基于卷积神经网络的鞋型识别方法

杨孟京，唐云祁*，姜晓佳
中国人民公安大学刑事科学技术学院，北京 100038

摘要 “监控＋鞋印”是目前公安机关刑事侦查的重要技战法，其基本原理是依据犯罪现场鞋印推断嫌疑人所穿鞋型，然后到周边监控视频中检索嫌疑鞋型。针对“监控＋鞋印”技战法自动化程度低下的问题，提出一种基于卷积神经网络的鞋型识别方法，实现对嫌疑鞋型的自动识别。根据鞋型识别独有特点，在 DeepID 的基础上设计卷积神经网络框架，并构建鞋型样本数据库(50 双鞋型样本，共计 160231 幅图像)。运用 Caffe 框架结合不同网络模型对鞋型图像数据进行训练和测试，实验设计的初始网络结构由两层卷积、两层池化、两层全连接组成。实验比对了不同的第一层全连接层输出元素数目对网络性能与训练效率的影响，又在不改变输出特征图大小的情况下比对了不同网络深度的实验结果，在优化模型的基础上引用重叠池化得到实验最优网络模型。实验结果表明，卷积神经网络对于鞋型有很好的识别效果，识别精度值最高达 96.06%，为鞋型识别提供了一种新的途径。

关键词 机器视觉；鞋型识别；卷积神经网络；重叠池化；刑事侦查
中图分类号 TP391 **文献标识码** A **doi**：10.3788/LOP56.191505

Novel Shoe Type Recognition Method Based on Convolutional Neural Network

Yang Mengjing, Tang Yunqi*, Jiang Xiaojia
Institute of Forensic Science, People's Public Security University of China, Beijing 100038, *China*

Abstract Criminal investigation is often conducted based on the surveillance video footage and crime-scene shoeprint identification. The basic principle of this method is to infer the type of shoe worn by the suspect based on the shoeprints identified at the crime scene and to subsequently search for the shoe in the surveillance video footage. To solve the problem of low automation associated with this criminal investigation method, a new shoe type recognition method using a convolutional neural network has been proposed in this study. According to the unique characteristics of shoe type recognition, a framework of convolutional neural network is designed on the basis of DeepID, and a shoe database containing 50 pairs of shoes and 160231 images is constructed. The experiments are conducted based on the Caffe framework using different network models. Initially, the network structure comprises two convolution layers, two pooling layers, and two full connection layers. Further, experiments are conducted to compare the effects of the number of output elements in the first layer of two full connection layers on the network performance and training efficiency, and the experimental results of different network depths are compared without changing the size of the output feature graph as well. Based on the optimization model, the optimal network model is obtained by using overlapping pooling. The experimental results denote that the proposed method achieves an excellent performance, with an accuracy of 96.06%. Therefore, the proposed method can be considered to be a promising new method for shoe type recognition.

Key words machine vision; shoe type recognition; convolutional neural network; overlapping pooling; criminal investigation
OCIS codes 150.1135; 100.4996; 070.5010; 330.5000

收稿日期：2019-03-01；**修回日期**：2019-03-25；**录用日期**：2019-03-27
基金项目：国家重点研发计划(2017YFC0822000)、国家自然科学基金(61503387，61772539)、上海市现场物证重点实验室开放课题、中央高校基本科研业务费项目(2018JKF217)

* **E-mail**：yunqit@163.com

1 引 言

随着视频监控技术的广泛应用，融合现场静态足迹及周围动态监控视频资料对犯罪嫌疑人进行追踪溯源的方法，为足迹检验的发展创造了新的契机。现有全国公安机关鞋样本查询系统可根据现场足迹得出鞋型图像，侦查人员在侦办案件工作中根据分析得出的鞋型图像在监控视频中查找犯罪嫌疑人。视频监控领域的不断建设与完善也使这一视频侦查技术得以发展成熟，使其在案件侦破中的作用日益显著。2015 年 1 月，广东云浮市公安局就根据“耐克”、“鸿星尔克”等品牌鞋型确定了作案人数，结合案发时空节点迅速缩小侦查范围，关联外围视频监控资料，锁定了技术开锁入室盗窃案的流窜盗窃犯罪团伙并将其成功抓捕[1]。2016 年 12 月，内蒙古呼和浩特市公安局在一起超市杀人案中利用与现场数枚血鞋印花纹一致的嫌疑鞋型进行视频侦查模拟实验，通过特征标示、拼接对比、重合比较等方法将可疑鞋图像作清晰化处理，利用特征评价对比分析处理后的图像与模拟鞋图像，分析结果表明两者种类特征相同，无本质性差异，由此可确定嫌疑鞋型为现场血鞋印同种鞋型。根据这种鞋型在视频监控中排查确认重点嫌疑人，进行轨迹追踪，为调整侦查重心提供了重要依据[2]。通过视频监控中获取的鞋型轮廓、品牌、花纹等鞋的排它属性特征可为侦破案件确定筛查范围，鞋的信息容量越丰富，视频侦查工作将愈发快捷高效。但目前查找锁定视频中的鞋型以及犯罪嫌疑人这一过程仍主要依靠公安技术人员人工进行，侦查效率低下，易错过侦查抓捕工作的最佳时机。因此，亟需一种以依据犯罪现场足迹得到的鞋型图像为基础，对犯罪现场周围不同场景监控视频中的鞋型图像进行自动匹配识别的方法，以提高鞋型分析自动化程度，为追踪锁定犯罪嫌疑人提供快速、有效的信息。

本文提出一种基于卷积神经网络对监控视频中的鞋型图像进行快速自动识别的方法。在案发时间段内，将犯罪现场周围视频中出现的行人进行检测并分割，进而对鞋型进行切割，建立临时数据库，然后将犯罪现场推断出的鞋型样本输入数据库进行分类识别，得到与样本相似的鞋型，进而查找到穿有目标鞋型的犯罪嫌疑人图像信息。由于特定时间段出现相似鞋型的概率不高，即使同时得到多双相似鞋型，也可协助公安技术人员根据鞋印推断犯罪嫌疑人的身高、体态等。在视频中自动检索人身并分割的算法已相对成熟[3-4]，故对鞋型进行切割也是可行的。本文将研究重点放在建立临时数据库后的分类识别算法。首先设立犯罪现场周围监控视频，在监控视频中采集实验数据，并对数据进行格式转换、分帧、鞋型区域切割及归一化处理，构建实验数据库，其中包含 50 类鞋型共计 160231 张数据图像；基于 Caffe 框架设计网络结构，搭建适配的网络模型，分析第一层全连接层输出元素数、网络深度对网络性能的影响，引用重叠池化对网络模型进行优化，通过调试网络参数等方法提高网络识别率与稳健性，得到准确、高效的鞋型识别网络，完善适用于鞋型识别的网络模型与方法。

2 相关研究工作

继监控普及后，融合足迹与犯罪现场周围监控视频对犯罪嫌疑人进行追踪溯源才逐渐开始被研究。袁楚平等[1]首次将犯罪现场提取的足迹输入全国公安机关鞋样本查询系统，通过比对查出鞋型图像，并在周围视频中搜寻嫌疑鞋型进而找到犯罪嫌疑人，结合后续的侦查工作打破侦查僵局，为破案提供重要线索。但文中对犯罪现场周围视频中鞋型采用人工识别的传统技术方法，非常耗费警力资源，且侦查效率低下，易错过侦查抓捕工作最佳时机。许磊等[2]对视频监控中可疑鞋图像与模拟实验视频中模拟鞋图像进行比对分析，发现可疑鞋图像与模拟鞋图像的特征基本相符，缩小了侦查范围。文中应用人工寻找特征、人工比对的方法，工作量较大，耗时费力，且易受主观因素影响，误差较大。鞋型外观形态的识别关键在于特征的提取与比对，只要对鞋型图像实现自动特征提取与智能决策即可实现该问题的自动化识别。自动化识别需要利用图像识别技术，即通过对图像的处理分析来识别不同模式目标或对象（如人脸、虹膜、花卉植物、遥感图像等）。近年来，图像识别技术不断发展，已经广泛应用于各个领域，其中人脸识别[5-7]、步态识别[8-9]等技术已相对成熟。

图像识别技术起源于 20 世纪 50 年代，最早应用于办公自动化任务中的字符识别，主要针对二维图像进行分析识别。20 世纪 60 年代，Roberts[10]从数字图像中提取三维特征，开创了理解三维场景的图像识别研究。而后的几十年中，前期图像识别技术所采用的高斯混合模型、k 均值聚类、支持向量机等浅层结构，可以解决医学、工业、交通、安防等各个领域的一些简单问题，但是难以解决复杂的真实环

境问题。后来随着深度神经网络的出现，特别是自2011年以来随着计算机计算能力的显著增强，各种算法的应用效果良好，出现了一大批基于神经网络的图像识别方法。自 Krizhevsky 等[11]在 ILSVRC 竞赛上提出卷积神经网络结构（AlexNet）后，卷积神经网络在图像分类和识别领域得到了广泛应用。卷积神经网络具有相对简单的网络拓扑结构，可直接处理图像二维矩阵，同时进行特征提取和模式分类，其利用权值共享，缩短了运行时间，适合不同领域的多分类任务[12]。2015年，牛津大学计算机视觉组和 DeepMind 公司共同研发了一种深度卷积网络 VGGNet[13]。VGGNet 相对于 AlexNet 网络深度更深，有16层和19层两种结构。2014时，VGGNet 网络在加深的同时减少了参数量，多个非线性操作使得网络判别性更好。同年，香港中文大学的 DeepID 项目[14]及 Facebook 的 DeepFace 项目[15]利用卷积神经网络结构，在 LFW 数据库上的人脸识别正确率分别达 97.45% 和 97.35%，只略低于人类视觉识别正确率（97.5%）[16]。He 等[17]提出的 ResNet 取得了 2015 年度的 ILSVRC 比赛冠军，ResNet 的核心是学习映射的残差，故为残差网络。相比于19层的 VGGNet，ResNet 的网络深度加深了7倍，解决了随着网络加深到一定程度后的识别精度下降问题。

鉴于卷积神经网络的良好应用结果，本文将卷积神经网络引入鞋型分析之中，不仅可以避免人工选择确定鞋型特征的主观因素影响，同时自动化识别匹配速度的大幅提升，可有效提高侦查工作效率，避免人工查找的误差，匹配结果可快速应用到下一阶段的侦查工作中。

3 卷积神经网络

卷积神经网络通常由卷积层、池化层、全连接层、输出层等构成，卷积层常与池化层相互交替使用。

每个卷积层会有多个卷积核，于是，数据经过卷积层会形成多个特征图，每个特征图由众多神经元组成，每个神经元与特征图的输入局部连接，通过对输入进行加权求和再加上偏置值得到输出。假设输出特征图为 a_i，则 a_i 的输出过程可以表示为[18]

$$a_i=\sigma(a_{i-1}*W_i+b_i), \tag{1}$$

式中：i 代表层数；a_{i-1} 为第 $i-1$ 层的输出（上一层的输出作为下一层的输入）；“$*$”代表将上一层的输出与卷积核第 i 层的权重 W_i 进行卷积操作；b_i 代表第 i 层的偏置；σ 为非线性激活函数，本文采用 ReLU 函数。

池化层常用的有最大池化和平均池化，主要作用是对数据进行降维，在降低计算量的同时增加了模型的泛化能力，加强了算法的稳健性。池化层一般设置在卷积层后，卷积层输出的特征图作为池化层的输入。经过池化操作后，输出的特征图数量不发生改变[19]，特征图大小变化公式为

$$O_{\text{utput}}=\frac{i_{\text{nput}}-k_{\text{ernel\_size}}}{s_{\text{tride}}}+1, \tag{2}$$

式中：O_{utput} 为池化后输出特征图大小；i_{nput} 为池化层输入特征图大小；$k_{\text{ernel\_size}}$ 为池化滑动窗口的大小；s_{tride} 为滑动步长。

经过多层卷积和池化之后，连接一个或多个全连接层。全连接层的作用是将卷积与池化操作后的输出整合为一维数组，对提取的特征进行分类，得到基于数据输入的概率分布 P，即

$$P(j)=P[L=l_j\mid a;(w,b)], \tag{3}$$

式中：L 表示损失函数；l_j 表示第 j 个标签类别；w 表示权重大小；b 表示偏置项值。卷积神经网络训练的最优模型是使损失函数最小。卷积神经网络的本质是对原始数据 a 进行多层滤波，以降低计算量，原始数据 a 经前向传播后得到的结果与预测结果的差异称为残差。

全连接层的每一个结点都与上一层的所有结点相连，用来把前边提取到的特征综合起来。由于其全相连的特性，一般全连接层的参数也是最多的，增大了计算量，所以现在常用卷积来代替全连接。

4 基于卷积神经网络的鞋型识别方法

4.1 初始网络模型

香港中文大学的 DeepID 项目[12]中 LFW 数据库上的人脸识别正确率达 97.45%，鞋型识别问题与人脸识别较为类似，根据鞋型识别独有的特点，参照 DeepID 项目的网络框架设置适用于鞋型识别的初始网络结构。DeepID 网络框架包含4层卷积，3层池化，2层全连接。其 Ip1 层连接了第4层卷积的同时还连接了第3层池化，故可兼顾局部与全局特征，但需要较多输出神经元。DeepID 网络可解决的数据集种类很多，而本实验模拟的数据库种类较少，考虑到计算量的问题，本文并未完全引用全连接策略，仅设置2层卷积、2层池化，其网络结构如图1所示。

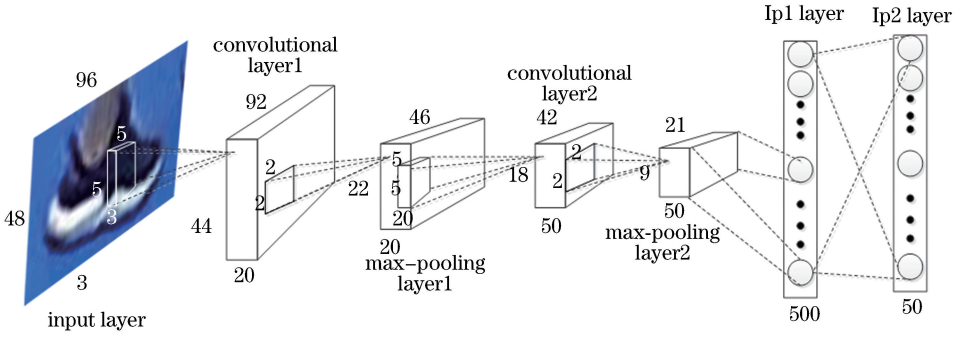

图 1 初始网络模型

Fig. 1 Initial network model

4.2 网络模型改进

4.2.1 Ip1 层输出改进

Ip1 层为全连接层，全连接层的目的是将网络学习到的特征映射到样本的标记空间中，但同时会损失部分特征。Ip2 层的输出作为分类层的输入，本文是对 50 类鞋型进行训练。Ip2 层的输出元素数不易改动，因此需探究 Ip1 层输出的元素个数对实验结果的影响。

4.2.2 网络深度加深

随着经典网络的发展，网络深度不断加深，其在 ImageNet 数据集的应用效果也越来越好。对于 Lenet 的 4 层网络结构和 AlexNet 的 8 层网络结构，Top-5 错误率为 16.4%；对于 VGG 的 19 层网络结构，Top-5 错误率降至 7.3%；对于 GoogLeNet[20] 的 21 层网络结构，Top-5 错误率低至 6.7%；后来 ResNet 的网络结构达到了 152 层，其 Top-5 错误率仅为 3.57%。虽然以上应用效果越来越好的影响因素还包括引入了 dropout、小卷积核、残差网络等，但是随着网络的不断加深，网络将具有更多的非线性函数，即使在不引入其他影响因素的情况下其判决函数判决性也会越来越好，可以达到更好的分类效果。

4.2.3 加入重叠池化

传统的池化策略是使池化滑动窗口大小与滑动步长相等，本文初始设置的网络结构的滑动窗口大小和步长均为 2。重叠池化与传统池化不同的是池化窗口大小大于滑动步长，相邻池化窗口之间会有重合区域。本文加入重叠池化，设置池化层的滑动窗口大小为 3 和 4。当训练或测试输入的图片较大时，由于传统池化策略是对输入图片进行单一映射，因此输出的图片特征维度仍旧较高。重叠池化将池化层的输出扩充为多级较小的特征，降低了池化层输出的特征维度[21]。假设原来池化层的输入大小为 $n\times n$，传统池化的池化窗口大小为 2×2，步长为 2，则输出的特征图大小 s 为 $(n-2)/2+1$，重叠池化的池化窗口大小为 3×3，步长为 2，则输出的特征图大小 s_1 为 $(n-3)/2+1$。显然，$s_1<s$，即池化窗口增大，输出的特征维度降低。

5 实验及结果分析

5.1 实验数据库的构建

本实验视频数据采集在足迹实验室进行，对 8 名志愿者(男性 6 名，女性 2 名)穿的 50 双鞋进行视频数据采集。在实验场地铺置浅蓝色地毯以减少背景干扰。为模拟视频监控中不同角度的鞋型并保证实验的可靠性，从不同角度(0°，5°，10°，15°，180°等)进行数据采集，同时通过调整室内灯光使视频数据不受光照强度的影响。志愿者依次在 1.2 m 宽的行走区域行走，行走路线如图 2 所示。标线为行走路线，标线首端为起点，沿着箭头来回行走。采集设备为海康威视监控摄像机，设置帧率为50 frame/s，每帧大小为 1920 pixel×1080 pixel，每类样本采集时长约为 80 s。

视频数据采集完成之后，运用 HMSTranscoder 软件对视频进行格式转换并截取每双鞋的视频素材作为一个样本视频，同时将视频亮度设定为 12，对比度为 15。运用 Matlab(R2016a)软件对每一个视频样本进行批量分帧处理，为确保数据量，对视频进行逐帧分帧，建立 50 个视频帧文件夹，命名为 1～50。由于视频帧尺寸较大，鞋型占有比例很小，为了排除无效的干扰并减小运算量，对视频帧中的鞋按 2∶1 的比例进行切割，并归一化为大小为 96 pixel×48 pixel 的图片保存。视频帧截取鞋型过程是模拟自动检测和切割，运用 Matlab(R2016a)编写程序，可以通过手动点击鞋型左上角位置来获得坐标，自动截取比例为 2∶1 的鞋型图片，并保存到指定文件夹。截取过程如图 3 所示。

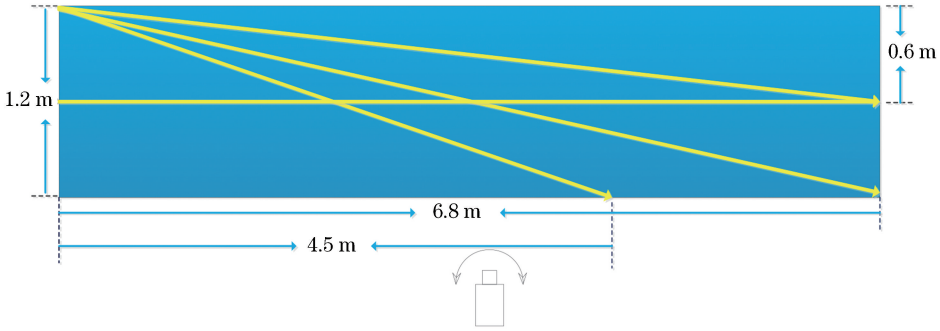

图 2 数据采集路线示意图

Fig. 2 Schematic of data acquisition route

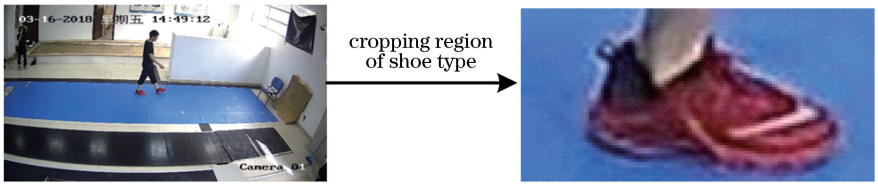

图 3 切割流程示意图

Fig. 3 Schematic of cutting process

视频帧处理过程中将左脚与右脚分开截取，对于左右脚重叠，有其他干扰及角度大于 45°的视频帧予以删除，对截取得到的 50 类鞋型图像数据进行筛选，挑选不同光照、角度样本图像共 160231 张。各类样本数据示例如图 4 所示。

图 4 各类实验数据示例

Fig. 4 Examples of various types of experimental data

5.2 实验配置及参数设置

本实验的硬件配置为 Intel(R)Core(TM)i7-8700K CPU @ 3.7 GHz，内存 16 GB，GPU 为 NVIDIA GeForce GTX1080。软件配置为 Windows 10，CUDA9.1 GPU 并行计算库，开源深度学习框架 Caffe。

本实验基于深度学习 Caffe 框架进行搭建，超参数设置默认选用小批量梯度下降优化算法，基础学习率为 0.0001，学习率调整策略设置为 inv。根据 inv 的学习策略：$l_r = b_{ase\_lr} \times (1 + g_{amma} \times i_{ter})^{-p}$，其中 l_r 为新的学习率大小，b_{ase_lr} 表示初始学习率大小，g_{amma} 表示学习率因子大小，p 为自定义的值，为防止学习率下降过快，设置学习率衰减系数为 0.0001，梯度更新权重为 0.9，权重衰减为 0.0005。

初始网络输入层的数据大小为 96 pixel×48 pixel，任意选取数据库 80%的样本作为训练集(共 128185 张)，20%的样本作为测试集(共 32046 张)。训练批量数目 b_{atch_size} 为 128，第一层卷积层选取 20 个大小为 5×5 的卷积核，步长 s_{tride} 为 1，输出 20 个大小为 92×44 的特征图。然后将卷积层得到的特征图引入 ReLu 非线性激活函数并进行标准化处理，保留特征的均值与方差，输出 20 个大小为 92×44 的特征图。利用滑动窗口大小为 2×2、步长为 2 的池化层，采用最大池化输出 20 个大小为 46×22 的特征图，将其作为第 2 卷积层的输入。第 2 层卷积选取 50 个大小为 5×5 的卷积核，步长为 1，边缘补充 p_{ad} 为 0，输出 50 个大小为 42×18 的特征图。窗口大小为 2×2、步长为 2 的池化层输出 50 个大小为 21×9 的特征图，将其作为第 1 层全连接 Ip1 层的输入，输出元素数设置为 500。最后一层是输出层，考虑到数据集种类设置输出神经元个数为 50。

5.3 实验结果与分析

5.3.1 Ip1 层输出元素个数对性能的影响

为了探究 Ip1 层输出元素个数对性能的影响，本文在初始模型上设置 Ip1 层输出元素个数分别为 300，500，1000，并对识别精度和训练时间进行比较

分析，实验结果如表 1 所示。

表 1　Ip1 层输出元素个数对性能的影响

Table 1　Effect of number of output elements in Ip1 layer on performance

Number of output elements	300	500	1000
Test accuracy /%	89.68	91.91	93.46
Train time /min	48	50	56

从表 1 中可以看出，随着 Ip1 层输出元素数的增加，鞋型的识别精度有所提高，错误率不断下降。这是由于随着输出的元素数增加，对鞋型提取的特征也相应增多，最后一层提取特征的选择性更加全面。从表 1 中也可以看出，随着输出元素数的增加，训练时间也会增加。因此，权衡网络性能及训练效率，在本实验前期训练中设置 Ip1 层输出元素为 500。

5.3.2　网络加深对性能的影响

由于利用两层 3×3 大小的卷积核堆叠的卷积层得到的感受野与大小为 5×5 的卷积核相同，但在达到相同感受野的情况下，采用较小卷积核可增加网络深度，提取特征更加精细。故本实验将原始网络5×5 大小的卷积核转换为两层 3×3 大小的卷积核，其他参数不变，实验效果如表 2 所示。

表 2　网络加深对性能的影响

Table 2　Effect of network depth on performance

k_{ernel_size}	Number of layers	Memory /MB	Train time /min	Accuracy /%	Loss
5×5	6	83.6	50	91.91	0.2807
3×3	8	150.4	78	95.81	0.1601

从表 2 中可以看到，在输出特征图数不变的情况下，通过运用多个尺寸较小的卷积核代替尺寸较大的独立卷积核，网络深度增加了两层，识别精度提高了 3.9%，损失值降低了 0.1206。这可能是由于增加网络层数后，网络具有更多的非线性函数，判决函数判决性更好。从表中可以看出，网络加深减缓了训练速度。这是由于网络的加深训练提高了网络的复杂程度，所需要的内存值也相应增加，因此增加了训练时间。

5.3.3　重叠池化对性能的影响

在加深网络结构后，识别精度值显著提升，在此基础上运用重叠池化代替传统的池化策略，在 pool1 层设置池化滑动窗口大小为 4×4，步长为 2，设置 pool2 层池化滑动窗口为 3×3，滑动步长为 2。实验结果如表 3 所示。

表 3　重叠池化对性能的影响

Table 3　Effect of overlapping pooling on performance

Pooling	Memory /MB	Train time /min	Accuracy /%
Original pooling	150.4	78	95.81
Overlapping pooling	145.9	72	96.06

从表 3 的实验结果可以看出，引用重叠池化后，模型训练所需内存减小，训练时间明显缩短，这是由于经过两层重叠池化输出的特征图大小为 20×8，比传统的池化输出 21×9 大小的特征图小，降低了输出特征维度。同时，池化窗口增大且步长不变，相比传统池化特征提取目的性更强，使得网络识别精度也提升了 0.25%。

5.3.4　综合实验结果与分析

综合以上分析，选用加深网络结构并运用重叠池化对搭建的鞋型数据库进行训练和分类识别时，其测试识别精度可达 96.06%。该网络下随着训练迭代次数的增加，识别精度和损失值的变化情况如图 5 所示。从图中可以看出，训练迭代次数达到 10 万的时候测试精度超过 80%，训练损失值降低到 0.1左右。训练迭代 10 万次后，识别精度在慢慢提高，损失值趋于收敛。

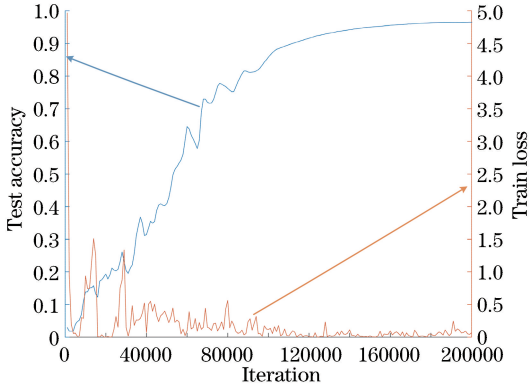

图 5　正确率及损失率变化图

Fig. 5　Variation curves of test accuracy and train loss

选用加深网络结构并运用重叠池化对搭建的鞋型数据库进行训练和分类识别后，识别错误率为 3.94%。识别错误的原因主要是颜色较为相近、图像模糊和光照不一等，部分错误识别的鞋型图片如图 6 所示。图 6(a)是将预测标签值为 4 的鞋型识别成标签值为 14 的鞋型，这主要是因为两类鞋型颜色较为相似，采集数据过程中行走过快导致鞋型图

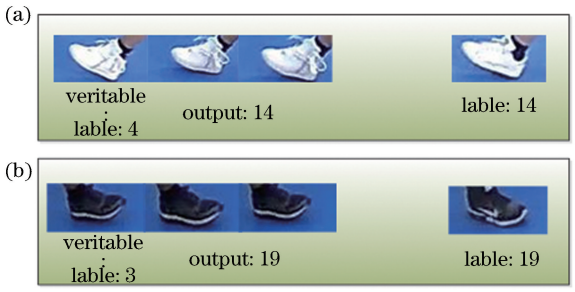

图 6 部分错误识别的鞋型图片。(a)标签为 4 的鞋型图片被错误识别为 14 示例;(b)标签为 3 的鞋型图片被错误识别为 19 示例

Fig. 6 Photographs of partial misidentify of shoe type. (a) Example of shoe image with label 4 being incorrectly identified as label 14; (b) example of shoe image with label 3 being incorrectly identified as label 19

像模糊。图 6(b)是将标签值为 3 的鞋型识别为标签值为 19 的鞋型,这主要是因为不同光照导致标签值为 19 的鞋型颜色与标签值为 3 的鞋型颜色相近。

6 结 论

基于大规模采集数据,建立数据库,设计适用于鞋型识别的卷积神经网络结构,设置合适的超参数。通过比较不同全连接的输出元素数对网络识别精度结果的影响,发现输出元素数的增加会提高识别精度,但训练时间会随之不断增加。权衡训练效率与精度后,设置 Ip1 层输出元素数目为 500。利用两个 3×3 卷积核替换 5×5 卷积核后,加深网络可显著提高模型精度。引用重叠池化代替传统的池化策略,识别精度可提高至 96.06%,实验结果充分验证了基于卷积神经网络进行鞋型识别方法的可行性。

将卷积神经网络应用于鞋型识别,取得了良好的识别效果。通过对初始设计网络的改进,在提高精度的同时也减少了参数量,加快了训练速度,提高了网络性能与识别效率。然而所采用的实验数据皆为纯蓝色背景下的鞋型,实际侦查案件中经常存在复杂背景、形变、部分遮挡等干扰。现阶段实验网络结构较为简单,且针对颜色相近的样本识别效果不佳。下一步将对网络进行改进,以提高对相似样本的识别精度,搭建适用于复杂场景下稳健的鞋型识别网络模型。

参 考 文 献

[1] Yuan C P, Yu S W. Preliminary study on the application of footprint analysis in video investigation [J]. Guangdong Gongan Keji, 2017, 25(2): 61-63, 74.
袁楚平, 余尚伟. 足迹分析在视频侦查工作中的运用初探[J]. 广东公安科技, 2017, 25(2): 61-63, 74.

[2] Xu L, Li Z H, Li Z G, *et al*. A murder case investigated and solved by applying the simulation experiment into the collected video [J]. Forensic Science and Technology, 2018, 43(4): 330-333.
许磊, 黎智辉, 李志刚, 等. 视频侦查模拟实验在案件侦破中的应用[J]. 刑事技术, 2018, 43(4): 330-333.

[3] Wang X, Liu Y, Li G Y. Moving object detection algorithm based on improved visual background extractor algorithm [J]. Laser & Optoelectronics Progress, 2019, 56(1): 011007.
王旭, 刘毅, 李国燕. 基于改进视觉背景提取算法的运动目标检测方法[J]. 激光与光电子学进展, 2019, 56(1): 011007.

[4] Chen C, Xuan S B, Xu J G. Pedestrian detection and segmentation under background clutter[J]. Computer Engineering and Applications, 2012, 48(30): 177-181.
陈超, 宣士斌, 徐俊格. 复杂背景下的行人检测与分割[J]. 计算机工程与应用, 2012, 48(30): 177-181.

[5] Wu C Y, Ding J J. Occluded face recognition using low-rank regression with generalized gradient direction[J]. Pattern Recognition, 2018, 80: 256-268.

[6] Weng R L, Lu J W, Hu J L, *et al*. Robust feature set matching for partial face recognition[C] // 2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, Australia. New York: IEEE, 2013: 601-608.

[7] Ali Akber Dewan M, Granger E, Marcialis G L, *et al*. Adaptive appearance model tracking for still-to-video face recognition [J]. Pattern Recognition, 2016, 49: 129-151.

[8] Alotaibi M, Mahmood A. Improved gait recognition based on specialized deep convolutional neural networks[C] // 2015 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), October 13-15, 2015, Washington DC., USA. New York: IEEE, 2015: 15888701.

[9] Wu Z F, Huang Y Z, Wang L, *et al*. A comprehensive study on cross-view gait based human identification with deep CNNs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 209-226.

[10] Roberts L G. Machine perception of three-dimensional solids [M]. New York: Garland Publishing, 1965.

[11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the 25th International Conference on Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada, USA. USA: NIPS, 2012: 1097-1105.

[12] Du J, Hu B L, Zhang Z F. Gastric carcinoma classification based on convolutional neural network and micro-hyperspectral imaging [J]. Acta Optica Sinica, 2018, 38(6): 0617001.
杜剑, 胡炳樑, 张周锋. 基于卷积神经网络与显微高光谱的胃癌组织分类方法研究[J]. 光学学报, 2018, 38(6): 0617001.

[13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. (2015-04-10) [2018-12-17]. https://arxiv.org/abs/1409.1556.

[14] Sun Y, Wang X G, Tang X O. Deeply learned face representations are sparse, selective, and robust [J/OL]. (2014-12-03) [2018-12-17]. https://arxiv.org/abs/1412.1265.

[15] Taigman Y, Yang M, Ranzato M, *et al*. DeepFace: closing the gap to human-level performance in face verification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1701-1708.

[16] Kumar N, Berg A C, Belhumeur P N, *et al*. Attribute and simile classifiers for face verification [C] // 2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE, 2009: 365-372.

[17] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.

[18] Li S M, Lei G Q, Fan R. Depth map super-resolution reconstruction based on convolutional neural networks [J]. Acta Optica Sinica, 2017, 37(12): 1210002.
李素梅, 雷国庆, 范如. 基于卷积神经网络的深度图超分辨率重建[J]. 光学学报, 2017, 37(12): 1210002.

[19] Xiao J S, Liu E Y, Zhu L, *et al*. Improved image super-resolution algorithm based on convolutional neural network [J]. Acta Optica Sinica, 2017, 37(3): 0318011.
肖进胜, 刘恩雨, 朱力, 等. 改进的基于卷积神经网络的图像超分辨率算法[J]. 光学学报, 2017, 37(3): 0318011.

[20] Szegedy C, Liu W, Jia Y Q, *et al*. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 15523970.

[21] Wang L. Research on human action recognition based on sparse spatio-temporal features [D]. Suzhou: Soochow University, 2015: 29-36.
王露. 基于稀疏时空特征的人体行为识别研究[D]. 苏州: 苏州大学, 2015: 29-36.