先进的方法的精度。与以前的方法相比，速度增益是识别先前的人脸对齐算法的基本组成部分，然后将它们合并到一个简化的公式中，形成通过梯度增强学习的高容量回归函数级联的结果。

我们表明，正如其他人有[8, 2]，面对齐可以用级联回归函数来解决。在我们的例子中，级联中的每个回归函数都有效地从初始估计估计形状和相对于这个初始估计索引的稀疏像素集的强度。我们的工作建立在过去十年的大量研究基础上，这些研究导致了人脸对齐[9, 4, 13, 7, 15, 1, 16, 18, 3, 6, 19]的重大进展。特别是，我们将一些成功的算法中存在的两个关键元素合并在我们学习到的回归函数中，我们现在详细介绍了这些元素。

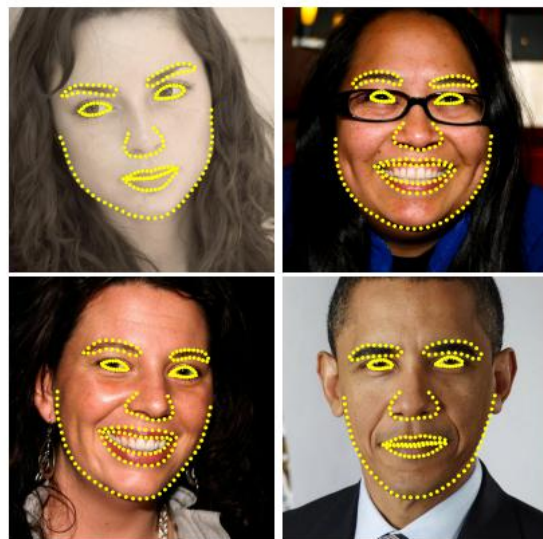


Figure 1. Selected results on the HFLRN dataset. An ensemble of random regression trees, in one millisecond, detects 194 landmarks on a face image.

第一个围绕着像素强度相对于当前形状的估计的索引。在人脸图像的矢量表示中提取的特征可能会由于形状变形和光照条件的变化等有害因素而有很大的变化。这使得使用这些特征进行准确的形状估计变得困难。困境是，我们需要可靠的特征来准确地预测形状，另一方面，我们需要准确地估计形状来提取可靠的特征。以前的工作[4, 9, 5, 8]和这项工作，使用迭代方法（级联）来处理这个问题。而不是基于在图像的全局坐标系中提取的特征来回归形状参数，而是将图像转换为基于当前形状估计的归一化坐标系，然后提取特征来预测形状参数的更新向量。这个过程通常会重复几次，直到收敛。

第二个是考虑如何对抗的困难

推理/预测问题。在测试时，一个对齐算法必须估计形状，一个高维向量，它最符合图像数据和我们的形状模型。该问题是非凸的，有许多局部最优。成功的算法[4, 9]通过假设估计的形状必须位于线性子空间来处理这个问题，这可以通过，例如，找到训练形状的主成分来实现。这一假设大大减少了在推理过程中考虑的潜在形状的数量，并有助于避免局部最优。最近的工作[8, 11, 2]使用了这样一个事实，即某一类回归变量可以保证在由训练形状定义的线性子空间中产生预测，并且不需要额外的约束。至关重要的是，我们的回归函数有这两个元素。

与这两个因素相关的是我们有效的回归函数学习。我们优化了一个适当的损失函数，并以一种数据驱动的方式执行特征选择。特别地，我们通过梯度增强和误差损失函数的平方来学习每个回归变量，这是我们在测试时最小化的相同的损失函数。稀疏像素集作为回归器的输入，通过梯度增强算法和输入像素对之间距离的先验概率的组合来选择。先验分布允许增强算法有效地探索大量的相关特征。结果是一个级联回归，可以定位面部地标初始化平均面部姿态。

本文的主要贡献是

1. 一种基于回归树集成的新对齐方法，在执行形状不变特征选择的同时最小化训练时间中我们希望在测试时最小化的损失函数。
2. 我们提出了一个方法的自然扩展，处理缺失或不确定的标签。
3. 定量和定性的结果证实了我们的方法可以产生高质量的预测，同时比之前最好的方法更有效（图1）。
4. 分析了训练数据的数量、部分标记数据的使用和综合数据对预测质量的影响。

2. 方法

本文提出了一种计算效率高的精确估计面部地标位置的算法。与之前的工作[8, 2]类似，我们提出的方法利用了一个级联的回归变量。在本节的其余部分中，我们将描述级联的各个组件的形式的细节，以及我们如何执行训练。

2.1. 回归变量的级联

首先，我们介绍了一些符号。让 $x_i \in \mathbb{R}^2$ 是图像 I 中第 i 个面部地标的 x, y 坐标。然后是向量 $S = (x_1^T, x_2^T, \dots, x_p^T)^T \in \mathbb{R}^{2p}$ 表示 I 中所有 p 个面部地标的坐标。在本文中，我们经常把向量 S 称为形状。我们使用 $\hat{S}^{(t)}$ 来表示我们对 S 的当前估计。每个回归器， $r_t(\cdot, \cdot)$ ，在级联中预测图像中的一个更新向量 $\hat{S}^{(t)}$ 它被添加到当前的形状估计值中 $\hat{S}^{(t)}$ 要改进估计，请执行以下操作：

$$\hat{S}^{(t+1)} = \hat{S}^{(t)} + r_t(I, \hat{S}^{(t)}) \quad (1)$$

级联的临界点是回归因子 r_t 根据特征进行预测，如像素强度值，从 I 中计算出来，并相对于当前的形状估计进行索引 $\hat{S}^{(t)}$ 。这在过程中引入了某种形式的几何不变性，随着级联的进行，人们可以更确定，人脸上的精确语义位置正在被索引。稍后我们将描述如何执行此索引。

请注意，如果进行初始估计，则确保由集成扩展的输出范围位于训练数据的线性子空间中 $\hat{S}^{(0)}$ 属于这个空间。因此，我们不需要对预测施加额外的约束，这大大简化了我们的方法。初始形状可以简单地选择为训练数据的平均形状，并根据通用人脸检测器的边界框输出进行缩放。

训练每个 r_t 我们使用了如[10]中所述的具有误差损失平方和的梯度树增强算法。现在我们给出了这个过程的确切细节。

2.2. 学习级联中的每个回归变量

假设我们有训练数据 $(I_1, S_1), \dots, (I_n, S_n)$ 每个 i 是一个人脸图像和 S_i 它的形状向量。来学习第一个回归函数 r_0 在级联中，我们从训练数据中创建一个人脸图像，一个初始形状估计和目标更新步骤，即， $(I_{T1}, \hat{S}_i^{(0)}, \Delta_i^{(0)}, S)$ ，其中

$$\pi_i \in \{1, \dots, n\} \quad (2)$$

$$\hat{S}_i^{(0)} \in \{S_1, \dots, S_n\} \setminus S_{\pi_i} \text{ and } \Delta_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)} \quad (3)$$

$$\pi_i \in \{1, \dots, n\} \quad (4)$$

对于 $i = 1, \dots, N$ ，其中 $N = n$ ，其中 R 是 $\Delta_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)}$ 数设置为 N ，其中 R 是 $\Delta_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)}$ 图像的每个初始形状估计都从 $\{S_1, \dots, S_n\}$ 中没有替代。

从这些数据中，我们学习到回归函数 r_0 （见算法1），使用梯度树与误差损失平方和的推进。然后更新训练三胞胎的集合

提供培训数据 $\pi_i, \hat{S}_i^{(1)}, \Delta_i^{(1)}, \dots, S$), 对于下一个回归变量 r_1 在级联中设置 ($t=0$)

$$\hat{S}_i^{(t+1)} = \hat{S}_i^{(t)} + r_t(I_{\pi_i}, \hat{S}_i^{(t)}) \quad (5)$$

$$\Delta S_i^{(t+1)} = S_{\pi_i} - \hat{S}_i^{(t+1)} \quad (6)$$

这个过程被迭代, 直到T回归器 $r_0, r_1, \dots, r_T - 1$ 当结合起来时, 可以提供足够的准确性水平。

如上所述, 每个回归变量 r_t 采用梯度增强树算法进行学习。应该记住, 使用了一个平方误差损失, 并且在最内层循环中计算的残差对应于在每个训练样本中计算的这个损失函数的梯度。在该算法的声明中包括一个学习速率参数 $0 < \nu \leq 1$, 也被称为收缩因子。设置 $\nu < 1$ 有助于对抗过度拟合, 通常会导致回归变量, 这比那些使用 $\nu = 1$ [10]学习的回归变量要好得多。

算法1学习 r_t 在梯度

有培训资料 $\pi_i, \{S_i^{(t)}, \Delta_i^{(t)}\}_{i=1}^N$ 和学习率 (收缩因子) $0 < \nu < 1$

1. 初始化

$$f_0(I, \hat{S}^{(t)}) = \arg \min_{\gamma \in \mathbb{R}^{2p}} \sum_{i=1}^N \|\Delta S_i^{(t)} - \gamma\|^2$$

2. 对于 $k = 1, \dots, K$:

(a) 设置为 $i = 1, \dots, N$

\dots, N

$$r_{ik} = S f \Delta_i^{(t)} - k - 1(I_{\pi_i}, \hat{S}_i^{(t)})$$

(b) 拟合一个回归树到目标 r_{ik} 给出了一个弱回归函数 $g_k(I, \hat{S}^{(t)})$.

(c) 更新

$$f_k(I, \hat{S}^{(t)}) = f_{k-1}(I, \hat{S}^{(t)}) + \nu g_k(I, \hat{S}^{(t)})$$

3. 输出 $r_t(I, \hat{S}^{(t)}) = f_k(I, \hat{S}^{(t)})$

3.2基于树的回归器

每个回归函数 r 的核心 t 是基于树的回归变量拟合到梯度增强算法中的剩余目标。我们现在回顾了训练每个回归树的最重要的实现细节。

3.12个形状不变的分割测试

在回归树中的每个分裂节点上, 我们根据两个像素强度差的阈值做出决定。测试中使用的像素在平均形状坐标系中定义的位置 u 和 v 。对于任意形状的人脸图像, 我们想索引与其形状位置相同的点, 因为 u 和 v 的平均形状。为了实现这一点, 在提取特征之前, 可以根据当前的形状估计将图像扭曲到平均形状。由于我们只使用图像的一个非常稀疏的表示, 因此与整个图像相比, 扭曲点的位置更有效。此外, 除了[2]所建议的局部平移外, 仅使用全局相似度变换就可以完成扭曲的粗略近似。

具体细节如下。让 k_u 是最接近 u 的平均形状的指数, 并将其与 u 的偏移量定义为

$$6x_u = u - \bar{x}_{k_u}$$

然后是一个形状的 S_i 在图像 I 中定义 i , 在 I 中的位置 i 这与 u 的平均形状图像定性相似

$$u \setminus = x_{i, k_u} + R_6 x_{\frac{1}{S_i}} u \quad (7)$$

s 在哪里 i 和 R_i 是变换 S 的相似度变换的尺度和旋转矩阵吗 $i \setminus S$ 到, 平均的形状。

尺度和旋转被发现是最小化的

$$\sum_{j=1}^p \|\bar{x}_j - (s_i R_i x_{i,j} + t_i)\|^2 \quad (8)$$

平均形状的面部地标点之间的平方和, \bar{x}_j 的, 和那些扭曲的形状。 $v \setminus$ 类似的定义。形式上, 每个分割都是一个包含3个参数 $e = (\tau, u, v)$ 的决策, 并应用于每个训练和测试示例为

$$h(I_{\pi_i}, \hat{S}_i^{(t)}, \theta) = \begin{cases} 1 & I_{\pi_i}(u') - I_{\pi_i}(v') > \tau \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

其中 $u \setminus$ 和 $v \setminus$ 是使用最佳扭曲的尺度和旋转矩阵来定义的吗 $\hat{S}_i^{(t)} \setminus S$ 根据等式(7)得出。

在实践中, 作业和本地翻译是在培训阶段确定的。计算相似度变换, 在测试时, 这个过程中计算成本最高的部分, 在级联的每个级别上只做一次。

2.3.2 选择节点分割

对于每棵回归树，我们用一个分段常数函数来近似底层函数，其中一个常数向量拟合于每个叶节点。为了训练回归树，我们在每个节点上随机生成一组候选分割，即 e 。*然后我们贪婪地从这些候选项中选择 e ，它使平方误差的和最小。如果 Q 是一个节点上的训练示例的索引的集合，这对应于最小化

$$E(Q, e) = \sum_{s \in \{l, r\}} \sum_{i \in Q_{e,s}} |r_i - \mu_{e,s}|^2 \quad (10)$$

其中 $Q_{e,l}$ 是由 e, r 诱导的决策而发送到左节点的例子的指数 i 是在梯度提升算法中为图像 i 计算的所有残差的向量，和

$$\mu_{e,s} = \frac{1}{|Q_{e,s}|} \sum_{i \in Q_{e,s}} r_i, \text{ 对于 } s \in \{l, r\} \quad (11)$$

最优分割可以非常有效地找到，因为如果一个人重新排列方程（10），并省略了不依赖于 e 的因素，那么我们就可以看到

$$\arg \min_e E(Q, e) = \arg \max_{s \in \{l, r\}} \sum_{i \in Q_{e,s}} |Q_{e,s} - \mu_{e,s}|^2$$

这里我们只需要计算 $\mu_{e,l}$ 当计算不同的 e 时，如 $\mu_{e,r}$ 可以从父节点上的目标的平均值中计算出来吗 $\mu_{e,l}$ 列举如下

$$\mu_{e,r} = \frac{|Q| \mu - |Q_{e,l}| \mu_{e,l}}{|Q_{e,r}|}$$

3.3.2 功能选择

每个节点的决策是基于一对像素的强度值差的阈值。这是一个相当简单的测试，但它比单一强度阈值更强大，因为它对全局照明的变化相对不敏感。不幸的是，使用像素差异的缺点是潜在的分割（特征）候选的数量在平均图像的像素数量是二次的。这使得如果不搜索大量的 e ，就很难找到好的 e 。然而，通过考虑到图像数据的结构，可以在一定程度上缓解这种限制因素。我们引入了一个指数先验

$$P(u, v) \propto e^{-\lambda |u - v|} \quad (12)$$

在分割中使用的像素之间的距离上，以鼓励选择更近的像素对。

我们发现，使用这个简单的先验可以减少对许多人脸数据集的预测误差。图4比较了有和没有此之前的所选择的特性，在这两种情况下，特性池的大小都固定为20。

2.4. 处理丢失的标签

方程（10）的目标可以很容易地扩展到处理在一些训练图像中没有标记一些地标的情况（或者我们对每个地标有一个不确定性的度量）。引入变量 $w_{i,j}$ 每个训练图像 i 和每个地标 j 均为2个 $[0, 1]$ 。设置 $w_{i,j} = 0$ 表示在第 i 张图像中没有标记，设置为1表示标记。然后，公式（10）可以更新为

$$E(Q, e) = \sum_{s \in \{l, r\}} \sum_{i \in Q_{e,s}} (r_i - \mu_{e,s})^T W_i (r_i - \mu_{e,s})$$

其中 W_i 是一个带有向量 $(w_{i1}, w_{i1}, w_{i2}, w_{i2}, \dots, w_{ip}, w_{ip})^T$ 对角线和

$$\mu_{\theta,s} = \left(\sum_{i \in Q_{\theta,s}} W_i \right)^{-1} \sum_{i \in Q_{\theta,s}} W_i r_i, \text{ for } s \in \{l, r\} \quad (13)$$

梯度增强算法也必须进行修改，以考虑到这些权重因子。这可以简单地用目标的加权平均值初始化集成模型，并将回归树拟合到算法1中的加权残差中，如下

$$r_{ik} = W_i (S \Delta_i^{(t)} f_k - 1(I_{T_i}, \hat{S}_i^{(t)})) \quad (14)$$

3. 实验

基线：为了准确地基准测试我们提出的方法的性能，我们使用回归树集合（ERT）创建了另外两个基线。第一种是基于随机特征选择（EF），另一种是更高级的基于相关性的特征选择（EF+CB），这是我们对[2]的重新实现。所有这三种方法的所有参数都是固定的。

EF使用了一个简单的实现的随机蕨类植物作为集合中的弱回归变量，是最快的训练。我们使用与[2]建议的相同的收缩方法来规范蕨类植物。

EF+CB使用一种基于相关性的特征选择方法来投射目标输出， r_{is} ，到一个随机方向 w ，并选择特征对 $(u, v) \in \{1, \dots, t\} \times \{1, \dots, I_i(u)\} \times \{1, \dots, I_i(v)\}$ 的训练数据与投影目标 w 的样本相关性最高 T_{ri} 。

参数：除非有规定，否则所有实验均采用以下固定参数设置进行。强回归变量的数量， r_t ，在级联中是 $T = 10$ 和每个 r_t 由 $K = 500$ 个弱回归变量 g 组成 k 。用来表示 g 的树木（或蕨类植物）的深度 k 设置为 $F = 5$ 。在级联 P 的每个级别上，从图像中采样400像素的位置。为了训练弱回归变量，我们随机抽取一对 P 像素位置

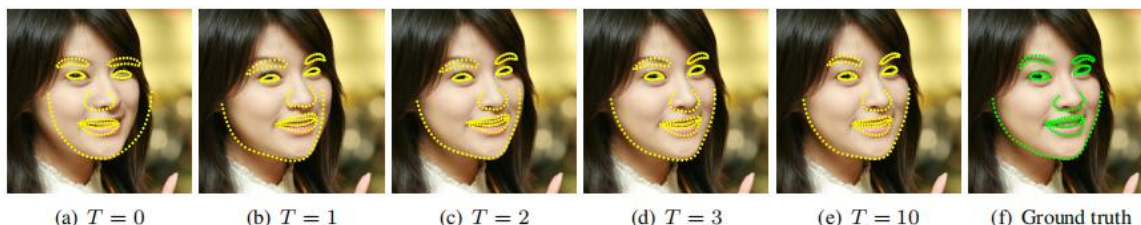


图2。以基本Viola & Jones[17]面探测器输出为中心的平均形状初始化的级联的不同级别的地标估计。在级联的第一级之后，误差已经大大减少了。

根据我们的先验，并选择一个随机阈值来创建一个潜在的分裂，如公式(9)所述。然后，通过重复这个过程 $S = 20$ 次，并选择一个能够优化我们的目标的过程，可以找到最佳的分裂。为了创建训练数据来学习我们的模型，我们为每个训练示例使用 $R = 20$ 个不同的初始化。

性能：该算法在单个图像上的运行时复杂度为常数 $O(1)$ (TKF)。训练时间的复杂性与训练数据的数量 $O(NDKFS)$ 呈线性关系，其中 N 为训练数据的数量， D 为目标的维数。在实践中，对于一个CPU，我们的算法大约需要一个小时来训练HELEN[12]数据集，并且在运行时每张图像只需要大约1毫秒。

数据库：报告的大多数实验结果都是针对HELEN[12]人脸数据库，我们发现这是最具挑战性的公开数据集。它由2330张图片组成，每一幅都有194个地标。根据作者的建议，我们使用2000张图片进行训练数据，其余的图片进行测试。

我们还报告了流行的LFPW[1]数据库的最终结果，该数据库包含1432张图片。不幸的是，我们只能下载778张训练图像和216张有效的测试图像，这使得我们的结果不能直接与之前在这个数据集上报告的结果相比较。

比较：表1是对我们与以前的算法相比的结果的总结。除了我们的基线之外，我们还将我们的结果与两种不同的活动形状模型，STASM[14]和CompASM[12]进行了比较。

	[14]	[12]	ef	EF+CB	EF+CB (5)	EF+CB (10)	电阻温度计
错误.	.111	.091	.069	.062	.059	.055	.049

表1。总结了不同算法的计算结果
海伦数据集。误差是每个地标到其地面真实位置的平均归一化距离。这些距离通过除以眼间距离来归一化。括号内的数字表示使用随机初始化运行回归算法的次数。如果没有显示任何数字，则用平均形状初始化该方法。在多重估计的情况下，估计的中值被选择作为地标的最终估计。

本文中描述的回归树集成显著地改进了结果

羊齿植物图3显示了在级联的不同级别上的平均误差，这表明ERT可以比其他基线更快地减少误差。请注意，我们还提供了多次运行EF+CB并取最终预测的中位数的结果。结果表明，该方法可以少计算一个数量级，可以获得与EF+CB相似的错误率。

我们还提供了被广泛使用的LFPW[1]数据集的结果（表2）。使用我们的EF+CB基线，我们无法复制[2]报告的数字。（这可能是因为我们无法获得整个数据集。）然而，我们的方法超过了之前在这个数据集上报告的大多数结果，只花费了任何其他方法所需的计算时间的一小部分。

	[1]	[2]	ef	EF+CB	EF+CB (5)	EF+CB (10)	电阻温度计
错误.	.040	.034	.051	.046	.043	.041	.038

表2。当应用于LFPW数据集时的不同方法的比较。有关这些数字的说明，请参见表1的标题说明。

特征选择：表4显示了使用公式(12)作为先验对在分割中使用的像素之间的距离的影响，而不是对最终结果的统一先验。参数 λ 决定了我们的特征中两个像素之间的有效最大距离，并在我们的实验中设置为0.1。在学习每个强回归因子 r 时，通过交叉验证来选择这个参数 t ，在级联中可能会导致更显著的改善。图4是在使用不同的先验时所选择的特征对的可视化。

统一指数
错误. 053. 049

表3。使用不同的先验进行特征选择对最终平均误差的影响。对定义一个特征的两个像素之间的欧氏距离应用指数先验，见式(12)。

正则化：在使用梯度增强算法时，需要小心地避免过拟合。为了获得较低的测试误差，需要进行某种形式的正则化。最简单的方法是收缩。这个

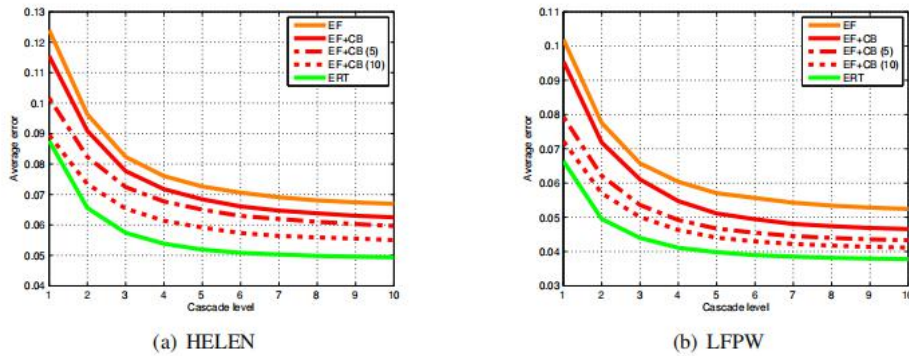


图3. 在HELEN (a)和LFPW (b)数据集上进行了不同方法的比较。EF是随机蕨类植物的集合，EF+CB是基于平均形状初始化的相关性特征选择的蕨类植物的集合。我们还提供了取[2]所建议的各种初始化（5和10）结果的中位数的结果。结果表明，所提出的仅以平均形状初始化的回归树集合（ERT）始终优于蕨类植物基线集合，并且可以在较少的计算量下达到相同的错误率。

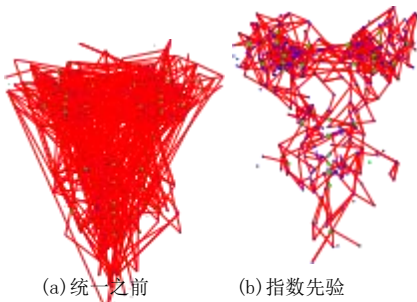


图4. 如果使用了不同的先验，则会选择不同的特征。指数先验使选择偏向于更接近的像素对。

涉及到将梯度增强算法中的学习速率 v 设置为小于1（这里我们设置了 $v = 0.1$ ）。正则化也可以通过平均多元回归树的预测来实现。这种方式， g_k 对应于一个随机的森林，而不是一棵树，我们设置了 $v = 1$ 。因此，在梯度增强算法的每次迭代中，我们不是拟合一个回归树，而是拟合多棵树（实验中为10棵），并对结果进行平均。（在所有的情况下，树的总数都是固定的。）

在偏差和方差权衡方面，梯度增强算法总是减少偏差，但增加方差。但通过收缩或平均的正则化方法通过学习多个重叠模型有效地减少了方差。

非正则化	收缩力	平均数
错误. 103	. 049	. 049

表4. 当应用不同形式的正则化时，在海伦数据集上的结果的比较。在相同的树总数下，当使用收缩或平均时，我们发现了类似的结果。

与更标准的收缩方法相比，我们使用平均正则化方法得到了类似的结果。然而，平均正则化具有更具可伸缩性的优点，因为它能够在训练期间实现并行化，这对于解决大规模问题尤为重要。

级联：在级联的每个级别的第二级

回归变量只能观察形状索引特征的固定和稀疏子集。基于当前估计对特征进行索引是一种简单的成本扭曲图像的粗糙方法。表5显示了使用和不使用级联时的最终错误率。我们发现使用这个迭代机制有了显著的改进，这之前报告的结果[8, 2]一致（为了进行公平的比较，我们将观察到的特征的总数固定在10 400个点）。×

#树	1	500	1	5000	10	500	×	×	×
错误.	085.	074.	049						

表5. 上述结果表明，使用级联回归变量相对于单一水平集成的重要性。

训练数据：为了测试我们的方法在训练图像数量方面的性能，我们从不同大小的训练数据的子集中训练不同的模型。表6总结了最终的结果，图5是级联中每一层的误差图。当我们有大量的训练示例时，使用许多层次的回归变量是最有用的。

我们重复了相同的实验，将增强样本的总数固定下来，但改变了用于从一个标记的面部示例生成训练示例的初始形状的组合和用于学习级联的注释图像的数量（表7）。

使用不同的初始值来扩充训练数据

#示例	100	200	500	1000	2000
错误	.090	.074	.059	.054	.049

表6. 关于训练例子的数量的最终错误率。当创建用于学习级联回归器的训练数据时，每个标记的人脸图像通过使用20个不同标记的人脸作为人脸形状的初始猜测而生成20个训练示例。

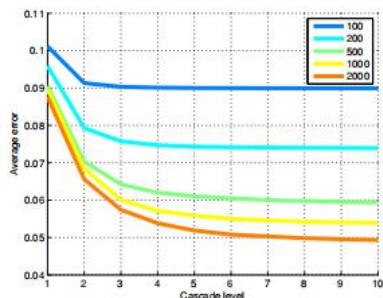


图5. 级联的每个级别的平均误差与使用的训练示例的数量相关绘制。当训练示例的数量很大时，使用许多级别的回归变量是最有用的。

#初始形状	100	200	500	1000	2000
	400	200	80	40	20
错误	.062	.057	.054	.052	.049

表7. 这里有效的训练例子的数量是固定的，但是我们使用不同的训练图像的数量和每个标记的人脸图像使用的初始形状的数量组合。

形状会根据形状来扩展数据集。我们的结果表明，这种类型的增强并不能完全弥补注释训练图像的缺乏。虽然通过增加训练图像的数量而获得的改进速度在前几百张图像后迅速减慢。

部分注释：表8显示了使用部分注释数据的结果。200个训练示例被完全注释，其余的只是部分注释。

#示例	200	200+1800 (25%)	200+1800 (50%)	2000
错误	.074	.067	.061	.049

表8. 使用部分标记数据的结果。200个例子总是有完整的注释。括号内的值表示观察到的地标的百分比。

结果表明，使用部分标记的数据可以得到较大的改进。然而，所显示的改进可能并不饱和，因为我们知道形状参数的潜在维度远低于地标参数的维度（194 2年）。×因此，通过明确利用地标位置之间的相关性，部分标签有可能有更显著的改进。注意，gra-

本文所描述的深度推进过程没有利用地标之间的相关性。这个问题可以在今后的工作中得到解决。

4. 结论

我们描述了如何使用回归树的集合从输入图像中提取的强度值的稀疏子集中回归出面部地标的位置。与以前的工作相比，所提出的框架在减少错误方面更快，也可以处理部分或不确定的标签。虽然我们的算法的主要组成部分将不同的目标维度视为自变量，但这项工作的一个自然扩展将是利用形状参数的相关性来进行更有效的训练和更好地使用部分标签。

致谢：这项工作已由瑞典战略研究基金会在VINST项目中获得资助。

参考文献

- [1] P. N. Belhumeur, D. W. 雅各布斯, D. J. 克里格曼和N. 库马尔. 使用范例的共识来定位面孔的某些部分. 在CVPR中, 第545-552页, 2011年. 1, 5
- [2] X. 曹, Y. 魏, F. 温和J. 太阳通过显式形状回归进行的人脸对齐. 在CVPR中, 第2887-2894页, 2012页. 1, 2, 3, 4, 5, 6
- [3] T. F. 饼干, M. Ionita, C. 林德纳和P. 索尔基于随机森林回归投票的鲁棒性和精确性的形状模型拟合. 在ECCV, 2012年. 1
- [4] T. F. Cootes, J. 泰勒, D. H. 库珀和J. 全麦的主动形状模型-它们的训练和应用. *计算机视觉和图像理解*, 61(1): 38-59, 1995. 1, 2
- [5] D. 克里斯蒂纳斯和T. F. 库茨增强的回归活动形状模型. 在BMVC中, 第79页. 1 - 79. 10, 2007. 1
- [6] M. 丹通, J. Gall, 法内利和L. V. 羊毛. 利用条件回归森林进行实时面部特征检测. 在CVPR, 2012年. 1
- [7] L. 丁和A. M. 马特内斯. 精确的面部特征. 在CVPR, 2008年. 1
- [8] P. Dollr, P. 焊接机和P. 佩罗纳. 级联姿态回归. 在CVPR, 第1078-108, 20105页. 1, 2, 6
- [9] G. J. 爱德华兹, T. F. coote和C. J. 泰勒主动外观模型的进展. 在ICCV中, 第137-142页, 1999页. 1, 2
- [10] T. 哈斯蒂, R. 提布希拉尼和J. H. 弗里德曼 *统计学习的要素: 数据挖掘、推理和预测*. 纽约: 施普林格-弗拉格, 2001年. 2, 3
- [11] V. 卡泽米和J. 苏里文使用基于零件的建模进行面对齐. 在BMVC中, 第27页. 1 - 27. 10, 2011. 2
- [12] V. 勒, J. 勃兰特, Z. 林, L. D. Bourdev和T. S. 黄交互式面部特征定位. 在ECCV中, 第679-692页, 2012页. 5
- [13] L. 梁, R. 肖, F. 温和J. 太阳通过基于组件的识别搜索进行的人脸对齐. 在ECCV中, 第72-85页, 2008年. 1

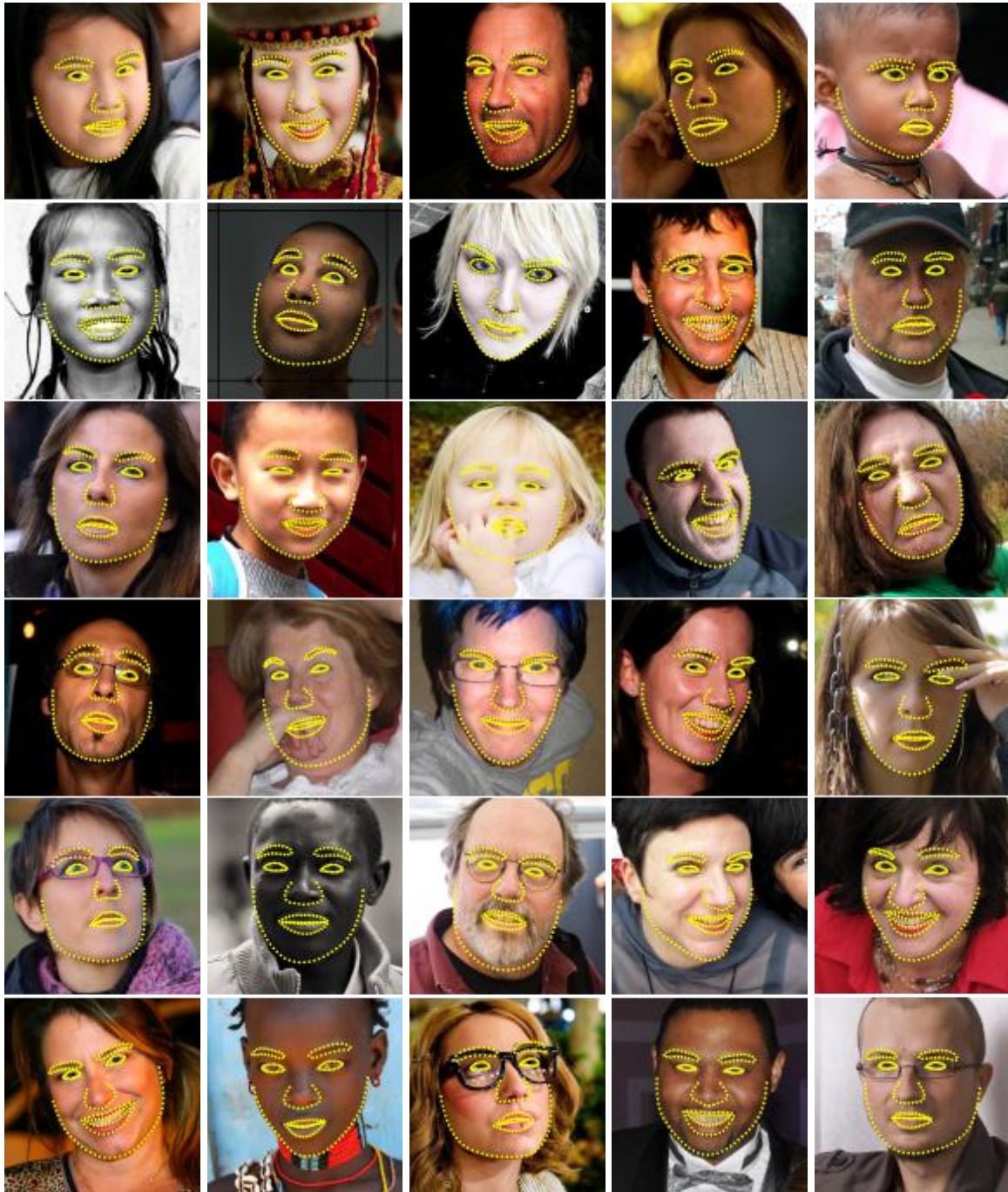


图6. 关于海伦数据库的最终结果。

- [14] S. 米尔博尔和F. 尼科尔斯取自父名用一个扩展的主动形状模型来定位面部特征。在ECCV, 第504-513页, 2008. [5](#)
- [15] J. 萨拉吉, S. Lucey和J. 科恩。通过正则化的地标均值位移来拟合可变形的模型。《国际计算机视觉杂志》, 91: 200-215, 2010. [1](#)
- [16] B. M. 史密斯和L. 张。具有非参数形状模型的关节面对齐。在ECCV中, 2012年, 第43-56页。 [1](#)
- [17] P. A. Viola和M. J. 琼斯健壮的实时人脸检测。在ICCV中, 第747页, 2001年。 [5](#)
- [18] X. 赵, X. 柴和S. 掸人关节面对准: 通过规范的重新拟合来修复不良的对准。在ECCV, 2012年。 [1](#)
- [19] X. 朱和D. 拉马南。在野外的人脸检测、姿态估计和地标定位。在CVPR中, 第2879-2886页。 [1](#)