

# 基于Hadoop平台的多模态人脸识别研究

李晓娜, 苏金善, 李瀚铭

伊犁师范大学, 新疆 伊宁

收稿日期: 2022年3月4日; 录用日期: 2022年4月4日; 发布日期: 2022年4月11日

## 摘要

随着现代人越来越喜欢用图像记录生活, 每日上传至互联网的图像都呈爆炸式增长。公安部门可以利用海量的图像数据实现网络追凶, 但现实中摄像头采集到的图像以及上传至网络的图像, 并非都是统一状态的图像, 而是包含各种状态的图像, 例如不同表情、不同动作、不同角度、不同程度的角度偏斜, 不同年龄, 以及有背景干扰的图像, 其中部分图像会因使用的设备不同, 导致上传至网络的图像既有灰白图像又有彩色图像。这些多模态图像为人脸的准确识别增加了难度, 要将实际中采集到的图像从如此复杂且规模庞大的数据集中匹配与识别出来, 是一件十分困难的事。针对以上问题, 提出将改进后的开源人脸识别库, 即face\_recognition库与Hadoop平台中的MapReduce进行结合, 在确保识别准确率的前提下提升人脸检测速度, 实现对大规模、多模态图像的有效识别。实验证明, 本文的方法能够有效解决大规模多模态图像的识别问题, 实时性高, 实用性强。

## 关键词

Hadoop, 多模态图像, face\_recognition库, AdaBoost算法, MapReduce

# Research on Multimodal Face Recognition Based on Hadoop Platform

Xiaona Li, Jinshan Su, Hanming Li

Yili Normal University, Yining Xinjiang

Received: Mar. 4<sup>th</sup>, 2022; accepted: Apr. 4<sup>th</sup>, 2022; published: Apr. 11<sup>th</sup>, 2022

## Abstract

The number of images uploaded to the Internet every day is exploding as modern people increasingly document their lives in images. Public security department can take advantage of the huge amounts of image data to realize the network after fierce, but in reality, the camera collected im-

ages and images uploaded to the Internet, are not all images of the united state, but contain images of the various states, such as different expressions and actions, different angles, different degree of skew Angle, different age, as well as the background interference image. Some of the images are uploaded to the network due to different devices, resulting in both gray and color images. These multi-modal images increase the difficulty of accurate face recognition, and it is very difficult to match and recognize the images collected in practice from such a complex and large data set. To solve the above problems, this paper proposes to combine the improved open source face\_recognition library, namely face\_recognition library, with MapReduce in Hadoop platform to improve face detection speed and realize effective recognition of large-scale and multi-modal images on the premise of ensuring recognition accuracy. Experimental results show that the proposed method can effectively solve the problem of large-scale multi-modal image recognition, with high real-time performance and practicability.

## Keywords

Hadoop, Multimodal Image, Face\_Recognition Library, AdaBoost Algorithm, MapReduce

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

现如今随着移动设备的普及与互联网的飞速发展,越来越多的人开始用图像来记录生活。合理利用这些图像数据将有助于构建更加和谐的社会。例如网络追凶,有的嫌疑人为了逃脱法律责任,常流窜多地,扩大公安干警的搜索范围,提升破案难度,也为社会的良好秩序增添了许多隐患。如果警方可以运用大数据人脸识别技术,将摄像头拍摄到的图像与地区数据库图像进行匹配,便可快速找到嫌疑人,提高办案效率,维护社会的良好治安。

在实际中,上传至互联网的图像不仅数据规模庞大,而且并不都是合适的人脸图像,这些图像是包含不同表情、不同姿态、不同光照以及有复杂背景干扰等影响因素的多模态图像,并且上传至网络数据库中的图像,一个人会有多张多模态图像。同样,摄像头抓拍到的人脸图像也存在多种角度、不同姿势、不同表情等情况,甚至由于设备不同采集到的图像既有可能是灰白图像,也可能是彩色图像。如何将实际场景中采集到的图像从多模态数据集中快速、准确识别出来,是本文研究的主要目的。

传统的单一服务器无法很好的处理大规模图像,而 Hadoop [1]作为目前最常用的大数据处理平台,它所搭建的集群只需要普通的计算机就可以实现,强大的并行处理能力非常适用于处理大规模图像。而且易于拓展、计算能力强、容错性高,价格低廉,非常适用广泛推广[2]。本文将 Hadoop 平台与传统的人脸识别技术相结合,可以让公安部门运用集群的并行数据处理能力,将摄像头采集到的多模态人脸图像与省或市的数据库多模态图像进行对比,只需很短的时间,就可以完成待识别图像与数据库图像的匹配工作。

## 2. 国内外研究现状

将 Hadoop 平台与人脸识别技术相结合,国内外也有一些研究人员对此进行了研究。Zhang B [3]对比了两种人脸识别的方法,第一种是将 Hadoop 中的 MapReduce 框架与支持向量机 SVM (Support Vector Machine)相结合,第二种将传统 LBP (Local Binary Patterns)算子和深度 LBP 算子进行融合,使其在人脸

识别过程中各自发挥优势,并且利用 BP (Back Propagation)和 RBF (Radial Basis Function)神经网络建立图像分类器进行识别,实验结果表明,Hadoop 与 SVM 分类器结合具有更好的人脸识别效果;Bogdanchikov A 等人[4]开发了一个新系统,将 Hadoop 与 HIPI 库结合,采集公共环境下的人脸图像,并对采集的图像应用人脸检测和识别算法,最终以可读的形式呈现结果;Villegas-Cortez J [5]等人使用了一种遗传算法来优化面部兴趣点,并将此种算法与基于内容的图像检索方法中的局部纹理分析技术相结合,运行在 54 个服务器搭建的 Hadoop 集群中,实现快速且准确的人脸识别;任静等人[6]在进行大规模人脸识别时,将离线处理与在线处理相结合。对大规模人脸图像预处理时进行离线处理,用户将待检测图像输入人脸识别系统进行识别时,采用在线处理;耿玉琴等人[7]针对传统人脸识别方法对处理大规模图像效率低的问题,提出将 PCA (Principal Component Analysis)算法和 Hadoop 平台中 MapReduce 框架相结合的新思路。首先对人脸图像使用 PCA 算法进行特征提取,之后使用 MapReduce 计算欧式距离[8],最后将识别出的内置图像信息进行存储。以上的方法虽然实现了大规模人脸识别,但是这些方法步骤较为复杂,识别效率较低,而且他们所采用的数据集图像涉及的实际场景中的图像较少,只能识别单一姿态变化的图像或者轻度姿态变化的图像。

本文提出了一种新的大规模人脸识别方法,将开源人脸识别库 face\_recognition [9] [10]库与 Hadoop 平台中的 MapReduce 框架相结合,face\_recognition 库中有着完整的人脸识别流程,并且对复杂环境中的人脸图像有着较好的识别准确率,大大简化了人脸识别过程,为了进一步提升人脸识别的效率,本文将 face\_recognition 库中的人脸检测算法 HOG 算法[11] (Histograms of Oriented Gradients)换成 AdaBoost 算法,该算法简单、检测速度快,对复杂环境中的图像检测准确率高,且具有较强的扩展性和稳定性。实验证明,本文的方法能够对大规模多模态图像进行有效识别,并且本文的方法提升了图像处理效率,实时性好、实用性强。

### 3. 人脸识别实现

#### 3.1. 改进 face\_recognition 库

face\_recognition 库是目前最简单的开源人脸识别库,具有完整的人脸识别功能,例如人脸检测、对齐、特征对比等。并且在开源数据集 LFW (Labeled Faces in the Wild)上有高达 99.38%的准确率,而 LFW 数据集中包含许多复杂场景中的图像,也含有各种状态的多模态图像。并且由 Python 语言编写,易于理解、维护和修改。将 face\_recognition 库与 MapReduce 相结合,在简化人脸识别过程的同时又保证了较高的识别准确率。

face\_recognition 库中的人脸检测算法采用的是 HOG 算法,该算法由于梯度的性质导致 HOG 算法对图像噪点敏感,且描述算子生成过程冗长,导致检测速度慢、实时性较差,在实际采集场景中,由于背景复杂、光照强度不同、拍摄角度不同、目标姿势变化等等因素的影响,导致人脸轮廓不清晰、损失部分人脸特征、人脸姿态扭曲等问题,导致 HOG 算法在检测人脸时检测不准。为了能够进一步提升人脸检测的效率和准确度,本文对 face\_recognition 库进行了改进,本文提出将 face\_recognition 库中的 HOG 人脸检测算法换成 AdaBoost 算法进行人脸检测,该方法检测准确率高、检测速度快、稳定性强,此算法相比其他人脸检测算法能更快的检测出人脸,也正是此方法的提出人脸检测技术才大量用于现实生活。

#### 3.2. 基于级联结构的 AdaBoost 分类器

AdaBoost 检测算法是利用扩展后的 Haar-like 特征[12]计算人脸图像的特征值,基于这些特征训练出弱分类器,之后在基于这些弱分类器训练出分类能力较强的强分类器,为了进一步提升检测的效率与准确率,再将强分类器通过串联的方式构成级联分类器,级联分类器能够更准确的检测出人脸。最初的

Haar-like 特征值的计算方式复杂且耗时长，后来提出的积分图法计算图像特征值，大大提升了人脸检测的速度。积分图如图 1 所示，其特征值计算定义如下：

已知检测图像  $I$ ，在点  $x, y$  处的积分图像值为：

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (1)$$

其中， $ii(x, y)$  是  $x, y$  处的积分图值， $i(x', y')$  是  $x', y'$  处的灰度值。

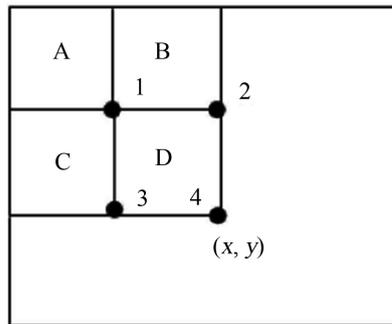


Figure 1. Integral figure  
图 1. 积分图

通过对积分图的使用，只需计算出点 (1,2,3,4) 处积分图的积分值，就可以将任何矩形特征的灰度和计算出来，大大加快了计算速度，提升了检测速率。计算公式如下：

$$Recsum(D) = ii(4) + ii(1) - ii(2) - ii(3) \quad (2)$$

其中， $Recsum(D)$  表示矩形区域  $D$  的像素灰度值的累加值。

由公式(2)可知，采用积分图法计算特征，只需计算 4 次积分图像的值，就能够得出矩形特征值，在提取图像特征值时，使用扩展后的矩形特征遍历整张图像，同时进行基础的加减运算即可完成特征值的计算。正是有了积分图法，才大大提升了特征值的计算速度，也加速了人脸检测的效率。之后用 AdaBoost 算法进行训练，训练过程如下所示：

1) 训练集  $X$  中共有  $m$  个样本，样本用  $y$  标注，若为人脸  $y = 1$ ，否则  $y = 0$ ，即

$$(x_1, y_1), \dots, (x_m, y_m), x_i \in X, y_i \in Y = \{0, 1\} \quad (3)$$

2) 初始化所有样本权重：

$$w_0(i) = \frac{1}{m} (i = 1, 2, \dots, m) \quad (4)$$

3) 挑选  $T$  个弱分类器需迭代  $T$  次，迭代次数  $t = 1, \dots, T$ ：

第  $t$  次循环中，第  $j$  个弱分类器的误差和

$$e_j = \sum_{i=1}^m w(j) |h_j(x_i) - y_i| \quad (5)$$

选择  $e_j$  值最小的弱分类器  $h_t(x)$ ；

计算弱分类器权重

$$\alpha_t = \log\left(\frac{1}{\beta_t}\right), \beta_t = \frac{e_t}{1 - e_t} \quad (6)$$

更新权重  $w_{t+1}(i) = w_t(i)\beta_t^{1-\theta_i}$  : 若样本  $i$  正确分类  $\theta_i = 0$ , 否则  $\theta_i = 1$  ;  
权重归一:

$$w_{t+1}(i) = \frac{w_t(i)}{\sum_{i=1}^m w_t(i)}, t = t+1 \tag{7}$$

4) 级联弱分类器形成强分类器:

$$h_x = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{其他} \end{cases} \tag{8}$$

其中,

$$\alpha_t = \log \frac{1}{\beta_t} = \log \frac{1-e_t}{e_t} = -\log e_t \tag{9}$$

强分类器对图像的检测仍然不够精准, 为了能进一步提升检测的精度, 将多个强分类器构成一个级联分类器, 级联分类中的前端强分类器可以快速排除多数非人脸区域, 进而加快检测速度, 后端的强分类器构造比较复杂, 用于对难识别的区域进行检测, 最终标记出人脸。图 2 为级联分类器的检测图。

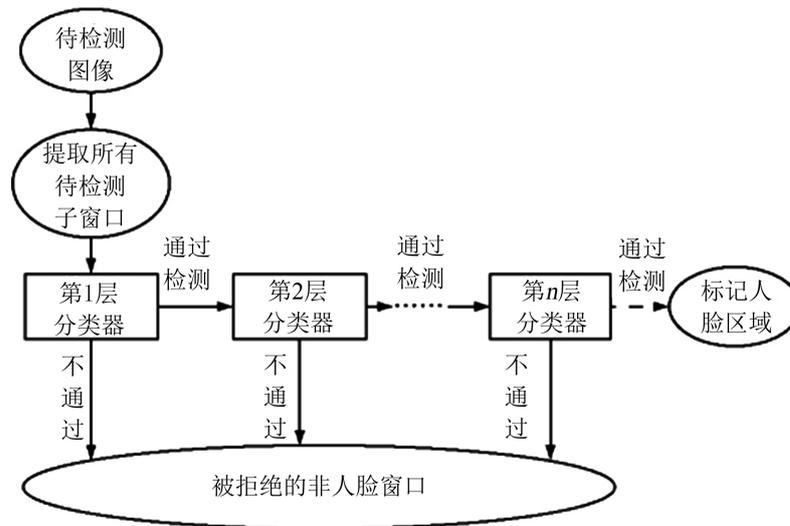


Figure 2. Face detection diagram of cascade classifier  
图 2. 级联分类器人脸检测图

### 3.3. 人脸对齐

检测到人脸后需要对人脸图像进行对齐, 因为检测到的多模态图像会存在不同程度的光照变化、姿势变化、复杂背景干扰等情况, 导致提取特征值时损失过多特征, 造成识别不准的问题。本文在人脸检测结束后进行人脸对齐, 便于特征提取时提取到尽可能多的特征。通过仿射变换对人脸进行对齐, 便于特征值的提取, 增加识别的准确度。

仿射变换是通过整合平移, 缩放, 翻转, 旋转和剪切一系列操作来实现的。在数学上可以用  $3 \times 3$  的矩阵  $M$  来表示, 其最后一行为  $(0,0,1)$ 。将原图像的点  $(x,y)$  经过仿射变换后变为新图像中的点  $(x',y')$ , 数学表达式如下:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & m_{12} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (10)$$

人脸对齐就是通过对人脸图像进行仿射变换来实现，人脸经过变换后仍保持特征点相对性不变的原则，例如将原图像以  $(x, y)$  为轴心顺时针旋转  $\theta$  弧度来完成旋转变换。其变换矩阵  $M$  为：

$$\begin{bmatrix} \cos \theta & -\sin \theta & x - x \cdot \cos \theta + y \cdot \sin \theta \\ \sin \theta & \cos \theta & y - x \cdot \sin \theta - y \cdot \cos \theta \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

相当于两次平移与一次原点旋转变换的复合，即先将轴心  $(x, y)$  移到原点，然后做旋转变换，最后将图片的左上角置为图片的原点，即：

$$M = \begin{bmatrix} 1 & 0 & -x \\ 0 & -1 & -y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & x \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

### 3.4. 提取特征值

本文的特征值提取由 Base64 编码实现，依据人脸对齐的关键点进行编码，将图像转成文本进行传输，传输图像数据根据 RFC4648 的定义，Base64 编码使用 “a~z”，“A~Z”，“0~9”，“+”，“/”，64 个 ASCII 字符及特殊字符后缀 “=” 来进行编码，Base64 以 3 个字节(24 位)为单位进行编码，6 位一组生成一个索引号，再通过 Base64 字母表查询该索引号，输出 4 个编码后的字符，针对不同类型的图片 Base64 也会生成不同的编码，这些编码也是人脸图像的面部特征值。

使用编码的方式既可以保护图像数据的信息安全，又使图像在上传时受到的网络影响较小。此外，Hadoop 平台实际上更适合文本传输，本文将图像转成文本进行传输，在一定程度上也加快了数据的读取速度。之后，我们将提取到的特征值，放入 Hadoop 平台中的 HDFS 中进行存储，由 MapReduce 框架完成待识别图像与内置数据库的匹配、识别。

## 4. MapReduce 运行

在 Hadoop 平台中，HDFS 实现了大规模数据的高效存储，而 MapReduce 则实现了大规模数据的高效计算。MapReduce 框架在普通服务器上就可以部署，容错性强、易编写、易扩展性、以及具有高吞吐量[13]。MapReduce 的关键程序就是 Map 和 Reduce 两部分，用户只需对这两部分进行编写即可，至于其他内容，如任务调度、错误处理、负载均衡等部分，都由 MapReduce 框架自动进行处理。MapReduce 的运行原理图如图 3 所示，其大体流程是先从 HDFS 中获得输入数据，再由 Map 任务对获取的数据进行切片分割，最后由 Reduce 汇总 Map 任务的结果并传回 HDFS。

在 MapReduce 框架中，首先对 HDFS 中读取的图像特征库进行数据切片分割，这里的切片不是实际意义上的切片，而是对待处理图像数据的路径、ID 号以及图像对应的编码进行记录，这里的编码就是图像的特征值。Map 读取图像数据进行切片，将切片后的数据分解为 Key/Value 形式的数据特征，其中键值 Key 表示图像特征在切片分割中的位置，键值 value 为图像编码形式的特征值。然后 map ( ) 函数调用这些 Key/Value 值，判断图像特征是否满足匹配条件，并将匹配的中间结果以 Key/Value 键值对的形式保存在本地系统中，其中 Key 值表示相似度，Value 值表示图像特征库中对应的内置图像。接着，对 Map 任务输出的中间结果进行整理，将整理好的中间结果传给 Reduce 任务。最后，Reduce 任务接收到 Map 输出的中间结果按照相似度大小对结果进行排序，并将排序好的结果存储于 HDFS 中。

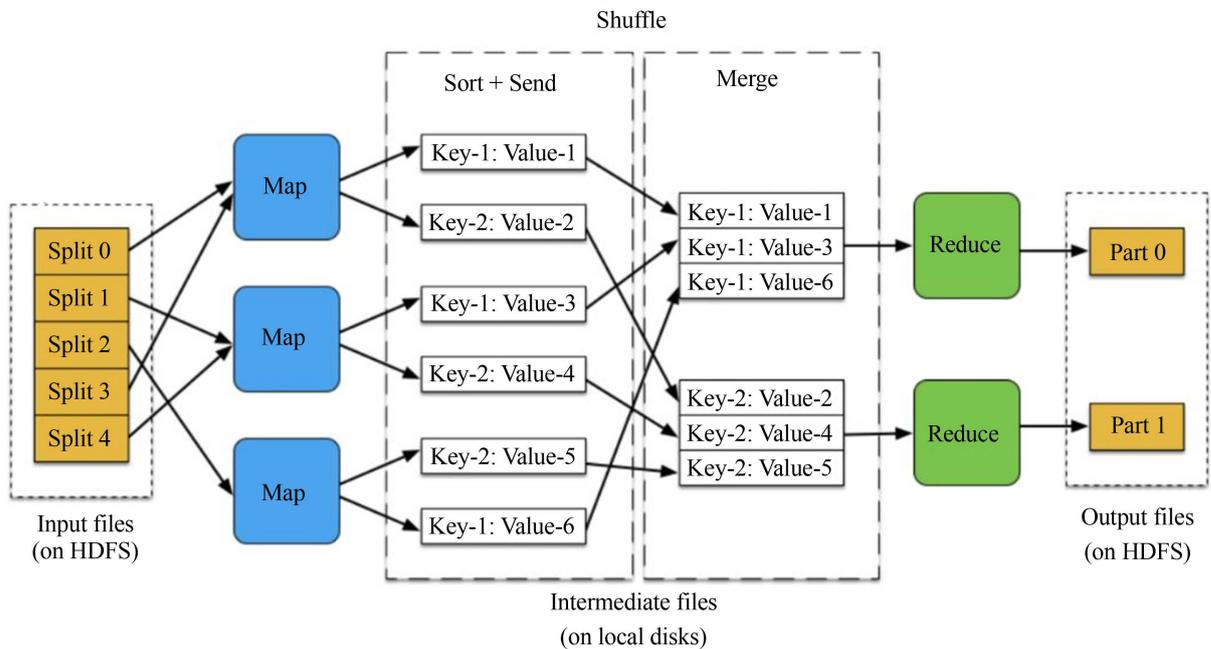


Figure 3. MapReduce operating principle diagram  
图 3. MapReduce 运行原理图

### 5. 实验结果与分析

本文在实验室一台 24 G 内存、GTX 1080Ti 处理器上安装一个 64 位的虚拟机 VMware Workstation，建立一个的小规模 Hadoop 集群系统，模拟多台机器同时工作的并行模式，也就是一台机器上运行多个系统，由于实验室条件有限，无法搭建真正的集群环境，因此本文的集群由三个节点构成，即一个 Master 主节点和两个 Slave 从节点构成。

#### 5.1. 实验环境

本次实验的硬件环境如表 1 所示，单机与集群的环境配置分别如表 2 和表 3 所示。

Table 1. Hardware environment  
表 1. 硬件环境

项	硬件配置
GPU	GeForce GTX 1080Ti
CPU	Intel XeonE5-2690
内存大小	24 GB
硬盘	1TSSD

Table 2. Environment configuration for a single machine  
表 2. 单机的环境配置

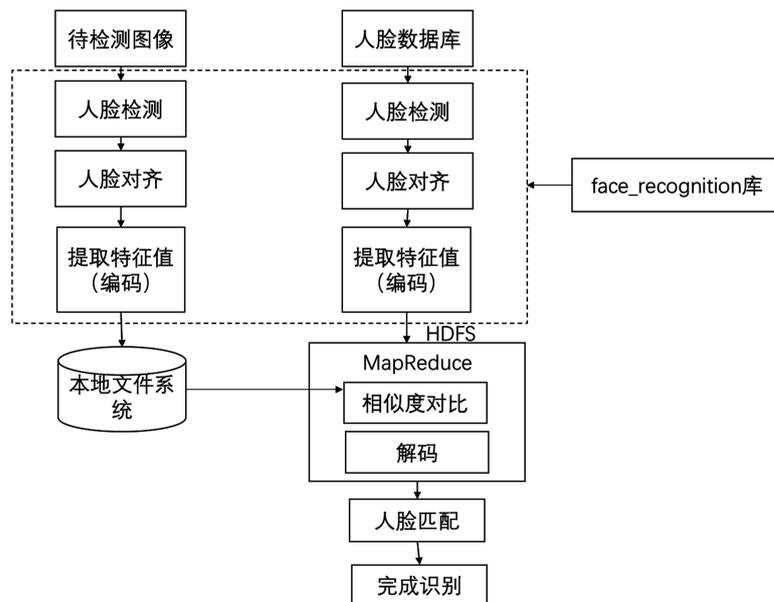
项	环境配置
操作系统	Windows10
使用的库	Opencv
实验对象	CelebA 数据集

**Table 3.** Environment configuration for the cluster  
**表 3.** 集群的环境配置

项	环境配置
操作系统	CentOS7
开发环境	Hadoop 平台
实验对象	CelebA 数据集
编程语言	Python3.6

### 5.2. 实验结果与分析

本文所提出的方案是将改进后的 face\_recognition 库与 Hadoop 平台下的核心子项目 MapReduce 相结合，解决复杂环境下的大规模多模态图像识别的问题。系统流程图如图 4 所示，主要是将采集到的待检测图像从大规模数据集中找出所有相同人脸，并进行识别。本实验所使用的数据库图像主要来自 CelebA 数据集，数据集中的每张图像都有对应的具体信息。先随机抽取 600 图像构成本实验所使用的数据集，进行一个小型实验，数据集包含不同光照强度、不同表情、不同面部角度、多种行为姿态、灰白图像以及有复杂背景干扰的多模态图像。由于现实中采集到的人脸图像不一定是良好环境下的正脸图像，也可能是含有不同光照、不同角度、不同表情、不同姿态等的多模态图像，因此待识别的图像也采用多模态图像。为了便于展示，本文只选取数据库中少量图像用于展示。



**Figure 4.** System flow chart  
**图 4.** 系统流程图

由图 5 可知，本文采用的图像数据集中包含有复杂背景干扰的图像，也有不同光照强度、灰白图、不同表情、不同姿态以及不同面部角度的图像。随机抽取不同角度、不同表情以及有复杂背景干扰的多模态图像作为待检测图像，比较符合实际场景中的采集情况。此外，由图 6 和图 7 实验结果可知，本文所提出的方案，不但可以实现“一对多”图像的识别，也可以实现“多对多”图像的识别，具有较大的实用性，在实际中可以大大提升警务人员抓捕嫌疑人的效率。

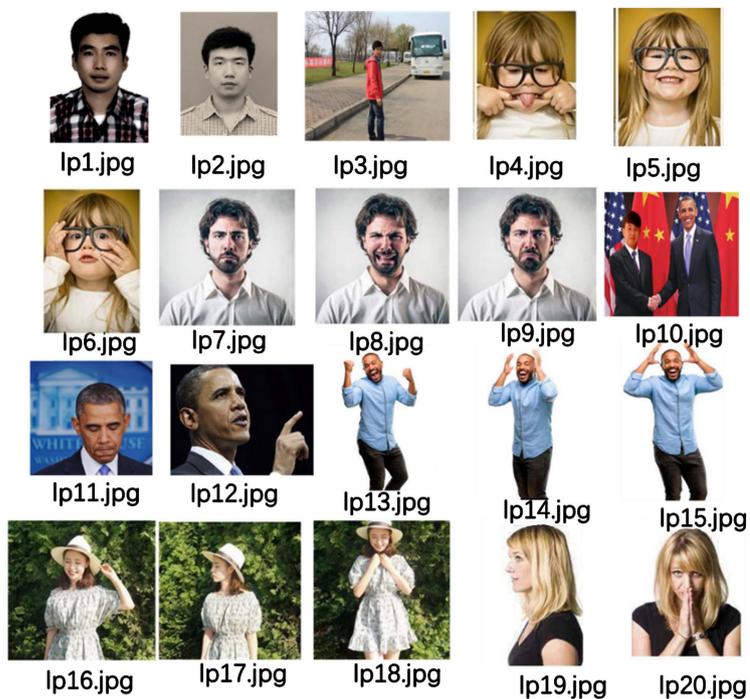


Figure 5. Part of the built-in image data set  
图 5. 部分内置图像数据集

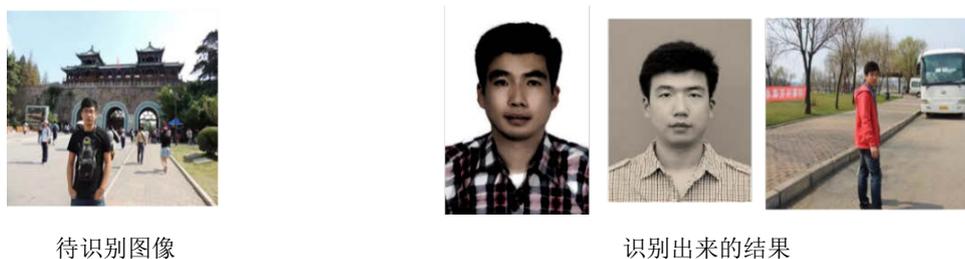


Figure 6. "One-to-many" multi-modal image recognition result graph  
图 6. "一对多"的多模态图像识别结果图

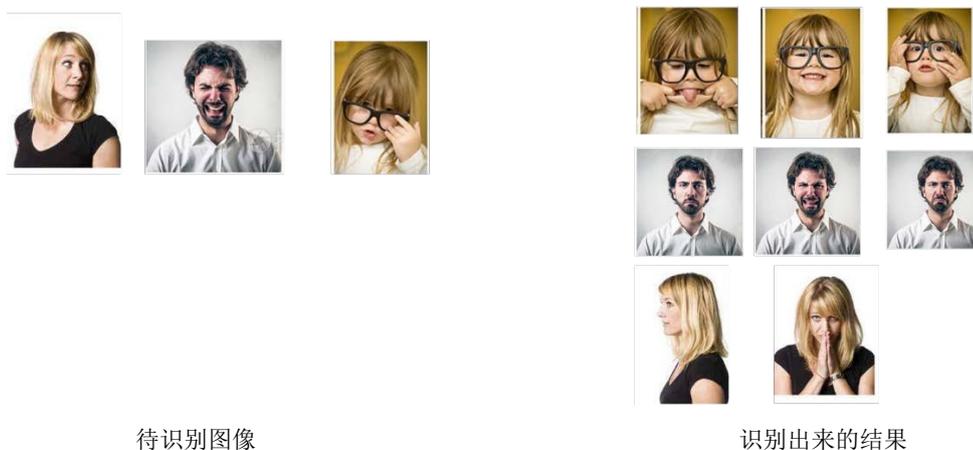


Figure 7. "Many-to-many" multimodal image recognition result graph  
图 7. "多对多"的多模态图像识别结果图

为了检查本文所提出方案的实时性，在其他条件都相同的前提下，将本文的算法在单机上进行一次实验，与集群在处理相同数量图像时的时间进行对比。对比图如图 8 所示。

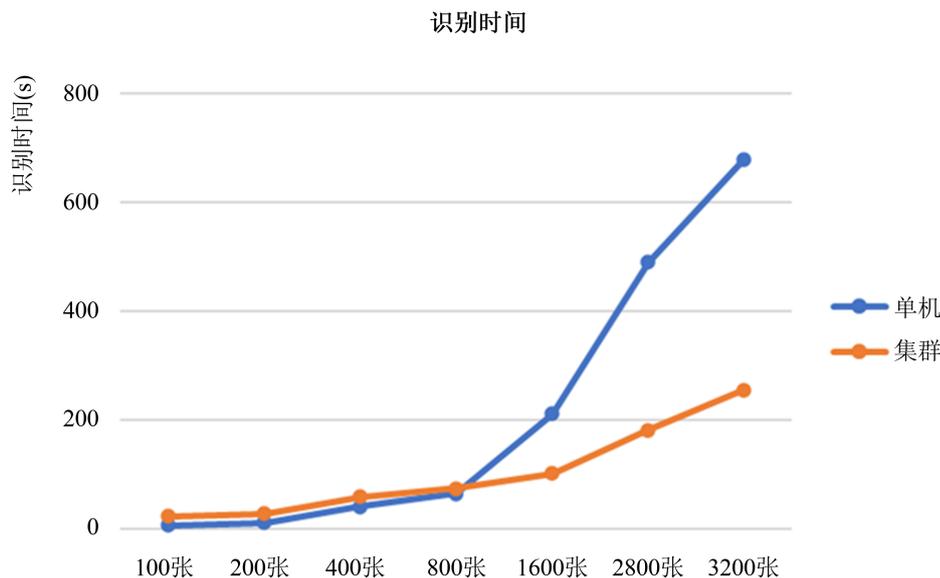


Figure 8. Comparison of single machine identification time and cluster identification time  
图 8. 单机与集群的识别时间对比图

由图 8 可以看出，当内置的数据集图片数据量低于 800 张时，单机运行的人脸识别时间少于集群识别的时间，但随着内置数据集的图像数量逐渐增多时，集群识别花费的时间更少。这是因为单机处理图像时，数据集存储在本地文件系统，且算法也在本地运行，因此少量级的图像数据处理速度更快。而当数据量增多时，单机操作系统负荷过重，其串行运行耗费时间较长。此时，集群的多节点并行处理能力，可以将数据量较多的数据集内的图像分配至多个节点并行处理，能够更快的实现大规模图像识别。

为了检测本文所提出的方法是否优于其他大规模人脸图像识别方法，将本文提出的方案与目前最为常用的大规模图像处理方法，也就是 Hadoop 与 PCA 结合方法进行对比，该方案在大规模图像处理中取得良好的识别效果。王沈括等人[14]使用数据集 ORL 与 FERET，将图像数据存储于 HDFS 中，提出使用 MapReduce 与 PCA 相结合的方法进行识别；耿玉琴等人使用 ORL 数据集将图像数据存储于 HBase 中，提出使用 MapReduce 与 PCA 相结合的方法进行识别。本文使用 CelebA 数据集，将图像数据存储于 HDFS 中，提出将改进的 face\_recognition 库与 MapReduce 相结合的方法。ORL 与 FERET 数据集都是含有多种姿态变化的灰度图像，而本文使用 CelebA 数据集中既有多种姿态变化的图像、又有彩色图像和灰度图像。虽然三种方法采用的数据集不同，但都是大规模图像识别中常用的数据集，且都有多种姿态变化的图像。

由图 9 可知，将图像数据存储于 HDFS 中的方法，识别时间少于存储于 HBase 中的方法。这是因为图像数据集存放在 HBase 中，加大了图像传输时的时间损耗，而数据存放在 HDFS 中，便于系统读取，减少了时间损耗。其次，由图 9 可知，两种数据存储于 HDFS 上的方法，在处理少量数据集时时间损耗相差不多，但随着图像数量的增多，本文的方法在图像处理中效率更高，这是因为改进后的 face\_recognition 库，提高了人脸图像的检测速度，特征提取时将图像转化成文本，更适合在 Hadoop 平台进行传输。最后，本文提出的方法具有较大的实用性，本文实验使用的数据集与待检测图像，涉及到的现实场景更多，包含不同表情、不同姿态、不同面部角度以及含有复杂背景干扰的图像，也含有彩色图像和黑白图像。由此可见，本文提出的方法的实用性更好、对大规模图像处理效率更高。

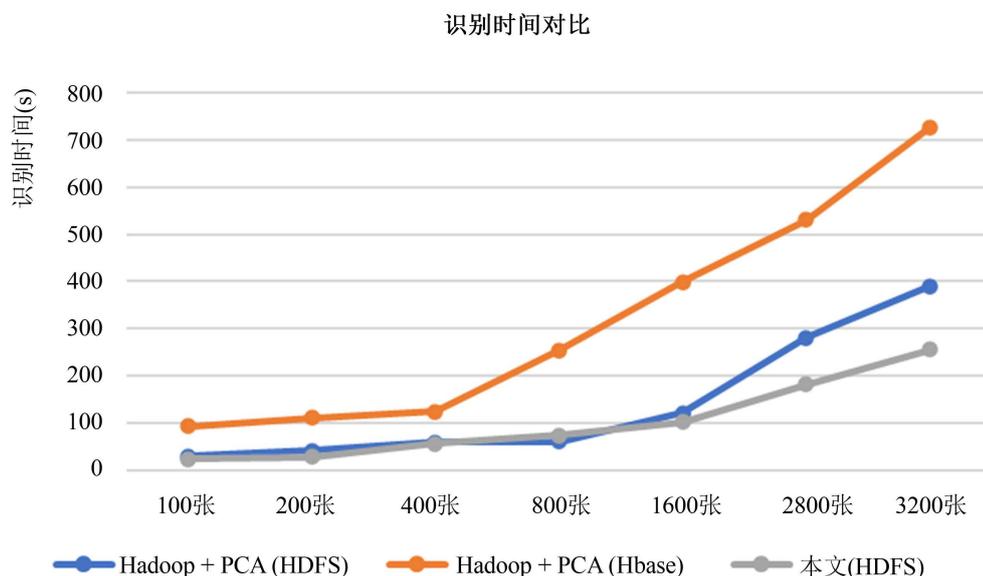


Figure 9. Face recognition time comparison graph based on Hadoop

图9. 基于 Hadoop 的人脸识别时间对比图

## 6. 结论

针对传统的人脸识别技术无法很好的处理大规模图像, 以及目前的大规模人脸识别方法, 大多只能对轻度姿态变化、单一模态的图像进行有效识别的问题, 本文将改进后的开源人脸识别库与 Hadoop 并行处理框架相结合, 可以将采集到的多模态图像从大规模多模态数据集中成功识别, 由于实验所用的图像较为符合实际中采集的图像, 因此本文的方法具有较大的实用性。本文的方法效率高、成本低、容错性强、易扩展, 非常适合在实际中应用普及, 有助于全国各地的公安干警快速抓捕逃窜的嫌疑人。

## 基金项目

面向网上追凶的部分遮挡人脸识别研究(2020YSYB004), 伊犁师范大学一般科研项目(校级)。

## 参考文献

- [1] Shaina, D. and Kumar, S. (2021) Big Data Analytics Using Apache Hadoop. *Turkish Journal of Computer and Mathematics Education*, **12**, 4664-4668.
- [2] 李夏风, 王忠林, 汪宏. 公安人脸大数据平台的建设与应用[J]. *警察技术*, 2019(2): 63-70.
- [3] Zhang, B. (2019) Distributed SVM Face Recognition Based on Hadoop. *Cluster Computing*, **22**, 827-834. <https://doi.org/10.1007/s10586-017-1330-5>
- [4] Bogdanchikov, A., Kariboz, D. and Meraliyev, M. (2019) Face Extraction and Recognition from Public Images Using HIPI. *2018 14th International Conference on Electronics Computer and Computation (ICECCO)*, Kaskelen, 29 November-1 December 2018, 206-212. <https://doi.org/10.1109/ICECCO.2018.8634718>
- [5] Villegas-Cortez, J., Benavides-Alvarez, C., Aviles-Cruz, C., et al. (2021) Interest Points Reduction Using Evolutionary Algorithms and CBIR for Face Recognition. *The Visual Computer*, **37**, 1883-1897. <https://doi.org/10.1007/s00371-020-01949-8>
- [6] 任静. 基于 Hadoop 云计算环境下人脸识别系统的研究与实现[J]. *电子设计工程*, 2019, 27(5): 116-120.
- [7] 耿玉琴. 基于 MapReduce 的人脸识别的研究[D]: [硕士学位论文]. 西安: 西安科技大学, 2017.
- [8] Wu, H., Cao, Y., Wei, H., et al. (2021) Face Recognition Based on Haar Like and Euclidean Distance. *Journal of Physics: Conference Series*, **1813**, Article ID: 012036. <https://doi.org/10.1088/1742-6596/1813/1/012036>
- [9] 汪鹏鹏. 基于 face recognition 的人脸识别平台研究及应用[D]: [硕士学位论文]. 成都: 西南交通大学, 2019.

- [10] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, **1**, 886-893.
- [11] 李澎林, 邹嘉程, 李伟. 基于 HOG 和特征描述子的人脸检测与跟踪[J]. 浙江工业大学学报, 2020, 48(2): 133-140.
- [12] 李静, 侯德文. 一种基于 Haar-like 和 AdaBoost 结合的人脸检测算法[J]. 山东师范大学学报(自然科学版), 2015(4): 34-37.
- [13] 丁文涛. 基于 Hadoop 的人脸识别系统研究与设计[D]: [硕士学位论文]. 镇江: 江苏科技大学, 2017.
- [14] 王沈括, 邬少飞, 张华杰, 夏宁. 基于 Hadoop 的人脸识别并行化方法的研究[J]. 电脑与电信, 2019(12): 15-19+22.