

第五次作业报告

——人脸比对实验报告
SA20218084 林睿江

一. 实验目的

判断两张人脸图片是否为同一人，用 0/1 表示并输出 txt 文件

二. 实验思想

为了实现人脸比对判断，主要思想是通过一系列处理获取人脸图片尽可能地调正、对齐、裁剪，再提取特征进行比对判断。实验的主要步骤可以分为 1. 框出人脸；2. 利用人脸的位置信息确定 68 个特征点的位置；3. 利用 68 个特征点将人脸图片调正并且截取对齐；4. 在 3 的基础上再提取特征，训练样本，做出判断。

三. 实验步骤与核心思路

1. 人脸检测与框出

核心思路：

Viola-Jones 算法是基于人脸的 haar 特征值进行人脸检测的方法，整体框架分成三个部分：

I. 提取目标的特征，实验中提取的是 haar 特征；

II. 分类器学习的过程，训练弱分类器成为最优弱分类器，多个最优弱分类器组合形成强分类器；

III. 强分类器进行级联，提高检测的速度和准确率；

haar 特征：

haar 特征仅与一个矩形区域内像素取值的和有关，其特征取值为白色区域对应图像的像素值之和减去黑色区域对应图像的像素值之和。人

脸的 harr 特征应用如下图所示：



为了降低计算复杂度引入积分图技术。积分图是一张与原图大小一致的图，其每个像素/点的取值为其原图左上角所有像素取值的和。

弱分类器的数学结构如下：

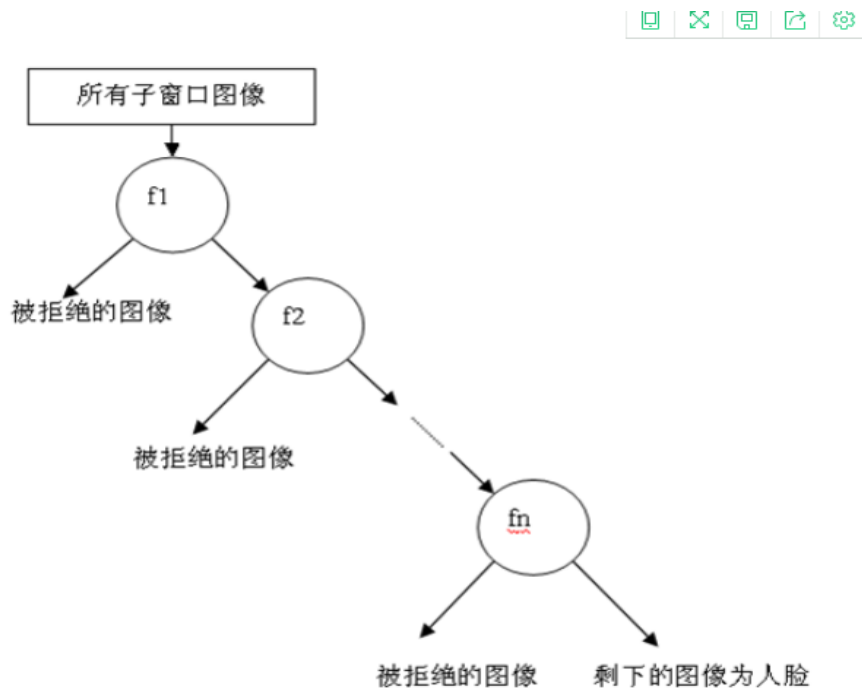
$$h(x, f, p, \theta) = \begin{cases} 1 & pf(x) < p\theta \\ 0 & \text{其他} \end{cases}$$

迭代并组合 T 个最优弱分类器得到强分类器，组合方式如下：

$$C(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{其他} \end{cases}$$

$$\alpha_t = \log \frac{1}{\beta_t}$$

最后通过级联方式将若干个强分类器由简单到复杂排列。



具体操作主要是经过方差归一化，伽玛校正‘硬’图像的预处理后，从类 haar 特征中训练弱分类器，使用 Adaboost 增强弱分类器和级联结构的人脸检测，在训练集训练模型后，再使用训练好的分类器进行人脸检测。

在实际的人脸检测过程中，由于训练集给出的图片背景较复杂，干扰因素较多，会造成无法检测出人脸或检测错误的情况，由于人脸框的获取对后面的实验进行十分重要，故也采取 matlab 库函数 Viola-Jones 进行人脸检测，以提高人脸检测的准确率。

2. 利用人脸的位置信息确定 68 个特征点的位置。(highlight)

核心思路：使用 matlab 复现论文《Face Alignment at 3000 FPS via Regressing Local Binary Features》Shaoqing Ren, Xudong Cao, Yichen Wei, Jian Sun; The IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), 2014, pp. 1685–1692, 这篇文章提出了一种高效、快速的人脸对齐算法，该算法在电脑上能达到 3000 FPS 的速度，在手机上也能达到 300 FPS 的速度，因此业界称其 face3000。算法的创新点在于局部二值特征以及训练这些局部二值特征的规则，即随机森林，连接所有根据随机森林得到的局部二值特征对线性回归器进行训练以得到最终结果。详细过程会在后面介绍。算法的总体概览如下：

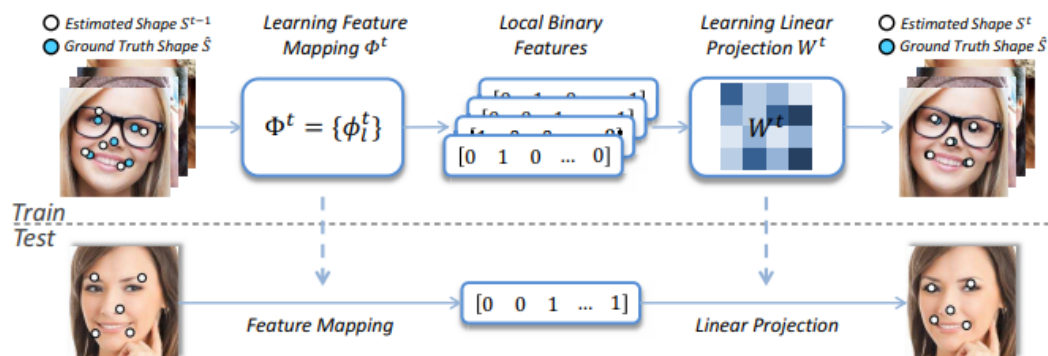


Figure 1. Overview of our approach. In the training phase, we begin by learning a feature mapping function $\Phi^t(I_i, S_i^{t-1})$ to generate local binary features. Given the features and target shape increments $\{\Delta S_i^t = S_i - S_i^{t-1}\}$, we learn a linear projection W^t by linear regression. In the testing phase, the shape increment is directly predicted and applied to update the current estimated shape.

给定一个初始形状 形状回归方法以级联的方式预测脸部形状 S 。

从一个初始的形状， S 通过一级一级估计形状增量进阶式改善的。

在一个通用形式下，一个形状增量在 t 级回归为：

$$\Delta S^t = W^t \Phi^t(I, S^{t-1})$$

其中 I 是输入图像，是来自于上一级的形状，是特征映射函数，是线性回归矩阵，应当注意的是依赖于 I 和 S ，在这种方式下学到的特征是指的是作为一个“形状编码”特征，通过将加入到，将回归带入到下一级。简单的说，该算法利用 random forest 和 shape

indexed features 得到 local binary features，然后利用 linear regression 求解 regressor，检测时利用训练得到的 feature mapping 和 linear projection 得到 update shape ΔS ，然后与上一个 stage 相加得到当前 stage 的 shape，之后不断迭代直到迭代结束。

在 LFPW 数据集上进行训练，并预先指定参数如下：

Param

绘图	变量	视图
1x1 struct 包含 15 个字段		
字段 ^	值	
<input checked="" type="checkbox"/> isparallel	1	
<input type="checkbox"/> augnumber_shift	0	
<input type="checkbox"/> augnumber_rotate	0	
<input type="checkbox"/> augnumber_scale	0	
<input type="checkbox"/> flipflag	0	
<input type="checkbox"/> ind_usedpts	1x68 double	
<input type="checkbox"/> augnumber	1	
<input type="checkbox"/> max_numfeats	[1000,1000,1000,500,500,500,400,400]	
<input type="checkbox"/> bagging_overlap	0.4000	
<input type="checkbox"/> max_raio_radius	[0.4000,0.3000,0.2000,0.1500,0.1200,0.100...	
<input type="checkbox"/> max_numtrees	5	
<input type="checkbox"/> max_depth	4	
<input type="checkbox"/> max_numthreshs	500	
<input type="checkbox"/> max_numstage	4	
<input type="checkbox"/> meanshape	68x2 double	

在 741 个训练人脸图象上面进行训练，结果保存在 “..\models\LBFRegModel.mat” 中，训练的结果在 224 个测试数据上可以达到 6.07 的均方误差。

```
extract local binary features...
时间已过 4.971598 秒。
predict landmark locations...
Mean Root Square Error for 224 Test Samples: 6.067964
时间已过 0.391725 秒。
```

对于给出的 LFW 数据集，由于未指定人脸的 ground truth，考

考虑使用 viola-jones 人脸检测算法，来标出人脸的 bounding box，并放入模型中预测出人脸的 68 个特征点。该方法依赖 bounding box 的选择，对于缺少 ground truth 的情况，人脸的 68 个特征点的均方误差为 51.8，在可以接受的范围内。（如图，红色为预测的 bbox，黄色为 ground truth，可见预测的 bbox 对于实际结果有较大影响）

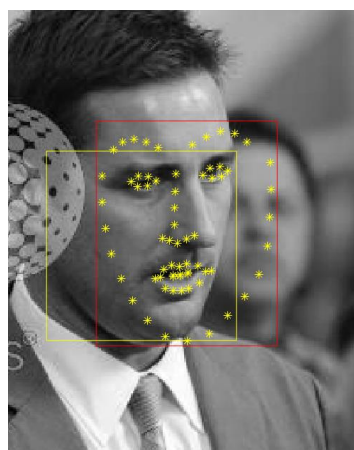
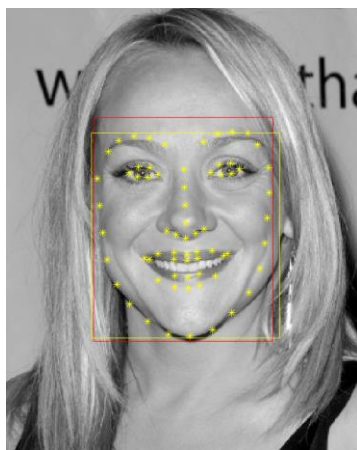
```
extract local binary features...
```

时间已过 5.833125 秒。

```
predict landmark locations...
```

Mean Root Square Error for 224 Test Samples: 51.817680

时间已过 0.371264 秒。



人脸 68 个特征点预测好了之后，就可以进行人脸对齐操作，即将人脸通过仿射变换映射到同一位置和尺度上，方便进行比较。

3. 人脸对齐和截取：

利用 68 个特征点将人脸图片调正并且截取对齐。

核心思路：利用 41 号和 47 号特征点（分别为左眼和右眼）计算出双眼中心，并以眼睛中心为轴进行旋转，旋转角度由双眼连线与水平的角度决定。

$$\text{angle} = \arctan\left(\frac{dy}{dx}\right)$$

计算出旋转矩阵 M （使用了 `cv2` 库函数下的 `getRotationMatrix2D` 函数），再将 M 使用到特征点和对应图像上获得对齐后的图像（`cv2` 的 `warpAffine` 函数）和特征点（根据旋转公式计算）。

截取脸部分，如果需要保证截取的图片大小相同的话，可以通过找到特征点横坐标和纵坐标的最小值 x_1, y_1 ，然后以其为标准取固定大小的图片，可以截取出脸部。程序中没有截取相同大小的图片，故直接截取 68 个特征点中最小值到最大值的部分。

这部分最后的结果，将所有图片进行处理后获得只包含脸部的对齐图片，方便后面进行特征提取和训练。

4. 特征提取和训练

在探索时尝试使用过 `matlab` 自带的 `CV toolbox` 进行特征提取和预测，但是在实际操作的过程中，来自同一个人的两张照片也无法匹配出一对相似的特征点。故选择其他的方法来计算两张图片的相似度。最后采用较为简单的灰度直方图特征。但是由于灰度直方图特征较为简单，且某些照片之间相似度较为一致。导致训练后的测试结果不太理想。

提取图片的灰度直方图特征，然后计算两张图片的灰度直方图的余弦相似度，保存下来作为训练数据。采用 Logistic Regression 的方法来训练判别器。训练数据为 $\{x_i, y_i\}$ 共 3200 组数据。训练出一个 w 用于预测。

四. 结果分析

将训练集和测试集按 15: 1 的比例分配。

模型的 precision=0.53, F1-score=0.43.

分析模型的准确率和召回率没有十分理想的原因主要在于:

1. 框脸算法的结果与希望得到的 ground truth 有一些差距, 特别在处理复杂背景等情况下的人脸图片时难以得到非常理想的结果;
2. 人脸位置确定 68 个特征点的算法准确率对框脸的 bbox 精确度有较大的要求。由于缺少完全吻合的 ground truth, 特性点标注不能完全准确;
3. 人脸对齐和裁剪并不能完全实现理想效果;
4. 尝试使用过 matlab 自带的 CV toolbox 进行特征提取和预测, 但是在实际操作的过程中, 来自同一个人的两张照片也无法匹配出一对相似的特征点。而采用较为简单的灰度直方图特征, 由于特征较为简单且某些照片之间相似度较为一致, 导致训练后的测试结果不太理想。

五. 参考资料

人脸检测部分参考资料:

<https://www.jianshu.com/p/024ad859c8de>

<https://blog.csdn.net/amazinguu/article/details/51505083>

68 个特征点提取部分参考资料:

《Face Alignment at 3000 FPS via Regressing Local Binary Features》Shaoqing Ren, Xudong Cao, Yichen Wei, Jian Sun; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1685-1692

人脸对齐部分参考资料:

<https://blog.csdn.net/u013713010/article/details/46047367>

<https://matthewearl.github.io/2015/07/28/switching-eds-with-python/>