



计算机视觉在智慧安防中的应用

陈志宏, 王明晓

(中国电信股份有限公司上海分公司, 上海 200120)

摘 要: 对人工智能领域中的计算机视觉和深度学习技术的发展进行了简要概述, 阐述了两者之间的关联, 并结合 5G 技术对其未来的发展趋势做了一定的展望。剖析了某小区人脸抓拍的实际案例, 重点介绍了计算机视觉技术在城市治理智慧安防行业中的应用, 包括流程设计、业务逻辑、算法原理等, 希望能够为相关领域的专家学者提供合理化建议与新的思路。

关键词: 人工智能; 计算机视觉; 深度学习; 智慧安防

中图分类号: D631.4

文献标识码: A

doi:10.11959/j.issn.1000-0801.2021195

Application of computer vision in intelligent security

CHEN Zhihong, WANG Mingxiao

Shanghai Branch of China Telecom Co., Ltd., Shanghai 200120, China

Abstract: The development of computer vision and deep learning technology in the field of artificial intelligence were summarized, the relationship between them was expounded, and its future development trend was proposed combining with 5G technology. The actual case of face capture in a community was analyzed, and the application of computer vision technology in the intelligent security industry of urban governance was focused on, including process design, business logic and algorithm principle, hoping to provide reasonable suggestions and new ideas for experts and scholars in related fields.

Key words: artificial intelligence, computer vision, deep learning, intelligent security

1 引言

近年来, 计算机视觉这个词逐渐开始流行, 1982 年 Marr^[1]提出了视觉计算的理论和方法, 认为人类的视觉主要是复原三维场景的可见几何表面, 提出可以通过计算的方式实现二维图像到三维结构的复原, 这标志着计算机视觉正式成为一门独立学科。随着社会经济的不断发展, 人们生

活质量逐渐提高, 新兴技术日新月异, 城市安防的需求也开始不断攀升, 对安防程度和安防效率提出了更高的要求。计算机视觉技术的出现正好可以填充这一技术的不足, 在智慧安防领域, 可以用机器代替人眼对目标进行识别、跟踪和测量, 并进一步做图像处理, 使其成为更适合人眼观察或传送给仪器检测的图像^[2]。

与此同时, 国家政策对深度学习、计算机视

觉等方面的支持,更是打造了一个比以往任何时代都有利于技术蓬勃发展的环境。计算机视觉正逐渐成为人工智能领域最重要的研究方向,因为视觉是人体获得信息最多的感官来源,曾经有实验证实视觉信息占人类获得全部信息的 83%^[3]。如今,各种新产品和新解决方案不断提出,行业之间的壁垒逐渐开始淡化,随着人工智能技术的不断演进、5G 技术的快速发展,计算机视觉行业生态在这些技术力量的驱动下,将会为城市治理和智慧安防领域注入全新活力。

2 深度学习与计算机视觉

2.1 前深度学习时代

在深度学习算法出来之前,计算机视觉算法大致可以分为 5 种方式,分别是特征感知、图像预处理、特征提取、特征筛选和推理预测与识别。而在早期的学习过程中,占优势的统计机器学习群体对特征是不大关心的^[4]。

计算机视觉也因此被一部分学者认为是机器学习在视觉领域的应用,在深度学习出现之前,计算机视觉都需要人们自主设计特征感知、提取、预处理和筛选的部分,但这无疑是一个相当困难的挑战。以往的识别方法,主要是把特征值的提取和用于分类存放特征值的容器分开设计,直到两者需要应用时再融合。这样的操作意味着首先需要有一个特征表达式或者特征值提取过程,再把表达收敛到的特征植入学习算法中进行分类学习。

举例来说,在早期的经典指纹识别算法中,主要在指纹图像上设立具有特殊几何属性的特征点,通过特征值匹配程度判断两个指纹图像是否是同一个。这种方式既需要丰富的经验进行特征值的人工设计和参数调优,同时还需要对领域内容非常了解。另外,上述方法不仅需要人工设计特征,还需要能够适应设计特征值的分类容器。这意味着,只有特征与分类容器契合,才能够达

到理想的结果。

2.2 深度学习与 5G 的到来

上述困难一直持续到端到端学习(end-to-end learning)概念在深度学习这一学科中被提出。这个概念指的是面向特定类型的特定对象,计算机能够通过一定的方法,自主学习什么种类的特征需要被查找,它能够描述对象最具显著特点的特征值。换句话说,通过深度学习端到端学习理念的网络模型,能够支持探索图像类别中的底层模式。特征提取和深度学习的方法差异如图 1 所示。

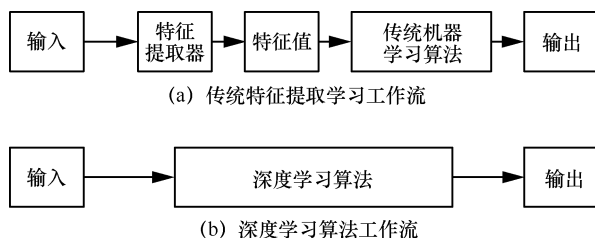


图 1 特征提取和深度学习的方法差异

在深度学习出现之后,理论上可以不通过人工筛选特征值、定义特征值,同时也不需要人工来描述这些特征值,算法能够代替人工完成这类工作。最具有代表性的是卷积神经网络。卷积是一个函数与一个函数通过翻转或平移等一系列相互关系,计算出新值,然后在连续空间进行积分运算和在离散空间进行求和运算的过程。实际上在计算机视觉学科中,卷积通常被视为一种抽象的过程,是将给定范围内的信息通过计算抽象出来的过程。

举例来说,对同一个对象,通过学习多个不同的卷积与函数对范围内的图像信息进行计算。根据需要可以使用不同的方法统计,如重点计算边缘位置,或者重点计算中心区域,这会使得结果变得更加多样化,从而同步实现多个函数的累积和。可以看出,采用这种方式的效率要远远高于之前人工设计特征提取模型识别图片的方式。



如此看来,卷积网络和深度学习应用在计算机视觉中也不算特别复杂,且在1998年时便已经有雏形了,但是直到最近才开始逐渐占据主流,原因如下。

其一,大数据技术的限制。如果卷积神经网络所设计的复杂程度不够,其真正应用所呈现的效果有可能不如传统特征提取模型。如果设计的复杂,就需要配套海量数据训练。人们很难看到已训练模型的内在机制,手动调参也相当困难,因为深度学习模型有数百万个参数——每个参数在训练过程中都需要调整。从某种意义上来说,深度学习模型是一个黑匣子,只有不断地向其中输入数据,持续训练,它才能够得以应用,而在大数据技术不够成熟的情况下,没有办法提供足够的训练数据,难以被广泛应用。

其二,运算能力的限制。卷积神经网络对机器的计算资源要求非常高,会进行海量的重复计算,在处理器不具备多核且计算能力不足的年代,是很难进行较深的卷积训练的。如今,计算机的GPU能力大大增强,结合大数据训练的卷积神经网络模型也不再遥不可及。

从另外一个方面看,5G的到来打破了这些条件的制约,其高速率、低时延以及广连接的特性,除了带来万物智联,也让过去各类难以实现的智能化场景成为可能。伴随5G建设的发展,高速率应用场景的要求渐渐被解决,各类边缘计算的技术更新又释放了算力资源的制约,可以预见,物联网最大的瓶颈将会被5G打破,5G会迅速催化物联网行业的更新迭代;而5G带来的物联网突破又会引起大数据行业的变革,因为大数据行业中90%以上的数据是由物联网中的智能感知设备所产生的;而大数据又是人工智能发展的三大要素之一,更多的数据会使人工智能更准确、更智能。

正因如此,深度学习与5G的到来,为计算机视觉技术带来了空前的发展机遇与进步空间。

3 计算机视觉在智慧安防中的应用

随着我国城市化进程的不断加快,城市资源的有限性与外来人口的恶性膨胀,人、财、物的大流动与社会控制力的相对弱化等一系列矛盾日益突出,引发了一系列的社会治安问题。由于警力不足、配置不当等原因,基层民警往往一警多能、一警多职,不但民警工作压力大,而且许多案件可能得不到及时的侦破与处理,治安管理往往处于被动局面^[5]。同时,社会治安防控体系涉及一切社会治安资源,不仅包括公安,也包括其他机关企事业单位以及各种群防群治组织等。各部门协调配合的好坏对于案件处理的效率存在重要影响。

如今,以计算机视觉识别为基础的生物体特征识别方式已逐渐开始成熟,并开始受到来自社会各界的广泛关注。这些特征包含面部特征,以及更细微的如人眼的虹膜、掌纹、声纹乃至步态等。尽管机器与生物体之间的交流还无法突破机械手段,各种信息和数据的录入手段还有待突破,但是机器能够通过视觉识别、生物特征检测等技术检测图像中的人是否是活体存在的,能够借助后台数据的对比鉴别图像中的人员身份信息,甚至识别图像中人员行动的轨迹和姿势,这无疑为城市安防治理提供了新的思路和方法。

目前,深度学习在计算机视觉中应用广泛且比较成功的案例为人脸识别,以下为上海市某小区应用人脸识别技术作为安防监控手段的真实案例。

3.1 流程设计

基于社区对可疑人员身份的检查需求,所设计的业务流程主要分为以下几个部分。

当人员进入摄像机监控范围,计算机视觉识别系统首先通过视频流分析,甄别当前画面中是否存在人像,并对人脸部分进行捕捉和抓拍,抓拍示意图如图2所示。

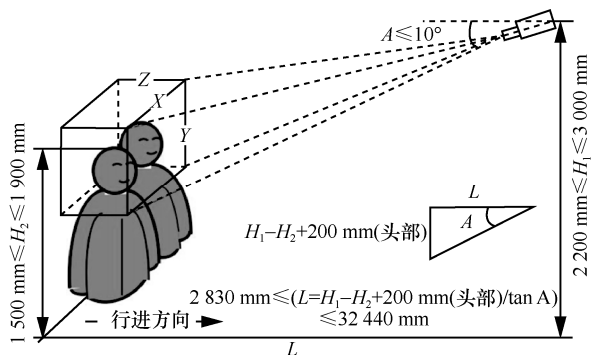


图2 抓拍示意图

系统会记录此次抓拍的时间和摄像机信息，并对抓拍到的人像做特征提取处理。同时，系统会将该特征在提前录入的“布控库”中做比对检索，如果对比结果高于设定阈值，识别系统将发出可疑人员身份提醒，并将提醒推送至网页端或移动客户端，便于保安及片区民警及时安排下一步动作。

3.2 算法原理

此次案例中设计的动态人像识别的算法识别过程主要包括视频解码、人脸检测、图像聚合、特征提取、检索比对5个模块。当摄像机将视频源传输给动态人像卡口系统时，系统实时分析视频中的人脸并进行布控报警，其算法识别原理如图3所示。

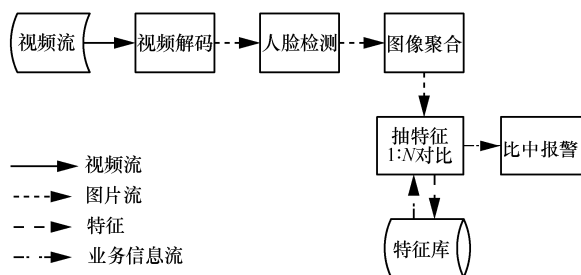


图3 算法识别原理

步骤1 进入视频解码模块。该模块首先会将不同码流的视频进行格式转化，然后以10~25 f/s的帧率进行解码处理，最终输出一帧帧的静态图片。

步骤2 进入人脸检测模块。一张图片中可能存在多张人脸，该模块会检测图片中的人脸信

息，从每一帧静态图片中一一识别人脸，并形成人脸特写图。

步骤3 进入图像聚合模块。由于人员一般不会主动配面对前端摄像机，所以人像系统如何精确识别行进人群中各种角度的人脸并进行精确报警是一件非常困难的事情。为此，在此次方案中添加图像聚合模块，将视频中同一人的多帧图片进行聚合，形成一条短暂的人脸轨迹（从进入监控画面到离开监控画面），把不同角度、姿态的同一人照片融合为一组图片。其优势在于，当小区内人流量较大时，容易出现人脸前后遮挡，导致同一人的轨迹被隔断误认为是不同的人的轨迹，重复计数。图像聚合模块能很好地处理高峰人脸遮挡情况，即使轨迹断掉，也可实时地将同一人的轨迹合并起来，从而有效避免大量的重复抓拍。

步骤4 进入特征抽取模块。算法将这一组图片进行特征抽取，形成一组特征码。该组特征码会存于数据库形成路人库，同时用于与布控库进行实时对比。

步骤5 进入检索比对模块。将一组特征码与布控库的特征值进行索引式对比，若比中则输出比对命中的信息，并在网站或手持终端进行提醒信息的实时推送。

3.3 算法训练

系统可根据布控库进行实时布控，当前端摄像机的抓拍人像和库中人像的相似度超过所设阈值时，系统会在第一时间推送消息提醒。为了保证功能稳定，还需要对系统进行性能评定，此次案例中选用业界常见的召回率与误报率作为关键指标。

在评测建模前首先定义3类数据集。

- 训练集 (training set): 用来训练学习算法。
- 开发集 (development set): 用于调整参数、特征选择等。有时也称为保持交叉验证集 (holdout cross validation set)。
- 测试集 (test set): 用于评估学习算法的性能，



该部分数据集是不能用于调整算法参数或者做特征选择的（该部分由开发集负责）。

在对上述开发与测试集定义后，系统可以通过模拟真实的人员识别场景，利用各异的算法学习参数展开练习，并快速检验各类算法的实验结论。在实验与练习的过程中，选择精确率（precision）针对预期的实验结论进行描述，它展示实验预期准确的样本中真实的正样本数量。预测为真则存在两种可能，一种是把真实结果预测为真实结果（true positive, TP），另一种是把错误结果预测为真实结果（false positive, FP）。而针对初始样本，则选用召回率（recall）作为测算依据。它表示的是初始样本中的真实样本被预测为真实的数量。与精确率的计算过程类似，同样会产生两种结果，一种是将初始的真实结果预测成真实结果（true positive, TP），另一种则是将初始的真实结果预测为错误结果（false negative, FN）。

精确率计算式：

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

召回率计算式：

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$F1$ 是 precision、recall 的调和平均值。

F 值计算式：

$$F1 = \frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} = \frac{\text{precision} + \text{recall}}{2 \times \text{precision} \times \text{recall}} \quad (3)$$

从式(3)中可以看出 $F1$ 得分综合了 precision 与 recall 的结果，因此当 $F1$ 较高时，则说明该算法比较有效^[6]。

通过与上述模型相结合，用于安防需求的计算机视觉识别系统理论上可以根据不同区域的安防级别不同，对自身算法进行自主调整适应当前的需求。例如，对于有些安防级别较低的区域，可以采用 F 值评分并不高的算法实现，降低对硬件、网络、计算资源的依赖和占用程度；反之对

于安防级别较高的区域，则可以部署 F 值较高的算法实现。

4 实施效果

针对城市安防的使用场景，本次方案中采用基于前端高清摄像机的计算机视觉识别系统，自动在实时视频中检测每一位来访人员，并截取最清晰的一帧图像存储。系统自动对视频中的每一帧图像逐帧做人像检测，当在不同帧识别出一个人时，系统将同一人在不同帧的位置形成一个轨迹，挑选这段轨迹中最清晰的一张作为抓拍照片。

在实际的使用场景中，人流量是影响人脸识别性能的一个关键外部因素。此次方案中设计的计算机视觉识别系统不仅可以处理画面中人脸逐一通过的简单场景，也可以在人流量高峰时期确保人像的实时捕获、不漏抓。在某小区的实际方案中，所设计的系统支持每路摄像机平均 1 万人/天、峰值 30 人/min 的过海量抓拍，保证人像信息及抓拍信息均被完整存储。所有存储的人像信息、抓拍信息均可以被实时查询。

在符合识别系统架设规范的环境下，人像抓拍情况的指标见表 1，分别为抓拍捕获率及误抓拍率。

以上基于人脸识别的计算机视觉识别技术的方案在社区及多类重点区域的安防布控落地时，能够自动完成识别、抓拍、对比，无须主动配合。从来访人员身份被识别到信息告知，8 s 内即可完成提醒，保安或片区民警可迅速响应，对于实战具有较强的指导意义。

表 1 人像识别系统参数指标

指标参数	指标参数说明	性能
人像抓拍捕获率	抓拍的有效人像数占实际通过人数（符合质量要求）的百分比	95%
人像误抓拍率	抓拍的无效人像占抓拍总人像数的百分比，无效人像包括非人脸	<0.1%

5 结束语

计算机视觉技术被认为是城市智慧安防的“眼睛”，能够采集、捕捉、分析各类图像及视频信息，无论安防程度还是安防效率，都传统人工无法企及的。诚然，这双“眼睛”也存在一定弊端，它不可避免地会产生大量的人脸影像数据，日积月累，除了会耗费政府本就紧张的算力、网络资源外，还会产生影像数据的隐私保护和监管问题。目前国内对这方面的政策仍然不够完善，但不可否认的是，现在国家对于深度学习、计算机视觉等人工智能领域方面的大力支持，打造了一个比以往任何时代都有利于技术蓬勃发展的环境。

如今，各种新产品和解决方案不断提出，行业间壁垒逐渐开始淡化。除了在安防领域的应用，计算机视觉还能够教育、医疗、金融领域以及诸如身份识别、刷脸支付等日常生活场景中提供技术支持。伴随5G带来的低时延、高速率以及大带宽的网络变革，未来类似医疗影像、自动驾驶等门槛极高的专业化场景也会逐渐被计算机视觉技术所攻克，相信计算机视觉行业生态在技术力量的驱动下，将会为城市治理、智慧安防以及更多的领域注入全新活力。

参考文献：

- [1] MARR D. Vision: a computational investigation into the human representation and processing of visual information[M]. London: MIT Press, 1981.
- [2] 陆志伟, 王明革, 赵瑞海. 计算机视觉技术的研究与应用[J]. 数字技术与应用, 2016(3): 109.
LU Z W, WANG M G, ZHAO R H. Research and application of computer vision technology[J]. Digital Technology and Application, 2016(3): 109.
- [3] TREICHLER D G. Are you missing the boat in training aids[J]. Film and AV Communication, 1967(1): 14-16.
- [4] 孙雨萌. 深度学习在计算机视觉分析中的应用分析[J]. 中国新通信, 2018, 20(23): 169-171.
SUN Y M. Analysis of deep learning and its application in computer vision[J]. China New Telecommunications, 2018, 20(23): 169-171.
- [5] 孙青龙. 浅析城市社会治安防控体系建设存在的问题及对策研究[C]//新型城镇化进程中的法律问题研究——第十届中部崛起法治论坛论文集. 合肥: 出版社不详, 2017: 1367-1378.
SUN Q L. Analysis of urban social security prevention and control system construction problems and countermeasures[C]//Research on legal Issues in the Process of new urbanization -- Proceedings of the 10th Central Rise of the Rule of Law Forum. Hefei: S.n., 2017: 1367-1378.
- [6] 金剑锋, 姚元, 王明晓, 等. 新零售时代, 基于5G和AIoT技术的智慧门店探索与应用[J]. 电信科学, 2020, 36(S1): 284-292.
JIN J F, YAO Y, WANG M X, et al. Exploration and application of smart stores based on 5G and AIoT technology in the new retail era[J]. Telecommunications Science, 2020, 36(S1): 284-292.

[作者简介]



陈志宏(1971—), 男, 中国电信股份有限公司上海分公司副总经理, 主要研究方向为通信技术、云网融合和人工智能。

王明晓(1991—), 男, 中国电信股份有限公司上海分公司业务创新中心产品经理, 主要研究方向为通信技术、物联网技术和人工智能。