



摘要

为解决当前严重的儿童失踪现象和低下的失踪寻回率问题，本文设计了一个基于卷积神经网络的智能儿童寻回系统。该系统通过融合两种不同的网络结构，成功将儿童识别技术与行人再识别技术相结合，实现了监控场景下的儿童识别，并通过截取特定行人图像，在视频序列中进行目标检索与匹配的功能，最后搭配应用端界面进行交互操作与结果显示，经系统验证识别准确率及各评估指标均达到 90% 以上。现监控网络通过移入此系统，可以提高公共场合监控资源的利用率，为当下儿童失踪寻回的社会难题提供一种智能高效的解决方案。

关键词：卷积神经网络, 儿童寻回, 网络融合, 儿童识别, 行人再识别



ABSTRACT

In order to solve the current serious problem of missing children and the phenomenon of low missing recovery rate, this paper proposes an intelligent child recovery system based on convolutional neural networks. The system through the fusion of two different network structure, success will be children's identification technology combined with pedestrians and identification technology, has realized the monitoring scenarios of children's recognition, and through specific pedestrian image, in the video sequence object retrieval and matching function, the collocation of application side interface interaction operation and the results show that the system verification identification accuracy and the assessment indicators are above 90%. By moving the monitoring network into this system, it can improve the utilization rate of monitoring resources in public places and provide an intelligent and efficient solution to the social problem of missing children.

Key Words : convolutional neural network, child recovery, network convergence, child identification, pedestrian re-identification



目录

摘要.....	I
ABSTRACT.....	II
1. 绪论.....	1
1.1 研究背景.....	1
1.2 研究现状.....	1
1.3 本文主要研究内容.....	2
2. 智能寻回系统设计.....	4
2.1 总体设计.....	4
2.2 算法设计.....	4
3. 卷积神经网络设计.....	6
3.1 儿童识别网络设计.....	6
3.2 行人再识别网络设计.....	7
3.3 儿童识别与再识别网络结合.....	8
3.4 数据集选择.....	10
3.4.1 儿童识别数据集.....	10
3.4.2 行人再识别数据集.....	11
4. 卷积神经网络与系统验证.....	12
4.1 神经网络训练与验证结果.....	12
4.2 系统应用验证与演示结果.....	14
5. 总结与展望.....	18
5.1 研究总结.....	18
5.2 未来展望.....	18
参考文献.....	19
致 谢.....	20



1. 绪论

1.1 研究背景

众所周知，我国儿童因迷路走失、离家出走、被拐卖等缘由失踪的现象严重，且儿童的失踪寻回率也不尽人意，迄今依旧还有大量失踪儿童不知身在何处，导致了巨大的社会隐患与家庭危害。失踪儿童寻回率的低下在于所掌握信息的不对称，归根结底就是无法在监控源头识别和检测出目标儿童，未能留下足够的线索信息进行追踪和找寻。

目前尽管大规模的监控网络被部署到各个公共场所，但巨大的监控影像数据量也使得图像信息堆积，导致信息收集、处理和分析的运算能力降低，整体的监控资源利用率不高。监控摄像作为最后记录儿童失踪前信息的有力证据，它不仅仅是提供某一特定时间和场景下的人员信息，人们更希望能够利用监控网络中多个摄像头进行协同工作，进一步对目标儿童的锁定。如何能够更迅速、更智能地识别出这些目标儿童，对儿童寻回至关重要。而侦查部门对于监控录像的调查都是基于一帧帧画面人为的逐一比对与核查，这种模式效率低、用时长、效果差，不仅会耗费大量的时间和精力，最致命的是在检索监控录像的过程中，也错过了失踪儿童寻回的黄金时间。

与此同时，随着社会科技的发展，人工智能技术得到迅猛发展，越来越多的生活场景得到更为广泛的应用，生活智能化和机器化已是大势所趋。机器学习利用概率论和统计学等原理，模拟人类的学习活动对信息进行学习，而机器学习中另一门新兴技术深度学习则通过建立神经网络，模拟复杂的人脑对信息进行分析和处理，在图像分类和识别等领域上取得了更为突出的研究成果。

因此，结合当下科技热点人工智能技术，针对当前儿童失踪的社会痛点和监控网络使用的局限性，本文将深度学习技术为出发点，切入到监控摄像的特定场景下，致力于实现一个智能儿童寻回系统。该系统涉及到目标识别和行人再识别技术，功能实现的关键在于将目标识别和行人再识别相结合。尽管当前识别技术依然受限于图像采集的优劣差异，但是很多网络结构上的优化设计不断提高着其精度和效率。本文所实现的智能系统，有效地解决了儿童识别的问题，通过理论分析和实验论证的方式校验了方法的有效性，对能够快速寻回失踪儿童以及提高儿童失踪寻回率、维护社会公共安全和稳定具有重要意义。

1.2 研究现状

纵观国内外研究现状，当前应用于监控场景下的智能识别技术主要有两个研究的大方向：行人目标识别和行人再识别，而这两种技术也主要有两种主流的实现方法：机器学习和深度学习。

基于机器学习的行人目标识别方法包括 Papageorgious 利用 Haar 特征和 SVM 向量机



进行多尺度特征图的识别^[1], Viola 和 Jones 利用级联 AdaBoost 分类器对 Haar 特征进行提取^[2]以及 Dalal 和 Triggs 提出的 HOG 特征结合线性 SVM 用于行人检测的方法^[3]。

基于深度学习的行人识别方法随着 SSD、Faster R-CNN 以及 YOLO 等一批结构复杂、设计合理的网络的出现,得到了快速应用,同时也有突出的成绩。著名的有 Yining Li 在卷积神经网络的基础上提取深度特征,结合上下文语义信息进行行人属性识别^[4],得到了很高的准确率。

行人再识别技术最早是用特征表示和度量学习等传统方法来实现的。特征表示是将目标行人表示为一个判断力强的特征向量。经典的有彭志勇等人提出的一种基于 HSV 模型和特征点匹配的行人再识别算法^[5]以及 Li 等人将定义的 11 种衣着属性传递给 SVM 向量机训练特征分类器,将特征进行融合后来表征行人^[6]。而度量学习则是计算样本间的相似度距离。Kostinger 等人提出的 KISSME 算法,通过似然比检测来衡量样本间的差异程度,再计算得到马氏距离^[7]则是度量学习范畴。

结合深度学习的行人再识别方法主要有 Lin 等人使用的 ResNet 作为基础神经网络对 Market1501 数据集进行训练^[8],而 Li 等人设计了上下文感知网络来学习行人的整个身体和某些部位特征,最后将二者进行特征融合进行识别^[9]。

而针对于儿童的识别技术,较为接近的是 Rothe^[10]等人提出了在人脸识别基础上进行年龄识别的实现方案,根据 IMDB-WIKI 数据集训练 VGG16 网络,先检测出人脸,利用脸部到纹理边缘等特征信息来判别所属的年龄,年龄误差能够控制在 1 岁以下。后面涌现的很多相关研究,也都是基于人脸识别结合年龄识别的技术方案实现的。

而在监控场景下,很难得到清晰的人脸图像。现有的行人目标识别技术没有根据特征再细分出儿童这一特定群体,而已有的儿童识别又是基于人脸实现的,与我们的实现方案差别较大,缺乏在行人目标识别基础上实现儿童识别的相关研究成果。而且在现阶段成果中,行人目标识别和行人再识别技术大多数是分开进行研究的,作为一个独立的技术分支,至今仍缺少二者相结合的相关实践。将这两种技术进行结合,并对某一特定群体进行研究,有助于衍生出一系列应用性更强的系统和应用。

1.3 本文主要研究内容

本文对应用于监控场景下的智能儿童寻回系统进行设计与实现,基于目标识别技术设计一个儿童识别的卷积神经网络用于儿童与成人特征的分,然后基于行人再识别技术设计一个行人图像检索的网络结构用于特征的计算与匹配,再将二者相结合,搭建具备识别与检索功能的系统。本文具体的研究内容包括以下几个方面:

(1) 监控视觉下的行人多为距离较远的拍摄角度,因此是无法辨认出行人清晰的脸部属性的。基于远视觉的行人识别,设计并优化出适合于小目标的神经网络结构。在局部结构上进行改进探索,并针对数据集进行筛选与清洗,探索出图像标注的技巧,进一步提高识别出儿童这一特定群体的准确率。



(2) 在搭建好的儿童识别与再识别网络的基础上，探索网络的结合技术。实际上这两种技术都是监控识别技术下的两个重要分支，可以说是相互联系和密不可分的。尝试将行人目标识别与行人再识别进行融合，实现在一个卷积神经网络下实现儿童识别和再识别的结合，设计一个完整的算法流程以保证整个系统在运转过程中的流畅和快捷。

2. 智能寻回系统设计

2.1 总体设计

智能儿童寻回系统总体由应用端窗口、卷积神经网络和监控摄像三部分组成。三部分作为独立的模块存在，相互联系构成一个闭环系统。应用端是基于 Python 开发的交互界面，主要用以监控物料的输入以及结果的显示。卷积神经网络作为整个寻回系统的核心，负责对待测视频进行运算处理，并可通过加载监控摄像头来获取画面图像进行识别。监控摄像作为唯一的硬件设备，根据运载的计算机性能进行选择，满足基本的帧率要求。具体的系统设计如下图所示。

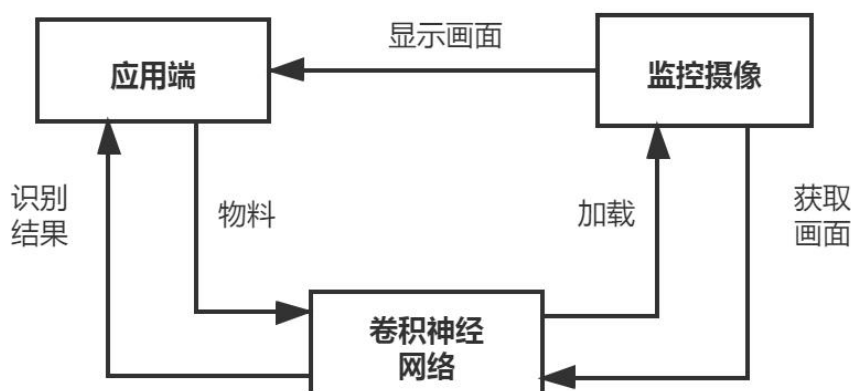


图 2.1 系统总体设计框图

由图 2.1 可看出，系统的数据流是双向传输的。信息的双向并行交换驱使系统能够实时地运行。应用端和监控摄像分别作为系统的两个图像输入端，可支持视频影像的检索和监控摄像的实时识别两种模式，方便系统的在线与离线使用。

2.2 算法设计

整个系统搭配卷积神经网络实现有序稳定的运转是整个算法设计思想的核心，为了实现儿童检测与行人再识别于一体，本小组设计了如下图的实现流程。

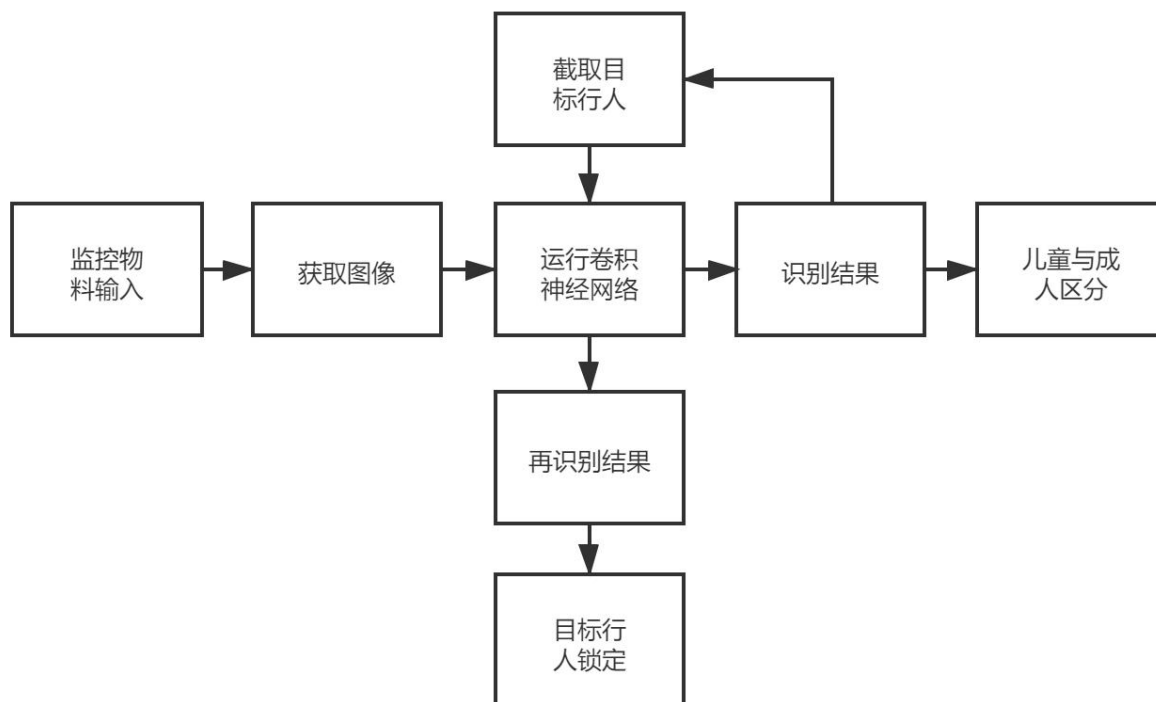


图 2.2 算法设计流程图

原始的待检测监控视频输入到系统中，加载卷积神经网络模型。对获取到的每一帧图像信息，神经网络的卷积结构进行特征的提取而后进行特征运算和分类，在画面中识别出儿童与成人，作为后面再识别系统的输入部分。根据显示的识别结果，需要人为对要检索的目标行人图像进行截取。神经网络分别对截取得到的目标图像和监控视频中的行人图像进行特征提取和特征计算，匹配出同一个目标行人，显示出再识别的结果。

整个设计的巧妙之处是神经网络的复用，两次识别都需要经过同一个神经网络结构层。网络结构层中包含了两种不同的网络模块，神经网络根据输入的不同，决定驱动不同的功能模块，控制最后显示结果的不同。

3.卷积神经网络设计

卷积神经网络是一种前馈型网络，包括输入层、卷积层、激励层、池化层和全连接层等结构，各层的神经元都只与上层相连，接收上一层的结果输出给下一层^[11]。卷积神经网络最大的特点是权重共享，利用同一个卷积核对特征进行提取，使得能够共用相同的参数，大大减少了参数量，简化网络结构。

3.1 儿童识别网络设计

行人检测作为目标识别中的一个关键技术，其针对场景画面做检测，在众多的目标里识别出行人目标。现较为成熟的神经网络结构是 Faster RCNN、SSD 和 YOLO 系列。由于设计的监控系统要求检测实时性，且儿童识别也比一般的行人识别要更为精确。YOLOv3 结构^[12]，在识别速度和准确率等性能上都能很好满足系统要求。它将识别问题统一为一个回归问题，利用端对端训练技术直接对输入处理进行输出，提升了整体的目标检测运算速度，使实时性成为可能。因此儿童识别网络，在 YOLOv3 神经网络结构的基础上，进行适度的优化设计，提高区分目标的精度。YOLOv3 具体网络结构如图 3.1 所示。

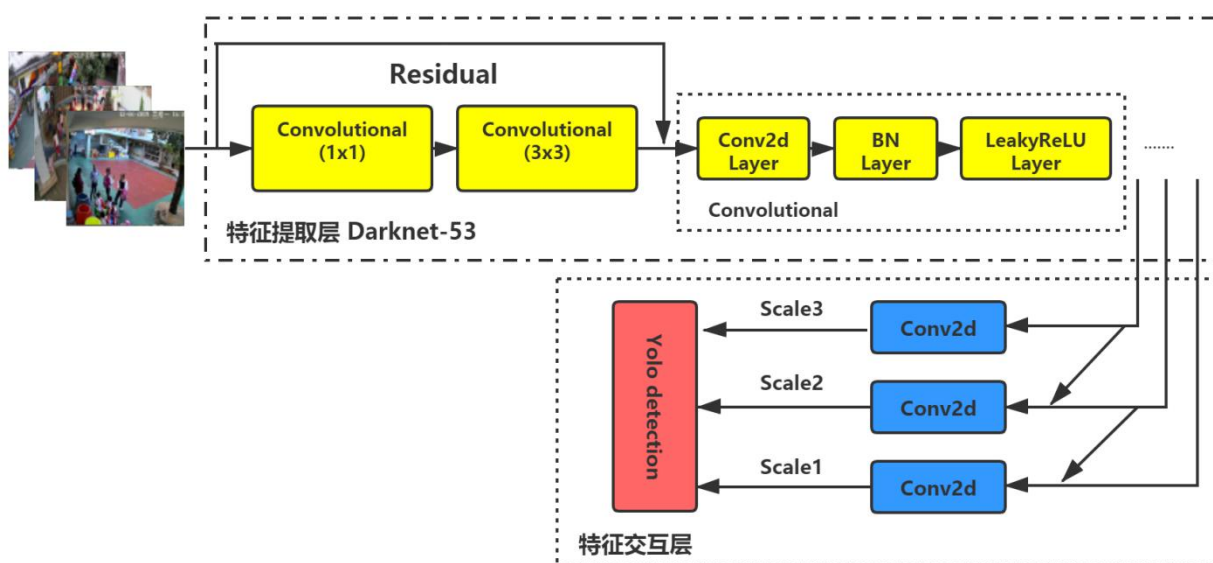


图 3.1 YOLOv3 网络结构图

YOLOv3 网络由特征提取层和特征交互层两部分构成。YOLOv3 前 74 层包含了 53 个相同的 1×1 和 3×3 的 Convolutional 卷积结构。每个 Convolutional 模块包括卷积层、BN 层和 Leaky ReLU 层，中间添加了大量的 Residual 结构，进行 shortcut 连接，作用是最优性能的维持。网络的 75 到 105 层则主要用于特征交互，最后输出三种不同尺度的预选框。每个尺度内，都是通过卷积操作实现特征图之间的局部特征交互。

相比较于其他网络结构，YOLOv3 的优化在于多尺度预测和更好的基础网络。YOLOv3 边界框使用聚类的方式得到多种尺度，可以适应不同大小的预测目标。而 Darknet-53 作为

YOLOv3 特征提取层的核心基础网络，由于引入了残差结构，其性能大大得到了提升。

本系统定位于监控视角下的小目标，为了达到更佳的实验效果，在原基础上进行了几处优化设计：

(1) 用于识别的行人属性除了衣着类型、颜色、身形等外，还需要分析目标的尺寸，合理配置预测框大小。这是利用标注数据的预测框辅助辨别的方法。在同一个图像画面内，成人与小孩之间的尺寸存在明显的高度差，二者预测框之间的比例关系可以充当区分儿童与成人的轮廓特征。在对数据进行标注时，注重目标的尺寸选择，尤其是不同类别预选框尺寸的落差对比。

(2) 网络中初始候选框的选择会直接影响训练速度和识别精度。利用 K-means 聚类算法对目标初始候选框的个数和尺寸维度进行聚类分析，选取优化后的聚类框作为网络的初始候选框^[13]。

3.2 行人再识别网络设计

行人再识别是利用计算机视觉技术判断图像或视频序列中是否存在特定行人的技术。对一台监控下某一特定行人，检索到该目标在其他监控下出现的所有图像，实现跨设备的行人检索。

目前行人再识别网络中一个高性能的 baseline^[14]是以 ResNet50 作为基础网络实现的。其常用的评估指标是 rank-n，表示检索结果中置信度最高的 n 张图有正确结果的概率，数值越大匹配度越高。行人再识别网络结构如图 3.2 所示：

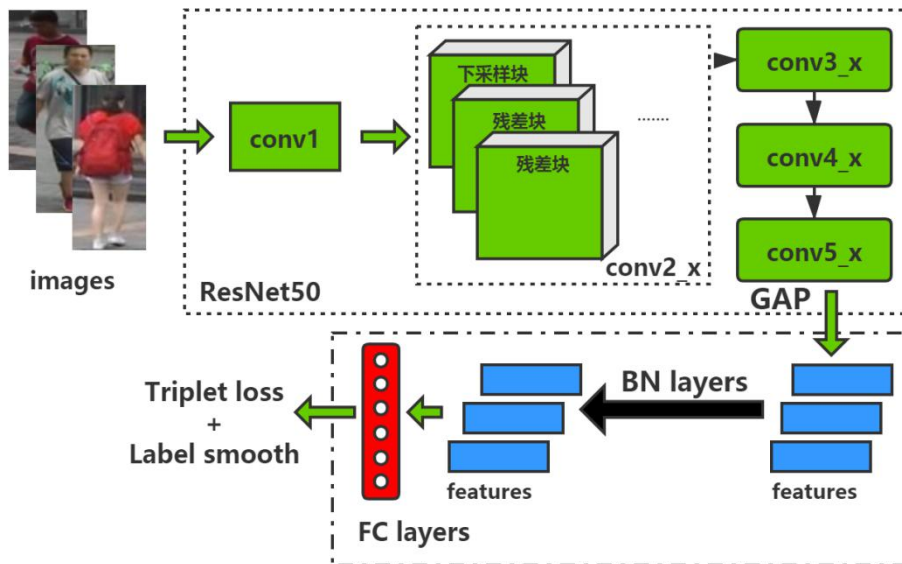


图 3.2 行人再识别网络结构图

ResNet50 包含有 1 个卷积层和含相同结构的 4 个模块，每个模块都由 1 个下采样块和



2 个残差块构成，对输入的特征图进行下采样，使长宽都缩减为原来的 $\frac{1}{2}$ ，不断减少网络的参数量。残差块利用 shortcut 连接进行残差学习，保持了输入与输出特征图尺寸的一致。而残差结构的设计，正是前面 YOLOv3 和 ResNet 系列巧妙之所在。

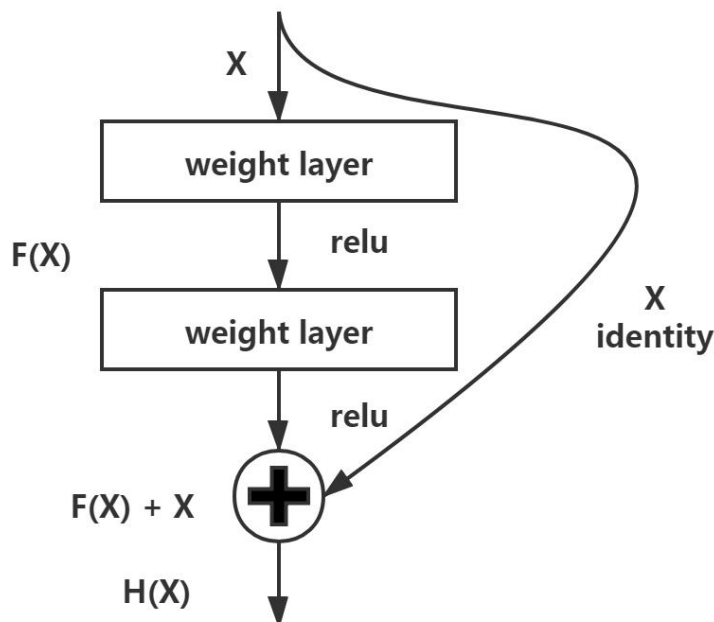


图 3.3 残差结构示意图

随着网络深度增加，网络性能可能不会随之提升。正如图 3.3 所示，如果网络最优的输出结果就是输入 x ，那么一般网络的优化目标就是 $H(x)=x$ ，而加入了右边的 shortcut 连接后，网络输出表示为：

$$H(x) = F(x) + x \tag{3.1}$$

$F(x)$ 为网络的残差，则此时网络的优化目标就转化成：

$$F(x) = H(x) - x \tag{3.2}$$

可以看出，只要将残差值优化成 0，就相当于构成了一个恒等映射 $H(x) = x$ ，网络性能就不会下降。网络的残差值优化成 0，要远远比将输出优化成输入本身容易，因此能够更好维持上一层的网络性能，避免了深层网络的梯度消失和退化。

3.3 儿童识别与再识别网络结合

目前行人再识别技术主要有两种应用场景：一是测试的输入是一对行人图像，输出二者的相似度来判断是否为同一个目标；二是输入单个目标行人，提取其特征，在视频序列中进行匹配。而实际中我们的儿童识别就是基于视频序列进行的，所以考虑到系统的一致

性，行人再识别技术也将定位于检测的同步显示。

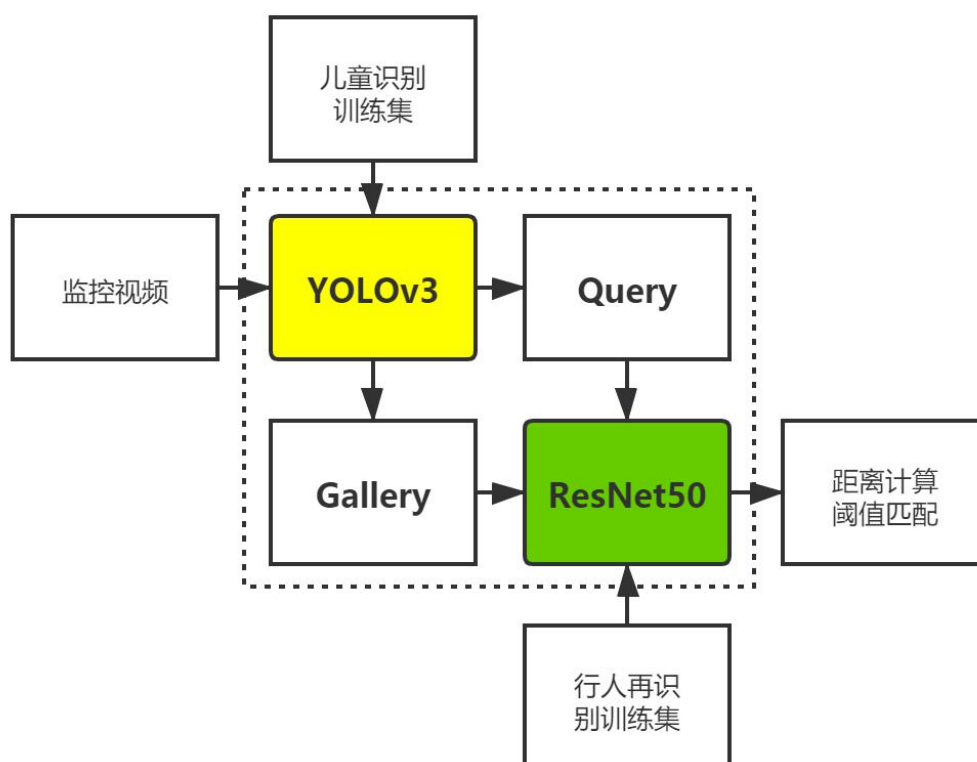


图 3.4 儿童识别与再识别网络结合示意图

整个网络架构设计如图 3.4 所示。两个网络的结合是单向连接的，信息只能沿着一个传输方向一步步被处理，总体经过了特征提取、特征识别、特征计算到特征匹配四个过程。视频数据先经过 YOLOv3 网络的处理，其起到的是检测与预测的作用，而 ResNet50 则根据 YOLOv3 处理后得到的预测结果进行计算和匹配。一个网络的输出结果，刚好是另一个网络的输入。

利用不同的数据集进行训练，使训练后的网络模型具备良好的特征识别和检索的能力。当有监控视频流传入至网络输入层后，便重载神经网络模型开始处理。深层网络将成千上万的参数一层一层往下传递，最后在每一帧图像里，输出儿童和成人两个不同的类别，根据识别到的结果画出预测框。每一个预测框内就包含有一个行人目标。这些只含有行人的预测框图像，正是 ResNet50 检测所需要的图样。将所有目标框存入 Gallery 中，先计算 Gallery 内各行人的特征向量，再计算 Query 待查询图像的特征向量与 Gallery 行人特征向量的欧式距离。

这里引入欧式距离是参考行人再识别的传统方法度量学习。距离度量常用于衡量个体特征间的相似程度，距离越远表明个体差异越大。而欧式距离是常见的距离度量方法之一，用于衡量空间各点之间的绝对距离。

n 维空间的欧式距离可表示为：



$$dist(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.3)$$

而在神经网络中，数据点是以向量矩阵形式存在的，矩阵的欧式距离可表示为：

$$dist = \sqrt{\begin{pmatrix} \|P_1\|^2 & \|P_1\|^2 & \cdots & \|P_1\|^2 \\ \|P_2\|^2 & \|P_2\|^2 & \cdots & \|P_2\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|P_m\|^2 & \|P_m\|^2 & \cdots & \|P_m\|^2 \end{pmatrix} + \begin{pmatrix} \|C_1\|^2 & \|C_1\|^2 & \cdots & \|C_1\|^2 \\ \|C_2\|^2 & \|C_2\|^2 & \cdots & \|C_2\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|C_n\|^2 & \|C_n\|^2 & \cdots & \|C_n\|^2 \end{pmatrix} - 2 \times PC^T} \quad (3.4)$$

转化为一般的欧式距离表示就是：

$$dist = \sqrt{P^2 + C^2 - 2 \times P \times C} = \sqrt{(P - C)^2} \quad (3.5)$$

最后 Query 与 Gallery 样本间的差异我们舍弃 Softmax 和交叉熵而引入 Triplet loss^[14] 来定义。Triplet loss 就是随机从训练集中选取一个样本 Anchor，再分别选取一个和 Anchor 同一类的样本 Positive 和不同类的样本 Negative，构成一个三元组。通过学习，让同类的特征表达之间的距离尽可能小，而异类的特征表达之间的距离尽可能大，并且使二者距离之间有一个最小的间隔 α ，可表示为：

$$\|f(x_i^a) - f(x_i^p)\|_2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2 \quad (3.6)$$

这个最小间隔，就是正负样本间的距离，也是程序用来判断是否为同一 ID 行人的阈值。最后再利用 Label smoothing 技术，对标签进行正则化处理，增大类间距离，减小类内距离，缓和网络的过拟合程度，公式表示为：

$$q_i = \begin{cases} 1 - \frac{N-1}{N} \varepsilon & (if \ i = y) \\ \frac{\varepsilon}{N} & (if \ i \neq y) \end{cases} \quad (3.7)$$

正是因为引入了 Triplet loss 和 Label Smoothing，再识别问题中类间相似、类内差异的难题才很好得到了解决。

3.4 数据集选择

两种不同的识别技术，依靠两种不同的网络结构实现，因此也需要两种不同的数据集。在数据集的选择上，除了依赖于已公开的数据集外，还通过网络爬取方式和图像变化等数据增强方式扩充了整个训练库。

3.4.1 儿童识别数据集

现有的 COCO、PASCAL_VOC 等目标识别数据集中，行人是作为众多目标中的一个分类，儿童与成人没有分开标注。而且我们所研究的系统应用于监控场景下，监控摄像普

遍像素不高，均为远景拍摄，行人在画面中所占比例较小，均为小目标样本，与里面近景高清大头照差异较大，因此不选用现有行人数据集充当系统的训练集。

通过在互联网各大网站上搜索和在监控录像中截取，最后收集到 4032 张相关数据，其中儿童与成人正样本共 2032 张，负样本 2000 张，共标注了儿童 8157 个，成人 5830 个。这些照片均为实际或模拟实际的监控场景。其中，3226 张作为神经网络的训练集，806 张作为测试集。

3.4.2 行人再识别数据集

目前大部分的行人再识别技术训练和测试都是基于在公开数据集上开展的，这些数据集均由若干个不同角度的监控摄像头拍摄得到，包含了众多截取好的行人图像，且对行人图像做好了标注处理。本文使用了 Market1501、CUHK03 和 MSMT17 三个数据集进行试验，三个数据集的部分行人图像样本如下图所示。

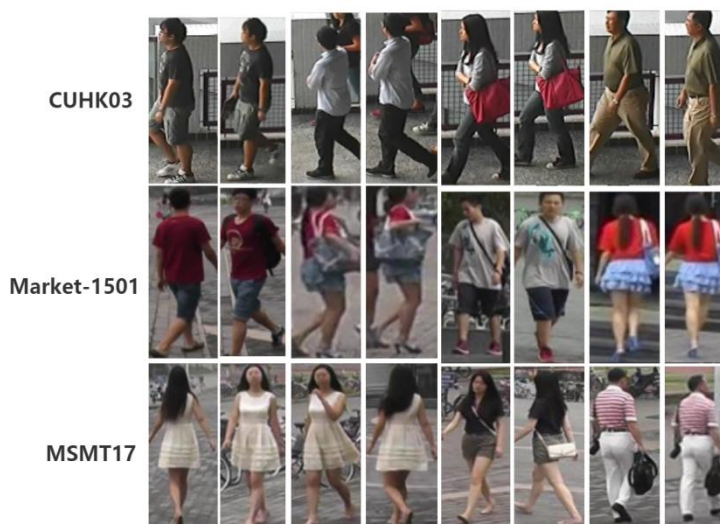


图 3.5 数据集的行人图像样本

为了提高行人再识别的跨域效果和泛化能力，更好地适应于现实中多变的生活与生产场景，我们对这三个数据集进行了联合训练。将 Market1501、CUHK03 和 MSMT17 共 7069 个行人合计 171822 张图片，统一转化为 Market1501 的命名格式，作为神经网络的训练集和测试集。这大大增加了数据集数量，使其囊括更多的实际画面和真实场景。

4. 卷积神经网络与系统验证

4.1 神经网络训练与验证结果

用上面处理好的数据集对设计的卷积神经网络进行训练与测试，针对训练过程记录了若干组不同情况的数据，如表 4.1 所示。

表 4.1 神经网络训练数据表

训练类型	Train loss	Test loss	Epoch time/s	mAP/%
原混合数据集	9.83	12.46	241.2	88.69
剔除现行人数据集	13.08	15.17	195.0	91.34
数据清洗后	12.74	14.50	184.8	93.05
网络优化后	10.31	13.95	89.4	93.18

训练一共分为四种不同的类型。在数据集的选择上，剔除混合数据集中原现有行人数据集，只保留模拟监控场景的数据，实验结果中训练损失下降到 13.08，数值上略有增大，原因在于数据集中存在许多背景相似的图像，导致信息量不足，无法进一步学习到更多特征，也致使训练的收敛速度变慢了，需要训练更长的时间才能达到相同的性能。但是经测试，整体的准确率却有所提升，表明系统正趋于数据集的特征表达，更有理由说明现有行人数据集不符合我们的系统要求。

而经过更进一步的数据清洗和网络优化后，多类别目标检测指标 mAP 最优时达到 93.18%，模型总体的平均精确度表现出了极高的性能，而儿童和成人的识别准确率也分别高达 94%和 92%，训练速度和准确率更是明显提升了约 54.2%和 5.1%，数据如图 4.1 所示。

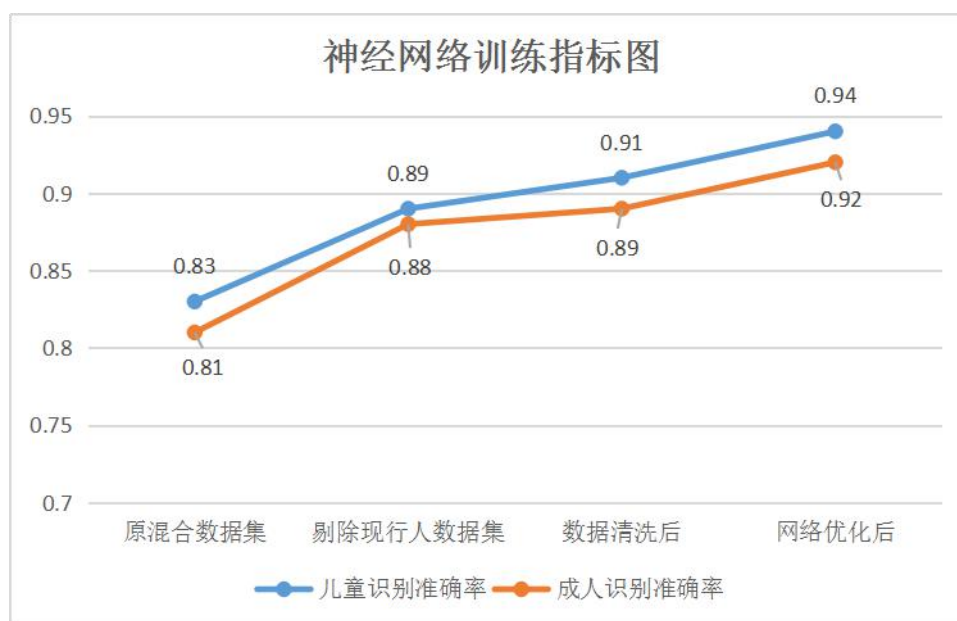


图 4.1 神经网络验证准确率数据图

儿童识别准确率和成人识别准确率反映着神经网络的内部性能，但是对于网络结构以及整个外部系统而言，是否能够在画面中准确识别出对应的类别才是有效评估效果的唯一指标。因此，为了能够更直观和清晰地验证网络实际效果，忽略准确率的动态起伏变化，除了识别准确率外，我们还定义儿童与成人的识别度和区分率这两个新概念作为评估指标。其具体的计算公式如下：

$$\text{识别度} = \frac{\sum \text{正确识别某一类别数量}}{\sum \text{画面中该类别总数量}} \quad (4.1)$$

$$\text{区分率} = \frac{\sum \text{正确识别对应类别}}{\sum \text{儿童数量} + \sum \text{成人数量}} \quad (4.2)$$

这里的识别度是指在一个监控画面中正确识别出某一类别的数量占画面内该类别总数的比例，而区分率是指在每一帧视频序列中能够正确区分出儿童和成人的精确度。

随机挑选 100 张混合着儿童与成人的监控场景图像，对神经网络进行验证。验证的结果如图 4.2 所示。

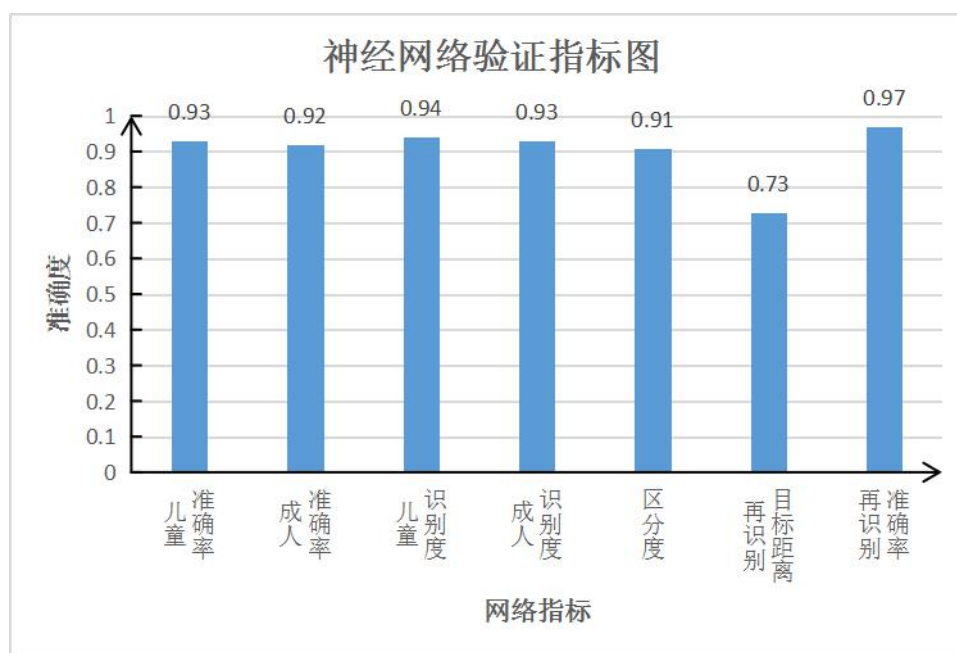


图 4.2 神经网络验证指标图

可以清晰看出，在儿童识别网络中，准确率和识别度均达到 90% 以上的正确率，而儿童与成人的区分率为 0.91，在混合人群的画面里，已基本具备精准识别和判断的能力。挑选其中一张实际的测试效果图，如图 4.3 所示。



图 4.3 实际测试效果图

在对行人再识别网络验证中, Query 与 Gallery 样本间计算得到的欧式距离平均为 0.73, 结果要远远小于类内距离阈值, 同类间的相似度分布较为接近。而再识别准确率为 0.97, 除去视频画面中一些光线暗淡、截取的图像不清晰、不完整和目标行人尺寸过小等极端情况造成无法检索到对应目标外, 一般情况下都能够再识别正确, 整体效果显著。

4.2 系统应用验证与演示结果

为了整个系统易于交互与显示, 我们将卷积神经网络搭载在应用端的窗口界面上, 整个智能儿童寻回系统以应用端方式呈现, 更有利于使用和推广。

整个图形界面是利用 Python 下的 tkinter 工具库设计的, 结合 OpenCV 图像处理技术, 将儿童识别网络和行人再识别网络都搭建成端对端的网络模型。在同一种编译环境下, 系统模块间的调度和交互速度可以达到最优。

本文设计的智能儿童寻回系统的主界面如图 4.4 所示。主界面包含可供用户操作的功能按键以及识别结果的展示窗口, 包括儿童识别和行人再识别两个主体功能, 其中儿童识别支持在线监控和视频监控两种识别模式。



图 4.4 智能儿童寻回系统应用界面

正常的系统操作是先进儿童识别的功能使用，选定待识别的监控视频，经过短暂的模型加载和计算时间后，应用界面上会显示出识别后的结果。儿童与成人两种不同的类别，会用不同颜色的预测框以示区别，如图 4.5 所示。



图 4.5 智能儿童寻回系统识别界面



然后在识别显示的过程中，利用应用窗口中的截取图像功能，对识别出的儿童或成人图像进行截取作为待检索的目标行人，截取得到的图像会统一保存在一个特定文件夹下，方便后面选择，效果如图 4.6 所示。



图 4.6 智能儿童寻回系统图像截取界面

最后是使用行人再识别功能。选择刚才截取的图像作为检索图像，在视频序列中判断是否含有特定目标，运算后的结果也会同步显示在窗口内。再识别结果如图 4.7 所示，待检索的目标已经被准确找出，通过其他标识标志。



图 4.7 智能儿童寻回系统行人再识别界面

至此，整个系统搭载着识别与再识别功能一脉相连，识别与检索能力相辅相成，搭配监控系统的使用，一整套实现流程下来，能够在海量的监控资源中一步步筛选出所需要的目标图像，极大地辅助了失踪儿童的高效寻回，大大节省了人力物力财力。



5.总结与展望

5.1 研究总结

本文的核心是实现儿童识别与行人重识别技术的结合，准确而高效将 YOLOv3 与 ResNet50 网络结构融合。由于系统应用于监控场景下，本文将 YOLOv3 网络结构进行适度的优化设计，使其更适用于小目标的检测。本文着重介绍两种技术所采用的神经网络结构，以及研究儿童识别与再识别网络结合后在系统中的处理机制和实现阶段，最后通过实验验证表明了该网络结构的有效性。验证数据表明，儿童与成人的识别准确率和区分率均达到 90%，行人重识别的准确率达到 95%，能够成功将儿童与成人进行区分，且在跨监控设备的行人检索中也能精确定位到目标。两个功能搭配起来使用，可以辅助失踪儿童的寻回，提高整体的失踪寻回率，具有重大的社会意义和实际应用价值。

5.2 未来展望

本文致力于将儿童识别与再识别技术相结合，提出了一种系统层面上的实现方案，取得了较好的效果，但是仍然存在一些问题在未来仍待解决：

(1) 神经网络对其他实际监控场景的普适能力较弱

本文训练所得的儿童识别准确率和成人识别准确率虽然较高，但是神经网络对其他监控场景的普适能力较弱。虽然很大部分原因是出自数据集的收集上，训练库的数据单一场景较多，网络模型缺乏丰富多样的监控场景学习，导致在其他场景的识别中会有所失误。但是除了需要进一步扩充数据集外，也还需要对神经网络作进一步的结构改善和优化设计，不断探索更多的改良技巧和方法，使其在对小目标的特征提取上能有更出色的效果。

(2) 真实监控场景下行人属性部分难以区分

在真实的监控场景下，摄像头所采集到的行人图像，因为光照不同、设备本身清晰度不同、拍摄角度不同等，会出现许多难以辨别的情况，再加上局部遮挡、背景干扰、行人姿态不同、衣着外观相似和空间位置不对齐等各式各样的图像问题，极易容易导致目标识别和行人再识别出现误判和错判的情况。



参考文献

- [1] Papageorgiou C , Poggio T . A Trainable System for Object Detection[J]. International Journal of Computer Vision, 2000, 38(1):15-33.
- [2] Viola P A , Jones M J . Rapid Object Detection using a Boosted Cascade of Simple Features[C]// Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001.
- [3] Dalal N , Triggs B . Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.
- [4] Li Y , Huang C , Loy C C , et al. Human Attribute Recognition by Deep Hierarchical Contexts[C]// European Conference on Computer Vision. Springer International Publishing, 2016.
- [5] 彭志勇,常发亮,刘洪彬,别秀德.基于 HSV 模型和特征点匹配的行人重识别算法[J].光电子·激光,2015,26(08):1575-1582.
- [6] Li A , Liu L , Wang K , et al. Clothing Attributes Assisted Person Reidentification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(5):869-878.
- [7] Kstinger M , Hirzer M , Wohlhart P , et al. Large Scale Metric Learning from Equivalence Constraints[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012.
- [8] Lin Y , Zheng L , Zheng Z , et al. Improving Person Re-identification by Attribute and Identity Learning[J]. 2017.
- [9] Li D , Chen X , Zhang Z , et al. Learning Deep Context-aware Features over Body and Latent Parts for Person Re-identification[J]. 2017.
- [10] Rothe R , Timofte R , Gool L V . DEX: Deep EXpectation of Apparent Age from a Single Image[C]// IEEE International Conference on Computer Vision Workshop. IEEE Computer Society, 2015.
- [11] 吴碧程,邓祥恩,张子憧,唐小煜.基于卷积神经网络的智能垃圾分类系统[J].物理实验,2019,39(11):44-49.
- [12] Redmon, Joseph & Farhadi, Ali. (2018). YOLOv3: An Incremental Improvement.
- [13] 鞠默然,罗海波,王仲博,何淼,常铮,惠斌.改进的 YOLO V3 算法及其在小目标检测中的应用[J].光学学报,2019,39(07):253-260.
- [14] Luo H , Gu Y , Liao X , et al. Bag of Tricks and A Strong Baseline for Deep Person Re-identification[J]. 2019.



致谢

匆匆四年过去，本科阶段的学习即将宣告结束。毕业论文的完成，也为自己最后的学生时代画上了一个完满的句点。回首来时的路，转瞬即逝而又百感交集。

感谢我的毕业论文导师——李军老师。从论题分析，到论文查阅，再到项目实现，最后到论文撰写，老师都给予了我很多的指导帮助以及改进意见。大学期间李军老师两次担任我的导师，我也深受老师严谨治学、求实创新的科研精神影响！我很感谢！

感谢我的父母及家人，这些年给予了我莫大的物质和精神上的支持。多亏了他们，13年背井离乡的求学之旅才能顺利完成。以前总想着能尽快独当一面，而现在，往后的担子就能交于我！我很感恩！

感谢我的舍友同学及兄弟们，在最喜悦的时候，有你们一起疯狂，而在最艰难的日子里，亦有你们的陪伴。四年来最珍贵的莫过于这种不离不弃的兄弟情义！我很庆幸！

感谢一路走来给予过我帮助和感动的人，正是别人这些一点一滴的善意，才让我即便很多次也仍有勇气往前，同样也时刻提醒着自己：但行好事，莫问前程。我很幸运！

感谢这四年来的一些匆匆过客，尽管终究只是旅途中彼此的过客，但曾经有过的一些交集，或许也会是往后一个很美好的回忆！我很满足！

最后，感谢陪伴了我四年的母校——华南师范大学。在这里走过的所有路、遇到的所有人、经历的所有事、获得的所有成绩、感受的所有思绪、享受的所有青春、领悟的所有成长，弥足宝贵且倍感珍惜。我很幸运，同样亦很知足！

也许，我并不知道路途的前方究竟有什么，但是，我们还是会迈出步伐。因为，我们仍在旅途之中。

张子憧

二〇二〇年四月三十日