

基于特征融合的合照人物位置信息自动化标注方法研究

王 宏¹ 轩保山¹
WANG Hong XUAN Baoshan

摘 要

针对合照中人物位置信息自动化标注效率低、易受人脸分辨率及身高差异影响等问题,文章提出了一种融合人脸检测、排位划分与人脸匹配的自动化人物位置信息标注方法。首先,利用YOLOv8模型对合照中的人脸进行检测并将人脸映射为二维平面点集;然后,对点集数据进行标准化处理,并通过肘部法则确定K-means聚类的最优聚类数量,用于区分不同排的人员。在此基础上,采用鲁棒最小二乘法对每一排进行直线拟合,以减少人物身高差异对排位的影响,并通过拟合直线的截距确定合照人物的具体位置,最后通过增强的人脸识别算法对合照中的人脸区域进行匹配,并将具体的人物姓名标注到对应的人脸区域上。由于合照中人脸区域分辨率较低,提出一个通过不同分辨率人脸图像进行特征融合以提高人脸识别准确率的方法。实验结果表明,特征融合后的人脸匹配准确率达到96.72%,准确率比单张图像高出5%以上。此方法能较为准确地完成合照中人物的自动化识别、分排与标注,对特定领域具有较高实用价值。

关键词

人脸检测与识别;位置信息;K-means;遗传算法;鲁棒最小二乘法

doi: 10.3969/j.issn.1672-9528.2025.08.045

0 引言

在日常生活和各类场景中,合照因其能够记录多人信息而广泛存在,如毕业合影、集体活动记录等。然而,由于合照中人物数量多、站位复杂,人工标注人物位置和信息不仅费时费力,而且容易出错。因此,如何开发一种高效、自动化的方法来检测、分排并标注合照中人物的位置信息成为一个具有重要意义的研究课题。

近年来,基于深度学习的人脸检测技术取得了显著进展。特别是YOLO系列模型凭借其高效、精准的目标检测能力,尤其是在多目标场景下漏检、误检都得到改善,因此被广泛应用于人脸检测中。董子平等^[1]针对在密集场景下多人脸检测容易漏检,小尺度人脸检测率不高的问题,通过引入FIT模块,获取小尺度特征,改进采样策略等,提出了基于YOLOv5s改进的多人脸检测算法YOLOv5s-MF,在WIDER FACE数据集上相较于原YOLOv5s提升了3.46%的AP值^[1]。另一方面,随着深度学习技术的发展,人脸匹配在生物识别领域得到了深入应用。邓良等人通过对人脸图像的均值哈希相似度和感知哈希相似度的加权来综合判断人脸匹配,在ORL标准人脸数据库上的匹配正确率到达

了96.2%^[2]。Facebook团队开发的DeepFace首次实现了在人脸识别中使用深度卷积神经网络进行端到端的特征学习,极大提高了人脸识别的准确性。而FaceNet和ArcFace,进一步优化了人脸特征提取^[3]。

针对低分辨率人脸匹配正确率稍低,本文提出通过不同分辨率人脸图像进行特征融合以提高人脸匹配准确率以及利用在合照中的一个先验知识即:在合照中,会尽量使每个人物的脸都显露出来,并且人物一般会按照一排一排阶梯站列,反映在合照中,人物所在的排越靠后,人脸的位置就越高。利用这些先验知识,无需相机内外参数或者深度学习模型就能对合照中的各个人物进行序位判定。

1 人脸检测与数据选取

YOLOv8由Ultralytics团队开发。YOLOv8延续了YOLO系列模型的传统,即在保持高检测精度的同时,追求更高的检测速度。YOLOv8的网络架构在前几代YOLO模型的基础上进行了进一步优化,采用了新的网络模块和连接方式,提高了模型的表达能力和特征提取效率^[4]。

通过YOLOv8对合照进行检测并将人脸映射为二维平面上的点集,点集中的每一个点都代表人脸的中心位置,这些点集代表的就是人物在图像中的位置信息。图1为1927年第五届索尔维会议物理学家合照,其中人物分布状况是合照中的常见情形。

1. 西安石油大学计算机学院 陕西西安 710065
[基金项目] 陕西省学位与研究生教育研究项目 (SXGER2023085)



图1 1927年索尔维会议合照

从图1中人脸分布情况可以看出,人脸的分布基本在水平方向,不难得出人脸坐标的散点图在仅有纵坐标值时分布得更加密集,而在更加密集的点集中,聚类算法会更加稳定,收敛也更快。因此参与聚类的数据选取仅有纵坐标,横坐标归0的人脸坐标点。

2 人物位置关系

2.1 K-menas 聚类算法与拟合

获取数据标准化的纵值数据后,需要进行聚类处理得到每一排的人脸坐标,K-means 聚类算法是最常用的聚类算法之一,算法简单,易于理解,效率高,并且合照上的人脸坐标分布相对稳定,这与 K-means 算法在数据分布点相对稳定的情况下可以产生较好的聚类效果相契合^[5],因此采用 K-means 算法进行聚类。

K-means 聚类算法的一个天然缺陷就是需要手动输入需要聚类的类别数 k ,也就是合照中人物站成的排数。所以需要一种方法来自适应地给出 k 值。有两种常用方法,分别是肘部法则和轮廓系数法,本文采用肘部法则。

正常情况下,合照中同一排的人物难免会出现某个人比平均值高很多或者低很多的情况出现,虽然直观上某人比平均水平高或者低半个头的高度时,但是映到像素上差值可能到达几百。这样的坐标值在进行最小二乘法拟合直线时就是一个异常点,为了减少异常点带来的影响采用鲁棒二乘法进行拟合^[6]。图2为传统直线拟合和鲁棒直线拟合的直观效果图,异常点位于图2顶部偏右,大幅偏离了其他数据点。

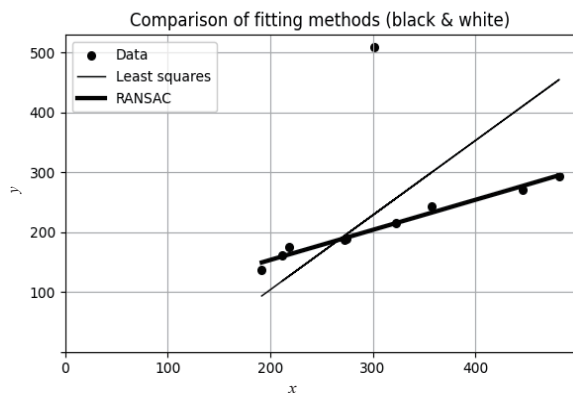


图2 鲁棒最小二乘法拟合异常点

可以看出传统拟合(细线)的直线受异常点影响较大,大幅度偏离数据点,而鲁棒直线(粗线)受到异常点影响较小,并未大幅度偏离数据点。通过聚类后的数据点拟合出的直线就可以得出每一排人物的具体位置。

3 特征融合进行人脸匹配

在人脸识别中,人脸匹配是除人脸检测之外的另一重要步骤,可以通过计算人脸图像的余弦相似度进行人脸匹配。在人脸匹配中,余弦相似度主要用于比较两个提取的特征向量之间的相似性,从而判断两张人脸是否属于同一个人。

由于人脸图像并不是刚性的,所以仅用单一人脸图像进行匹配会导致准确率低且不稳定。针对这一问题,本文研究中首先对两张人脸图像进行特征融合,即提取两张人脸图像特征并进行加权计算后作为该人的匹配数据,以此为基础建立人脸匹配数据库,人脸特征加权公式为:

$$F(f_1, f_2) = a \times f_1 + b \times f_2 \quad (1)$$

式中: F 为加权特征; f_1 为初始分辨率较低的图像特征; f_2 为初始分辨率较高图像特征;权重 a 对应初始图像分辨率较低的图像;权重 b 对应初始图像分辨率较高的图像($0 < a < b < 1$, $a+b=1$)。

人脸图像加权特征余弦相似度匹配流程如图3所示,首先提取一个人的特征时输入两张人脸图像,采用 Arcface 模型提取特征,之后提取两张图像的特征值 f_1 和 f_2 ,加权作为综合特征值 F ,对待匹配的图像进行同样的提取特征操作得到 f_3 ,计算 F 和 f_3 的余弦相似度 Cos 。采用余弦相似度匹配时, Arcface 模型默认阈值并不固定,一般为0.400~0.800,通常在低分辨率下设置为0.400。设定合适的阈值 M , Cos 大于 M 认为是同一人,当 Cos 大于0.400小于 M 时择优匹配,当小于0.400时,由于合照中的人脸区域较为模糊,因此即使小于0.400也可能为同一人。

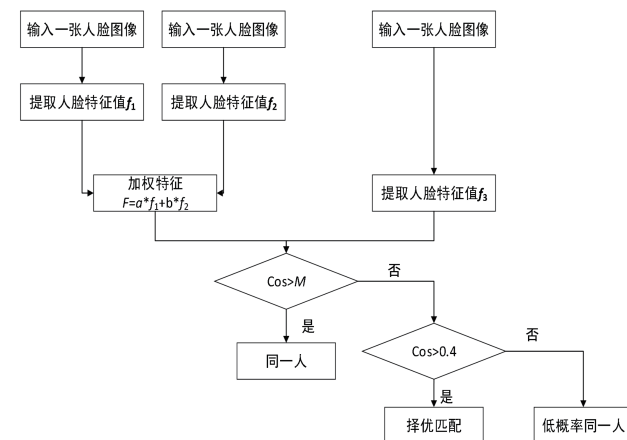


图3 加权特征余弦相似度匹配算法

从图3可以看出人脸匹配的准确率与权重系数 a 、 b 和阈值 M 有关。本文采用遗传算法优化权重 a 、 b 并寻找合适的阈值 M 。由于采用余弦相似度来进行匹配,而余弦相似度

的值越大代表越接近，因此需要同一人有3张不同的人脸图像，其中两张用来加权提取特征，并和第3张图像计算相似度，通过不断优化 a 、 b 使得同一个人的不同人脸图像之间的余弦相似度达到最大，即在数据集上的平均余弦相似度最大。

规定 a 、 b 的精度为 0.01，使用 LFW 标准人脸库，筛选出每人有 3 张图像的子文件夹，并使用 YOLOv8 进行人脸检测并截取人脸区域图像保存，由于 Arcface 模型训练时采用 112 px×112 px 分辨率图像，因此对小于 112 px×112 px 的人脸图像采用 LANCZOS4 插值方法进行 2 倍放大。共筛选出 739 人，一共 2 217 张人脸图像，随机选取 60% 的数据即 443 人的图像，分别提取特征值作为优化数据库。图 4~6 为实验使用的部分样例。



图 4 样例 1



图 5 样例 2



图 6 样例 3

传统的遗传算法采用二进制编码染色体来表示权值 a 和 b ，在优化迭代的过程中会大量地进行解码，编码以及翻转二进制位的变异操作，尤其是在高精度时，需要更多的迭代，计算效率较低，因此本文采用实数编码以追求更高的计算效率。实数编码与传统的二进制编码不同如图 7 实数编码所示^[7]。

4 实验结果及分析

实验环境为：Windows10 操作系统，配备 Intel Core i7 9th，16 GB 内存，NVIDIA GeForce 1660ti，6 GB 显存。实验平台为 PyCharm，Python 版本 3.8.0，PyTorch-GPU 版本 2.01，CUDA 版本为 11.8。

4.1 人脸检测模型及聚类结果

YOLOv8 模型训练的关键参数如表 1 所示，其余参数均为默认值。

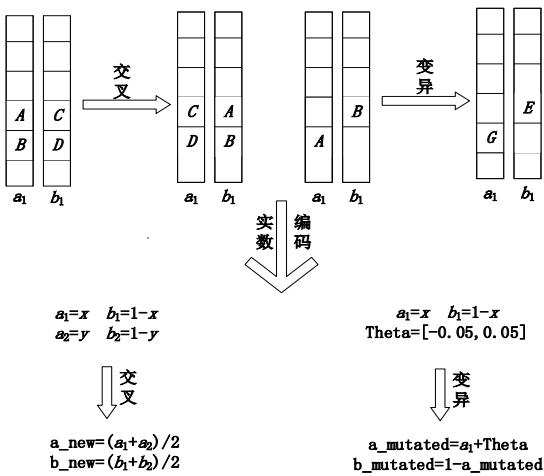


图 7 实数编码

表 1 模型训练参数

参数	设置
Batch_Size（每次更新批量数）	8
Epochs（训练迭代次数）	100
Images_Size（图片大小）	640

训练所用数据集为 Roboflow 网站上的公开数据集。共有 人脸图片 3 521 张。

4.2 权值优化结果

在筛选过的 LFW 数据库上使用遗传算法对参数 a 、 b 进行优化，设置交叉率为 0.5，变异率为 0.03，种群数量为 200。经过迭代优化得出：当 $a=0.41$ ， $b=0.59$ 时，计算出的平均余弦相似度最大。

在 Schroff 等人^[8]提出的 FaceNet 一文中，优化阈值的选择方法基于统计分析，找出一个百分位数作为阈值，以保证大多数样本被判定为匹配。本文选择了余弦相似度的 95% 分数位作为阈值，以确保绝大多数样本的匹配结果高于此值。在 $a=0.41$ 和 $b=0.59$ 的权重下，找出了一个实验性阈值 $M=0.516$ ，在参与实验 443 人中的 70% 即 310 人的数据中，有 95% 的加权余弦相似度超过了该值，在剩余的 30% 即 133 人的数据中，有 93.74% 的人加权余弦相似度超过了该值，因此采用 $M=0.516$ 作为阈值。

在 739 人剩余的 30% 数据即 296 人的图像进行测试，具体操作为：首先在每人 3 张图片的文件夹中进行拆分，分为两个部分，一部分有两张图像，用来加权提取称为加权部分；另一部分仅有一张图像用来和加权部分进行余弦相似度匹配称为匹配部分^[9]，选取匹配文件夹中的 10% 的数据，即 31 人的匹配部分进行互相随机交换，并打上 False 标签，未调换的文件夹打上 True 标签。最后将对应的匹配部分和加权部分计算余弦相似度 Cos，Cos 大于阈值 M 记为 True，Cos 小于阈值 M ，记为 False，最后将实验标签与原始标签对比得出准确率^[10]。单一图像（ $a=1, b=0$ ； $a=0, b=1$ ），优化加权图像

($a=0.41, b=0.59$), 和平均加权图像($a=0.5, b=0.5$)的平均余弦相似度以及准确率如表2实验结果所示。

表2 实验结果表

权值	平均余弦相似度	准确率/%
$a=1, b=0$	0.625	90.35
$a=0, b=1$	0.632	91.51
$a=0.5, b=0.5$	0.645	95.26
$a=0.41, b=0.59$	0.660	96.72

通过表2可以看出, 经过加权提取特征后, 平均余弦相似度比单一图像和平均加权图像的平均余弦相似度分别高了5.60%、4.43%、2.32%。在阈值 $M=0.516$ 认为是同一人的情况下, 优化加权的准确率比单一图像分别高出了6.37%、5.21%, 比平均加权图像高出了1.54%。

4.3 人脸匹配及标注结果

图8为2008年NBA全明星赛西部阵容的合照标注。搜集到了本届全明星赛全部26名球员每人2张人脸图像共52张, 使用YOLOv8模型提取人脸保存, 再使用Arcface模型在权重 $a=0.41, b=0.59$ 的情况下加权提取各人物的特征并和名字对应一一存储, 形成人脸数据库来进行匹配实验, 人物头顶的2.1表示按照从前往后, 从左往右的顺序是第二排第一个, 图像的左上角标注rows: 2, sum: 12表示图中的总排数为2, 总人数为12, row 1:5表示第一排共有5人, 以此类推。从图8标注结果来看, 只有位置为2.6的标注结果错误, 实际标注为比尔·拉塞尔, 应该标注为卡梅隆·安东尼。



图8 NBA全明星赛西部阵容合照标注结果图

图9为顶视图, 其中半径相同的实心圆代表为同一排。

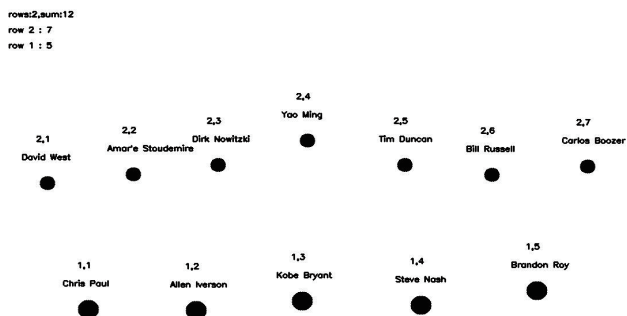


图9 NBA全明星赛西部阵容合照顶视图

5 结语

本文提出了融合人脸检测、排位划分与人脸匹配的自动化方法, 用于合照中人物的排位检测和标注。实验结果表明, 该方法能够有效地对照片中的人物进行自动分排, 能够准确识别并匹配同一个人, 并在合照中快速指出某人的具体位置, 减少了传统人工操作中的复杂性和误差。

参考文献:

- [1] 董子平, 陈世国, 廖国清. 基于YOLOv5s的密集多人脸检测算法[J]. 计算机工程与科学, 2023, 45(10): 1838-1846.
- [2] 邓良, 许庚林, 李梦杰, 等. 基于深度学习与多哈希相似度加权实现快速人脸识别[J]. 计算机科学, 2020, 47(9): 163-168.
- [3] 杨雅莉, 黎英, 章育涛, 等. 面向人脸识别的多模态研究方法综述[J/OL]. 计算机应用, 1-15[2024-11-11]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20240816.1952.003.html>.
- [4] 赵会鹏, 曹景胜, 潘迪敬, 等. 改进YOLOv8算法的交通标志小目标检测[J]. 现代电子技术, 2024, 47(20): 141-147.
- [5] 李柏, 郭晨, 李博, 等. 基于K-means聚类算法的洁净厂房无线网络AP布局优化设计[J]. 现代工业经济和信息化, 2024, 14(9): 127-129.
- [6] 王炜, 曹新宇, 何森. 分布鲁棒最小二乘问题的割平面算法[J]. 辽宁师范大学学报(自然科学版), 2017, 40(3): 293-296.
- [7] 曹源, 王雯婧, 宿帅, 等. 基于遗传算法的重载车轨耦合动力学模型校正[J/OL]. 铁道科学与工程学报, 1-12[2024-11-27]. <https://doi.org/10.19713/j.cnki.43-1423/u.T20241586>.
- [8] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 815-823.
- [9] 党宏社, 孟饶辰, 高宛蓉. 基于双流特征交叉融合 Efficient Transformer 的人脸表情识别[J/OL]. 计算机工程与应用, 1-9[2024-11-27]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240822.1436.005.html>.
- [10] LUO C, ZHANG J, ZHANG X Q. Tensor multi-view clustering method for natural image segmentation[J]. Expert systems with applications, 2025, 260: 125431.

【作者简介】

王宏(1968—), 男, 陕西西安人, 硕士, 副教授, 研究生导师, 研究方向: 人工智能、大数据应用、计算机视觉、物联网技术。

轩保山(1999—), 男, 河南周口人, 硕士研究生, 研究方向: 模式识别、计算机视觉。

(收稿日期: 2025-03-24 修回日期: 2025-07-31)