

基于深度学习的运动目标交接算法^①



曹建荣^{1,2}, 武欣莹¹, 吕俊杰¹, 王亚萌¹, 杨红娟^{1,2}, 张旭¹

¹(山东建筑大学 信息与电气工程学院, 济南 250101)

²(山东省智能建筑技术重点实验室, 济南 250101)

通讯作者: 曹建荣, E-mail: 317370808@qq.com

摘要: 针对多摄像机非重叠视域下存在的运动目标不连续性和不确定性的问题, 提出一种基于深度学习的运动行人目标的交接算法. 首先基于深度卷积神经网络构建人脸特征提取模型, 对人脸特征提取模型进行训练, 获得精确的人脸特征. 然后比较两种常用的相似度度量方法, 选择其中一种更适合的相似度度量方法, 以完成最优的人脸匹配过程, 提高人脸匹配的准确率. 最后通过对不同摄像机下的人脸进行特征匹配找到最匹配的人脸, 实现运动目标的交接. 实验表明, 深度神经网络可以减少运动目标丢失的概率, 准确地提取到运动目标的人脸特征, 有效完成多摄像机下运动目标的交接跟踪任务.

关键词: 深度学习; 目标交接; 特征提取; 相似度度量

引用格式: 曹建荣, 武欣莹, 吕俊杰, 王亚萌, 杨红娟, 张旭. 基于深度学习的运动目标交接算法. 计算机系统应用, 2020, 29(12): 126-134. <http://www.c-s-a.org.cn/1003-3254/7675.html>

Algorithm of Moving Target Handover Based on Deep Learning

CAO Jian-Rong^{1,2}, WU Xin-Ying¹, LYU Jun-Jie¹, WANG Ya-Meng¹, YANG Hong-Juan^{1,2}, ZHANG Xu¹

¹(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

²(Shandong Provincial Key Laboratory of Intelligent Building Technology, Jinan 250101, China)

Abstract: In view of the discontinuity and uncertainty of moving objects in the non overlapping view of multiple cameras, a handover algorithm of moving pedestrian target based on deep learning is proposed. Firstly, a face feature extraction model is constructed based on deep convolution neural network, and the face feature extraction model is trained to obtain accurate face features. Then compare two common similarity measurement methods, choose a more suitable similarity measurement method to complete the optimal face matching process, and then improve the accuracy of face matching. Finally, the most matching face can be found by feature matching under different cameras to realize the handover of moving objects. Experiments show that the depth neural network can accurately extract the facial features of moving objects, achieve the accurate matching process of human faces, and effectively complete the task of moving object tracking under multi cameras.

Key words: deep-learning; target handover; feature extraction; similarity measurement

随着监控场景复杂度的增加和视频监控系统规模的不断扩大, 基于智能视频分析的多摄像机下的运动目标交接技术逐渐成为计算机视觉领域热门的研究方

向之一. 运动目标交接的目的是分别在相邻的摄像机下确认运动目标的身份等各项参数指标, 由此利用毗邻的多台摄像机实现对该运动中目标对象的不间断追踪.

① 基金项目: 山东省重点研发计划 (2019GSF111054, 2019GGX104095); 山东省重大科技创新工程 (2019JZZY010120)

Foundation item: Key Research and Development Program of Shandong Province (2019GSF111054, 2019GGX104095); Major Science and Technology Innovation Engineering of Shandong Province (2019JZZY010120)

收稿时间: 2020-04-06; 修改时间: 2020-04-28, 2020-05-10; 采用时间: 2020-05-13; csa 在线出版时间: 2020-11-30

以目标的特征融合为基础的目标交接一般应用于无重叠视域的多摄像机连续跟踪. 在进行目标的特征提取时, 陆兴华等^[1]通过统计灰度直方图的信息可以提取到目标的特征, 但目标和背景的颜色差异有可能带来特征匹配误差; Liang 等提出使用 LOMO (Local Maximal Occurrence) 算法^[2]将手工特征和深度特征两种特征交融在同一个深度网络中完成目标的交接, 但网络训练速度较慢. 而常用的特征相似度度量有余弦相似度、欧式相似度、MLAPG 相似度^[3]和 NLML 相似度^[4].

以空间模型为基础的运动目标交接一般应用于有重叠视域的多摄像机连续跟踪. 张正本等^[5]提出了一种采用卡尔曼一致滤波对空间内多个目标状态进行一致性估计的方法解决分布式多运动目标交接问题, 但该方法对测量模型要求高, 实用性不高. 阳小燕等^[6]提出了一种改进的基于视野分界线的多摄像机运动目标跟踪算法, 但当视野分界线被遮挡时会影响目标交接的准确性. Ur-Rehman 等^[7]为了解决多个运动目标在交接时所产生的遮挡问题, 采用聚类模型理论作为指导方法进行研究, 但该模型的收敛速度较慢, 受限于样本容量大小.

以上两种传统的目标交接方法可以根据目标的特征和空间模型特征完成重叠或非重叠视域的多摄像机连续跟踪, 但以上模型较为复杂且无法兼顾目标交接的实时性和准确性. 近年来, 深度学习算法在运动目标交接和跟踪中展现出强大的速度优势, 因此本文以人脸为研究对象并利用深度学习的方法完成运动目标交接, 在保持算法原有的速度优势的同时, 还可以缓解由目标相互遮挡引起的目标不连续和不确定问题, 提高交接的精度.

本文通过对不同摄像机下获取的人脸图像进行匹配实现运动目标在不同摄像机中的目标交接匹配, 实现对运动目标的连续跟踪. 首先利用深度学习检测方法行人运动目标, 并提取行人的人脸特征; 然后通过合适的相似度度量方法对不同摄像机中的人脸进行特征相似度匹配, 找到最匹配的人脸, 有效地实现毗邻摄像机对运动目标的目标交接算法.

1 整体算法介绍

本文整体算法流程图如图 1 所示, 算法的主要结构如下.

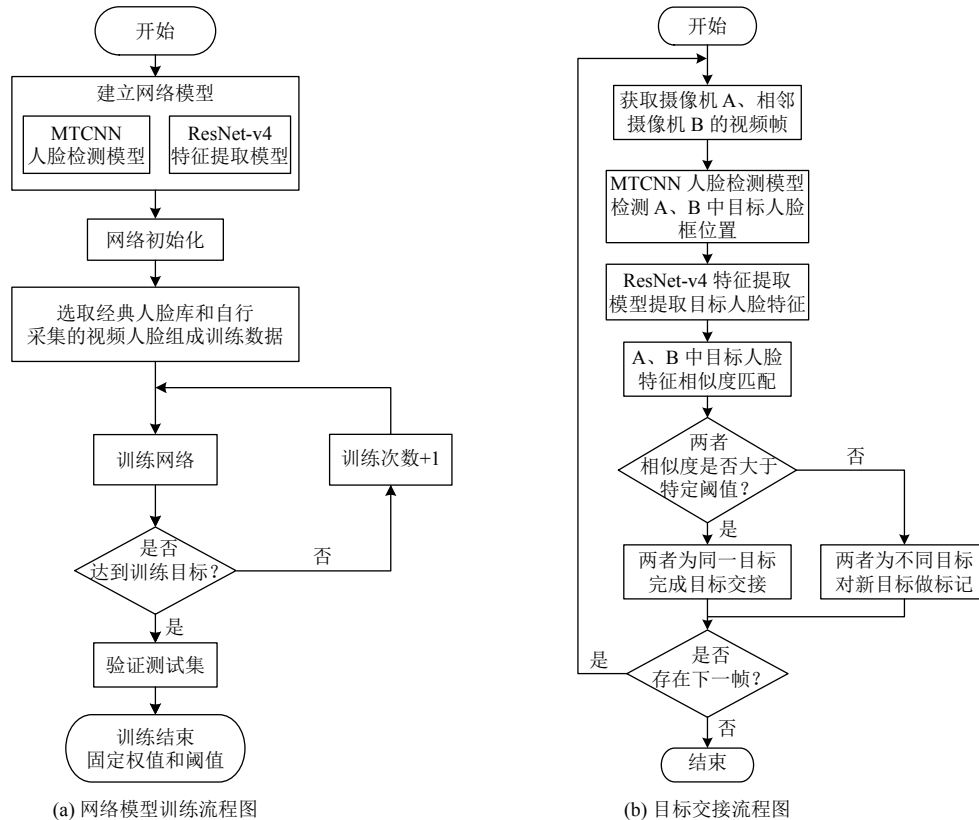


图 1 整体算法流程图

(1) 首先, 运动目标作为非刚体易发生相互遮挡且其他生物特征在匹配时容易产生匹配误差, 因此本文选择人脸作为研究对象, 利用深度学习提取人脸的特征;

(2) 然后, 开始网络模型的训练, 训练流程如图 1(a) 所示. 分别建立 MTCNN 人脸检测模型和 ResNet-v4 特征提取模型, 选取经典人脸库和自行采集的视频人脸库组成训练数据, 对模型进行训练. 根据本文训练数据的特点对模型进行调整和改进, 确定网络的权值和阈值, 提高模型的自适应性和训练速度;

(3) 最后, 将摄像机 A、相邻摄像机 B 获取的视频帧输入训练达标的模型中实现目标交接. 交接流程如图 1(b) 所示. 根据人脸检测模型检测到人脸的位置, 根据特征提取模型提取检测到的人脸特征, 利用相似度度量将人脸特征在相邻摄像头下进行人脸特征匹配,

并循环获取摄像机 A、B 视频帧不断进行检测和匹配过程, 最终实现所有目标的交接.

2 人脸检测算法

人脸检测旨在有行人运动目标的视频监控序列中判断是否存在人脸, 并同时给出人脸的数量、位置等参数, 是运动目标交接的首要环节. 本文利用 MTCNN (Multi-Task Convolutional Neural Network) 网络进行人脸检测. 该网络由图像金字塔及三阶段级联 CNN 组成, 将人脸检测、人脸位置标定和人脸特征点检测同时进行^[8]. 级联网络包括 P-Net (Proposal Network) 网络、R-Net (Refine Network) 网络及 O-Net (Output Network) 网络, MTCNN 3 个 CNN 子网络结构如图 2 所示.

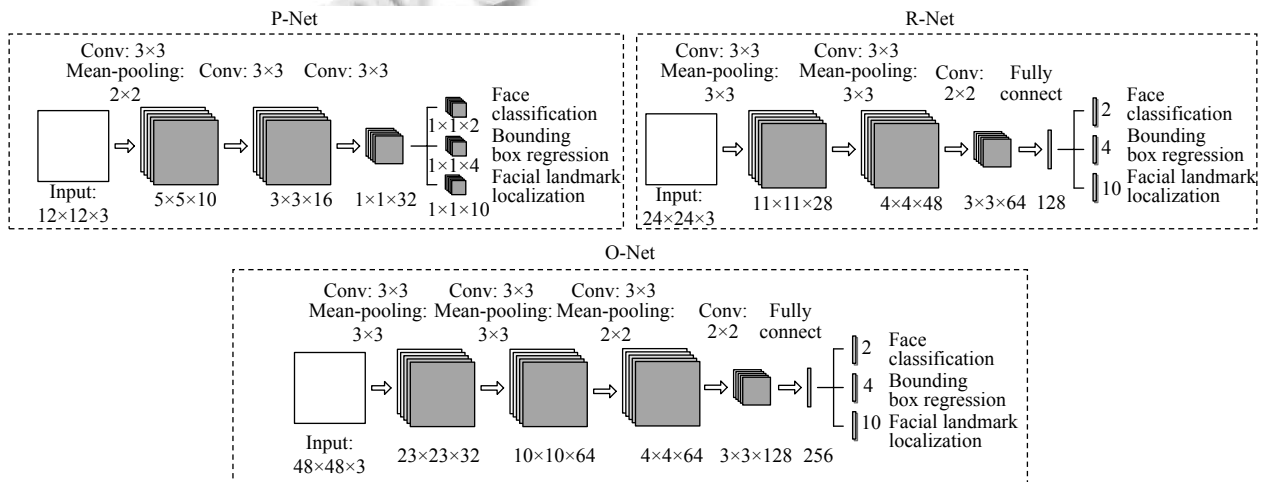


图 2 MTCNN 3 个子网络结构图

该算法的主要流程如下:

1) 首先将图片按照不同比例缩放成大小不同的图片, 建立图像金字塔, 最后得到的最小的图像最短边要大于等于 12;

2) 然后将 12x12 的图片输入 P-Net 网络, 通过计算生成大量的人脸候选框和边框回归向量, 利用边框回归的方法来矫正生成的人脸框, 利用非极大值抑制合并重叠的候选框;

3) 其次将 P-Net 网络输出的图像缩放至 24x24, 输入至 R-Net 网络中, 进一步筛掉大量非人脸框, 继续使用边框回归和非极大值抑制的方法合并候选框, 达到高精度过滤和人脸区域优化的效果;

4) 最后将 R-Net 网络输出的图像缩放至 48x48,

输入至 O-Net 网络中, 最终输出回归框分类、回归框位置及人脸特征点的位置.

由上可知, MTCNN 网络为了兼顾性能及准确率, 避免滑动窗口加分类器等传统方法带来的性能消耗, 先使用简单网络生成有一定可能性的人脸区域候选框, 然后再使用更复杂的网络进行细分类和更高精度的人脸框回归, 并让这一步递归执行来构成三层网络, 完成快速高效的人脸检测.

3 人脸特征提取算法

人脸特征提取是多摄像机运动目标交接方法中的关键部分, 主要分为基于传统和深度学习的特征提取方法. 传统方法侧重于通过人的肤色、纹理和几何特

征等方面提取特征. 例如, HOG 特征提取算法通过计算图像局部区域的梯度方向直方图来获取图像特征^[9]; LBP 算法通过区域块匹配的方法提取人脸图像的纹理特征^[10]. Haar-like 特征算法利用人脸器官间的几何关系来提取人脸特征^[11].

基于深度学习的卷积神经网络模型主要有 DeepFace^[12]、DeepID^[13]、FaceNet^[14] 模型等, 其算法的精准度已经优于人眼观测的效果. 最新研究成果表明, 深度卷积神经网络通过不断的训练提取图片中的特征, 具有较强的客观性和选择性, 因此本文选择深度学习的方法搭建人脸特征提取模型.

3.1 ResNet-v4 深度学习框架

本文运用深度学习方法对人脸特征进行提取时需要神经网络模型进行训练, 主要参考 VGGNet^[15] 与 GoogleLeNet^[16] 的结构模型, 使用 ResNet-v4^[17] 的模型提取人脸的特征, 该模型可以避免训练集的准确率随网络层数的加深而下降的风险. 最后依据本文训练数据的特点, 进一步对模型进行完善, 完善之后的训练主网络结构图如图 3 所示.

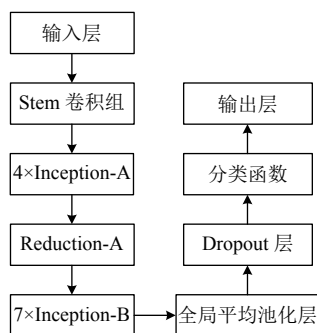


图3 训练主网络结构图

图 3 展示了由 1 个输入层、13 个卷积组、1 个全局平均池化层、1 个 Dropout 层及 1 个输出层组成的改进的网络模型. 开始先向模型中输入 3 张 160×160 尺寸的训练图片, 经过 Stem 卷积组后生成 18×18 的特征图共 384 个; 其次通过卷积组 Inception-A 共 4 个, 得到 18×18 的特征图形共 384 个; 之后通过卷积组 Reduction-A 1 个, 得到 9×9 的特征图形共 1024 个; 再通过卷积组 Inception-B 共 7 个, 得到 9×9 的特征图形共 1024 个; 最后经过平均池化层, 得到 1×1 的特征图形共 1024 个.

该残差网络相对于之前的深度学习网络具有很多改进和优点. 在参数数量方面, 用 2 个小的卷积核串联

代替较大的卷积核可以降低计算量的同时增加神经网络的表征能力. 例如, 在该网络中将 2 个 3×3 小卷积核串联代替 1 个 5×5 的大卷积核, 训练参数由 25 个变为 18 个; 在网络宽度方面, 为了提高网络模型对尺寸的自适应能力, 将卷积组中 1×1 、 3×3 的卷积核和 3×3 的池化层堆叠在一起增加网络结构的宽度; 在卷积降维方面, 卷积网络的增宽会急速地增加卷积通道的数量, 为了避免其对神经网络模型的运算速率产生影响, 将 Inception-B 卷积组与 1×1 的卷积核进行卷积计算, 这样能够大大降低大卷积核的通道数.

3.2 人脸特征提取数据集

在网络的框架构建完成之后, 还要处理和分析用于人脸特征匹配的数据集. 当今网络化时代有大量用于人脸识别的开源数据集. 例如, CelebA (Celeb Faces Attributes Dataset) 数据集^[18] 是由香港中文大学汤晓鸥教授实验室公布的大型人脸属性数据集, 该数据集包含了 200 000 张人脸图片, 人脸属性有 40 多种; 为了探究人脸识别在非限制环境中产生的问题, LFW 数据集^[19] 应运而生, 这一数据集共包含人脸图像 13 000 多张, 涉及人数为 5 749 人, 每张人脸均被标注上了对应的人名, 通过定义无监督、限制与非限制 3 个标准协议来评估人脸匹配的性能, 是标准的人脸匹配测试数据集.

本文采用 LFW 标准人脸匹配数据集作为测试数据集, 训练数据集是根据目标交接要求, 通过网络收集到 8 400 张大小为 160×160 的人脸图片, 并将其分为 400 组不同的类别, 完成人脸特征提取数据集的制作.

3.3 人脸特征提取模型损失函数

本文的损失函数借鉴文献 [20], 采用一种 Softmax 损失与中心损失相结合的方法. 这种方法能够降低同类内部的差异, 同时能够提高类别之间的差异, 以此来增强特征提取时模型的辨别与泛化能力.

Softmax 的损失函数主要用来处理多分类问题, 它的公式如下所示:

$$L_{\text{Softmax}} = - \sum_{i=1}^N y_i \ln \hat{y}_i \quad (1)$$

其中, L_{Softmax} 表示 Softmax 的损失函数; i 为图片数量; \hat{y}_i 表示 Softmax 的输出向量的 y 第 i 个值, 输出的是预测结果; y_i 表示个体的真实类别, 取值为 0 或 1, 即真实标签对应位置的那个值为 1, 其他都为 0.

中心损失函数为一种经典的聚类算法, 当卷积神

经网络具备多个特征时,每一批次都能够通过计算得到多个特征中心,同时也能够都得到对应的的损失函数.式(2)为中心损失函数的表达形式:

$$L_{\text{Centre}} = \frac{1}{2N} \sum_{i=1}^N \|x_i - C_{y_i}\|_2^2 \quad (2)$$

其中, L_{Centre} 表示中心损失函数; x_i 为第 i 张图片的特征值; C_{y_i} 为第 i 张图片所处类别的中心,即 y_i 所属类别的特征值的中间点.

分类的特征中心 C_{y_i} 随着深度特征的变化而变化,初次训练时需要随机设定 C_{y_i} 的值,单次训练时需要利用式(3)对 C_{y_i} 进行更新:

$$\Delta C_j = \frac{1}{N} \cdot \frac{\sum_{i=1}^N \delta(y_i = j) \cdot (C_j - x_i)}{1 + \sum_{i=1}^N \delta(y_i = j)} \quad (3)$$

其中, ΔC_j 表示分类中心的变化量; y_i 代表类别个体的实际所属,当它与 C_j 表现出一样的类别时,对 C_j 进行迭代更新;当它与 C_j 表现出不一样的类别的时候,不进行迭代更新.

通过中心损失与 Softmax 损失这两种函数的合作监督,构建总损失函数.总损失函数计算公式如式(4):

$$\begin{aligned} L &= L_{\text{Softmax}} + \lambda L_{\text{Centre}} \\ &= - \sum_{i=1}^N y_i \ln \hat{y}_i + \lambda \frac{1}{2N} \sum_{i=1}^N \|x_i - C_{y_i}\|_2^2 \end{aligned} \quad (4)$$

式中, L 表示总损失函数; λ 的值由人为设置,它是一个使两种损失相互平衡的超参,选择合适的 λ 可以增强网络特征的识别能力.当 $\lambda = 0$ 时,可以将式(4)的函数认为是仅有 Softmax 损失的情况. λ 取值越大种类之间越分散,反之越集中.本文对文献[19]进行分析,取 $\lambda = 0.9$.

通过以上 Softmax loss 和 Center loss 联合监督的方法可以解决三重损失中采样复杂的问题,也可以解决对比损失中无法采集合适样本对的问题,可使模型的泛化能力大大提升.

4 人脸相似度匹配方法

构建好人脸特征提取框架后,还需要通过相似度度量方法对人脸的相似度进行计算,完成人脸匹配.最常用的相似度测量方法是余弦距离相似度,最常用的距离度量是欧氏距离,很多相似度的度量和距离度量的方法都是在这两种方法基础上的变化和推演而来

的.文中将余弦距离与欧氏距离进行对比分析,对最佳的度量算法进行选用.

(1) 欧氏距离 (Euclidean Distance)

欧氏距离对个体差异的分析通常取决于维度数值的大小,主要通过数值表现的差异来判断个体间的类别.

例如空间是 N 维的,那么 N 维向量 $x(a_1, a_2, a_3, \dots, a_n), y(b_1, b_2, b_3, \dots, b_n)$ 之间的欧式距离为 d_N . 其计算公式如下所示:

$$d_N = \|x - y\|_2 = \sqrt{\sum_{k=1}^N (a_k - b_k)^2} \quad (5)$$

(2) 余弦距离 (Cosine Distance)

在 N 维空间中,还可以通过余弦距离来得到两个向量组成角度的余弦值,余弦距离大多数情况是从向量的方向上来分辨差别,依据向量在方向上的差异来判断两者是否为同一类别^[8].

假设在 N 维空间中,计算 N 维向量 $x(a_1, a_2, a_3, \dots, a_n), y(b_1, b_2, b_3, \dots, b_n)$ 之间的夹角余弦公式. 其计算公式如下所示:

$$d_N(\cos \theta) = \frac{\sum_{k=1}^N a_k b_k}{\sqrt{\sum_{k=1}^N a_k^2} \sqrt{\sum_{k=1}^N b_k^2}} \quad (6)$$

已知向量之间的夹角的余弦值位于 $[-1, 1]$ 这一范围内,两个不同向量在 N 维空间的方向相同时为正向,1 为最大值;而方向相异时为负向, -1 为最小值. 夹角余弦越大表示两个向量的夹角越小,说明人脸图像的差异越小. 反之表示两向量的夹角越大,说明人脸图像的差异越大.

深度神经网络的最后往往连接不同的分类函数,用来输出区别人脸类别的预估概率可能性. 往往要筛除掉全连接一层,同时把最后一层卷积层当作人脸的“向量表示”,用来表示人脸的“特征”. 假设两张人脸图像的“向量表示”表示为 $x(a_1, a_2, a_3, \dots, a_n), y(b_1, b, b_3, \dots, b_n)$ 将其带入式(5)或式(6)可以计算出不同人脸图片的相似程度.

5 运动目标交接算法

多摄像机间的监控区域主要根据人脸面部清晰度和监控视觉特性这两个方面来划分,图4所展示的是摄像机监控下的视域范围.

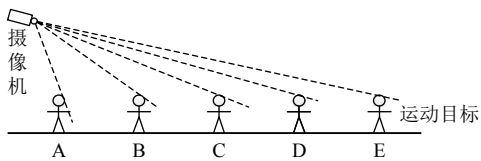


图4 摄像机监控下的视域范围图

如图4可知,假如运动目标的监控视域随着监控角度的变化可以等分为A—E这4个监控区域,由于摄像机具有景深的特点,在以上4个监控区域中,监控区域B—D这一视域范围内是捕捉人脸图片的最优范围.所以,本文以D线作为视线边界并由此开始捕捉运动目标的人脸图片,将人脸捕捉间隔设置为5帧,并对同一目标捕获3张图片加以保存.

相邻两摄像机在非重叠视域下的运动目标交接步骤如下:

- 1) 当运动目标通过第一个监控摄像机的监控区域时,对所有目标人脸进行检测并标注不同的ID,相同目标人脸图像抓取3张并保存;
- 2) 若目标行人运动到邻近的摄像机的视域范围内,该摄像机随即对人脸图片进行捕获;
- 3) 分别对两台相邻摄像机检测到的人脸进行人脸对齐和特征提取;
- 4) 最后,将第二台摄像机采集到的人脸与前一

摄像机的目标人脸图片进行对比分析,根据分析结果对不同ID的人脸特征相似度进行平均值计算;

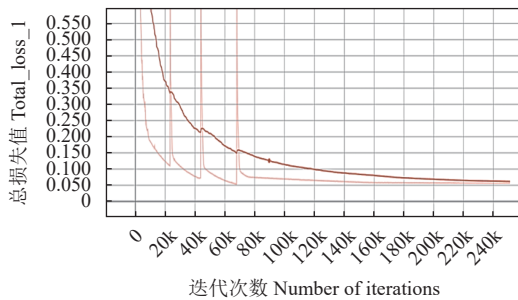
5) 假设特征相似度计算选用余弦距离:已知设定阈值 a 大于设定阈值 b ,对后一台摄像机与前一台摄像机分别采集到的人脸图片进行对比分析计算,若对比结果相似度的值比预定的阈值 a 大,那么可认为二者是相同的对象,完成目标交接;

6) 若计算结果的相似度比预定的阈值 a 小,则选出计算结果里最大的一个值,若该值比预定的阈值 b 大,那么可认为二者是相同的对象,若比阈值 b 小,那么可认为二者是不相同的对象,并对这个新的目标做一次标记.

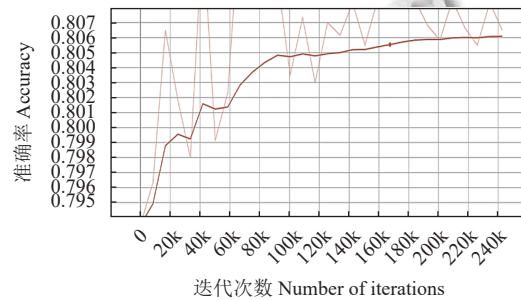
6 运动目标交接实验分析

6.1 人脸特征提取模型训练

人脸特征提取模型以3.1节中所搭建的深度学习框架为依据,搭建所需要的实验分析平台.训练共进行30轮,设置批次大小为60.其中每经过10轮训练,将学习率更改为上一阶段的10%,从第10轮开始分别设定成0.005、0.0005、0.00005,神经元的保留率在Dropout层上被设定为0.8.最终得到如图5所示的训练结果.



(a) 网络总损失值与迭代次数的关系图



(b) 网络测试集准确率与迭代次数的关系图

— 实际趋势 — 规范化后的趋势

图5 网络训练结果图

由图5中的网络训练结果图分析得到,总损失值会因为训练的不断进行而慢慢变小,而测试集的精准度因此有所升高.但是由于实验采集到的人脸模型训练数据量不够大,最终训练的精确度达到80%以上.实验证明,该网络模型的性能和自适应能力可以满足数据集训练的要求,从正确率趋势可以看到,准确率会随着训练的轮数和数据量的增加而得到进一步的提升.

6.2 人脸相似度度量方法的确定

根据6.1节训练完成人脸的特征提取模型性能之后,还要确定人脸匹配的判定阈值以及适用的相似度的度量方法.本节随机选取脸部状态不同情况下的5张相同人脸和5张不同人脸,分别采用欧式距离与余弦距离这两种度量方法,他们得到的人脸的特征“距离”结果分别如表1与表2所示.

表1 欧式距离相同/不同人脸的相似度量

相同人脸	R1	R2	R3	R4	R5	不同人脸	A	B	C	D	E
R1	0.0000	0.5716	0.6255	0.9059	0.5038	A	0.0000	1.0216	0.9661	1.0660	1.0540
R2	0.5716	0.0000	0.6255	0.9223	0.6560	B	1.0216	0.0000	1.0159	1.0278	1.0163
R3	0.6255	0.6255	0.0000	0.6785	0.7340	C	0.9661	1.0159	0.0000	1.0862	0.9956
R4	0.9059	0.9223	0.6785	0.0000	0.9142	D	1.0660	1.0278	1.0862	0.0000	1.0408
R5	0.5038	0.6560	0.7340	0.9142	0.0000	E	1.0540	1.0163	0.9956	1.0408	0.0000

表2 余弦距离相同/不同人脸的相似度量

相同人脸	R1	R2	R3	R4	R5	不同人脸	A	B	C	D	E
R1	1.0000	0.9085	0.6479	0.9024	0.9085	A	1.0000	0.3237	0.3201	0.2994	0.1538
R2	0.9085	1.0000	0.6525	0.8035	0.6613	B	0.3237	1.0000	0.4165	0.3030	0.2489
R3	0.6479	0.6525	1.0000	0.8380	0.6525	C	0.3201	0.4165	1.0000	0.3030	0.2489
R4	0.8024	0.8035	0.8380	1.0000	0.8035	D	0.2994	0.3030	0.1583	1.0000	0.1953
R5	0.9085	0.6613	0.6525	0.8035	1.0000	E	0.1538	0.2489	0.2629	0.1953	1.0000

通过表1能够得到如下结论:在欧式距离算法得到的结果中,相同的目标人脸特征之间的“距离”大多都比1小,相异的目标人脸特征之间的“距离”大多都比0.8大.而通过表2能够得到如下结论:在余弦距离算法得到的结果中,相同人脸特征之间的“距离”大多都比0.45小,相异人脸特征之间的“距离”大多都比0.55小.通过对两种距离算法的比较可以得出,余弦距离的计算模型比欧式距离得到的重合度数值更小,并且前者衡量的是空间向量的夹角,更侧重于从方向上区分差异.综合衡量以上内容,挑选余弦距离算法来计算人脸特征的相似程度.

在确定完相似度的计算算法之后,还需要确定人脸匹配的判定阈值.本文根据在不同的情况下的相同人脸和不同人脸的图像各400张来计算人脸匹配的阈值.开始对相同的人脸对象特征进行相似度的计算,得出的数据结果中,针对小于0.45的那部分数值再取平均,得到其平均值是0.439;接着对不同的人脸对象特征进行相似度的计算,在得出的数值结果中,针对大于0.55的那部分数值再取平均,得到其平均值是0.561.综上所述,本文确定人脸匹配的阈值 $a=0.561$,阈值 $b=0.439$.

6.3 运动目标交接实验结果分析

本文中选择的两个相邻的摄像机分别为视域边界不重合的“走廊东”与“走廊中”的摄像机,通过完成对运动目标的人脸检测和特征提取的过程,来实现对运动目标交接算法的验证.实验结果如图6至图9.

如图6,展示的是“走廊中”的摄像机的监控范围.由图6(a)可知视频中出现了5个不同的行人,在目标

进入监控区域后获取视频中行人的人脸图像,同时记录下运动目标的ID,保留5个不同ID行人的人脸图片,人脸ID标记为1、2、3、4、5.由1.2节的目标交接原理可知,每张相同的ID保留3张,因此总共需要保留15张人脸图像.



(a) 人脸区域图像提取 a 图



(b) 人脸区域图像提取 b 图

图6 “走廊中”人脸区域图像提取图

如图7,展示的是监控摄像机位于“走廊东”的监控视域.由图可知当运动目标在走进该摄像机的监视范围之前,对目标运动对象展开检测和再次追踪,图中目标对象在交接之前被设置的ID分别为1、2.

如图8,展示的是目标进入“走廊东”摄像机监控的

视域范围内. 由图可知, 目标对象进入到监视范围后, 摄像机随即对人脸进行检测和特征提取, 然后与图 6(b) 的对应的人脸特征进行对比匹配. 若对比前后二者的结果显示匹配不成功, 则对这一目标对象进行连续的追踪; 若对比前后二者的结果显示匹配成功, 则对运动目标身份展开再一次的分析确认, 然后替换掉该目标人脸的 ID, 如图中展示, 人脸的 ID 分别从 1、2 变换成为 4、5.



图 7 目标再次跟踪图



图 8 运动目标交接图

最终再把两台摄像机, 即“走廊东”与“走廊中”的摄像机分别采集到的人脸特征进行相似度的分析对比. 其相似程度的比较结果如图 9 所示. 由图中结果可知, 对于同一个运动目标来说, 两者之间的余弦距离相对更大, 差异更小; 当两场景的人脸图像是不同目标时余弦距离相对更小, 差异更大. 因此可以由余弦相似度对比值完成人脸匹配的过程, 进而实现运动目标在不同摄像机下的交接.



图 9 人脸图像相似度对比图

6.4 算法比较

最后, 为进一步体现本文算法的优越性, 参考文献 [21] 的实验方法选择搭建好的双摄像机系统环境, 对不同交接算法的同一时间的视频帧进行了两组统计验证, 对比结果如表 3 所示.

从表 3 中可以看出, 在相同实验环境下本文的交

接算法的准确率高其他两种算法, 有更好的目标交接准确性. 主要由于本文的目标交接算法中的深度学习网络模型对数据波动较为敏感, 网络模型会根据本文数据特点及时调整, 从而保证交接的速度和准确率, 但深度学习的目标交接实现必须要建立在大量训练样本的基础上. 另外本文可以解决由目标遮挡或光照干

扰引起的目标不连续和不确定问题,使得整体的实验效果较为理想。

表3 不同算法目标交接结果对比

方法	一组			二组		
	正确数	错误数	准确率(%)	正确数	错误数	准确率(%)
文献[22]	111	89	55.0	119	81	59.5
文献[6]	165	35	82.5	170	30	85.0
本文算法	167	33	83.5	175	25	87.5

7 结论与展望

本文重点研究了在相邻摄像机间视域边界不重叠情况下的运动目标交接技术。首先利用深度学习检测运动行人的脸,搭建以人脸为研究对象的特征提取模型;之后比较两种相似度量方法,选取最优的度量方法对人脸特征进行相似度匹配;同时计算了大量不同情况下的相同与不同的人脸特征“距离”,通过数据拟合确定了本文人脸匹配的最优判定阈值;最后通过对不同摄像机下人脸进行特征匹配找到最匹配的人脸,实现运动目标的交接。从仿真结果可以发现,深度神经网络模型可以精确地提取运动目标的人脸特征,余弦距离的相似度计算模型更符合本文的实验要求,该算法有效地完成了多摄像机下运动目标的交接跟踪任务。从实际应用中,本文使用的方法相比较于传统的运动目标交接的方法减少了算法的复杂度和计算量,大大提高了运动目标特征提取的精确度,可以实现对运动目标的确定性和连续性跟踪,适用于复杂多变环境下的视频监控系统。

参考文献

- 陆兴华, 刘铭原, 龙庆佳, 等. 基于灰度直方图的运动目标特征检测算法. 计算机与现代化, 2019, (6): 71-75.
- Liang Z, Chi J, Hua R, *et al.* Person re-identification algorithm based on the fusion of deep feature and LOMO feature. International Society for Optics and Photonics, 2019.
- 周智恒, 刘楷怡, 黄俊楚, 等. 一种基于等距度量学习策略的行人重识别改进算法. 电子与信息学报, 2019, 41(2): 477-483.
- Huang SY, Lu JW, Zhou J, *et al.* Nonlinear local metric learning for person re-identification. arXiv: 1511.05169, 2015.
- 张正本, 马同伟. 基于卡尔曼一致滤波和分布式协作网络控制的多目标跟踪. 电子测量与仪器学报, 2018, 32(3): 127-134.

- 6 阳小燕, 任家富, 刘长江. 基于视野分界线的多摄像机运动目标跟踪算法. 计算机工程与应用, 2016, 52(9): 208-212.
- 7 Ur-Rehman A, Naqvi SM, Mihaylova L, *et al.* Multi-target tracking and occlusion handling with learned variational Bayesian clusters and a social force model. IEEE Transactions on Signal Processing, 2016, 64(5): 1320-1335. [doi: 10.1109/TSP.2015.2504340]
- 8 王亚萌. 基于深度学习的多摄像机目标交接跟踪技术研究 [硕士学位论文]. 济南: 山东建筑大学, 2019.
- 9 Ge PH, Hu YP. Vehicle type classification based on improved HOG_SVM. Proceedings of the 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019). Dalian, China. 2019. 656-663.
- 10 张弢, 柳雨农, 任帅, 等. 基于LBP人脸纹理特征的差分直方图移位无损信息隐藏算法. 计算机应用研究, 2020, 37(6): 1774-1778.
- 11 Yan ZG, Yang F, Wang J, *et al.* Face orientation detection in video stream based on Harr-like feature and LQV classifier for civil video surveillance. Proceedings of IET International Conference on Smart and Sustainable City 2013. Shanghai, China. 2013. 161-165.
- 12 Cate H, Dalvi F, Hussain Z. DeepFace: Face generation using deep learning. arXiv: 1701.01876, 2017.
- 13 王晶晶, 刘蓉, 侯雨静. 基于DeepID的人脸检测与识别算法研究. 电脑知识与技术, 2018, 14(17): 220-221.
- 14 李淑. 基于卷积神经网络的视频人脸检测与识别. 电脑知识与技术, 2018, 14(21): 210-211, 216.
- 15 喻丽春, 刘金清. 基于改进的VGGNet算法的人脸识别. 长春工业大学学报, 2018, 39(4): 378-385.
- 16 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. arXiv: 1409.4842, 2014.
- 17 Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, inception-ResNet and the impact of residual connections on learning. arXiv: 1602.07261, 2016.
- 18 Karras T, Aila T, Laine S, *et al.* Progressive growing of GANs for improved quality, stability, and variation. arXiv: 1710.10196, 2017.
- 19 Wen YD, Zhang KP, Li ZF, *et al.* A discriminative feature learning approach for deep face recognition. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 499-515.
- 20 余成波, 田桐, 熊递恩, 等. 中心损失与Softmax损失联合监督下的人脸识别. 重庆大学学报, 2018, 41(5): 92-100.
- 21 王标. 多摄像机目标跟踪算法研究与实现 [硕士学位论文]. 武汉: 武汉理工大学, 2017.
- 22 梅树起, 原魁, 张怀相. 一种基于投影不变量的目标跟踪方法. 机器人, 2007, 29(1): 45-50.