

融合全局与局部特征的鞋印特征提取网络设计

辛一冉¹, 唐云祁^{1*}, 蔡能斌²¹中国人民公安大学侦查学院, 北京 100038;²上海市现场物证重点实验室, 上海 200083

摘要 为进一步解决残缺、模糊鞋印的检索困难问题,设计一种使用全局与局部特征联合表达的特征提取网络。一方面,在全局特征分支上对多尺度的鞋印全局特征进行快速归一化加权融合,并对其所有输出分别计算损失;另一方面,在局部特征分支上利用分部特征提取(PCB)模块将鞋印特征图分为3个部分,分别提取3个部分的局部特征并计算其损失。在训练阶段,将全局特征分支与局部特征分支的所有损失相加进行联合表达;在测试阶段,将两分支拼接后的输出直接展平作为待检索鞋印的描述符,并将其与样本库鞋印描述符的余弦距离作为相似性评分。实验结果表明,所提方法大幅降低模型的参数量及计算成本,并在 CSS-200、CS-Database 和 FID-300 这3个鞋印数据集上取得较高的准确率,且在 CSS-200 和 CS-Database(Dust)数据集上的 top1% 取得较好的准确率,分别为 94.5% 和 95.45%。

关键词 图像处理;深度学习;鞋印检索;低质量鞋印;特征融合;EfficientNet

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP213313

Design of Shoe Print Feature Extraction Network Integrating Global and Local Features

Xin Yiran¹, Tang Yunqi^{1*}, Cai Nengbin²¹School of Investigation, People's Public Security University of China, Beijing 100038, China;²Shanghai Key Laboratory of Crime Scene Evidence, Shanghai 200083, China

Abstract A feature extraction network using global and local features is designed to tackle the issue of retrieving incomplete and fuzzy shoe prints. The global features of the multiscale shoe print are normalized and weighted, and the losses of all their outputs are calculated; moreover, the part-based Conv baseline (PCB) module is used to divide the shoe print feature map into three parts, extract the local features of the three parts, and calculate their losses. During the training phase, all of the global feature branch and local feature branch losses are added to express them collectively. The output of the two branches after splicing is directly flattened as the shoe print descriptor to be retrieved in the test phase, and the cosine distance between it and the descriptor of the sample library shoe print is used as the similarity score. The experimental results show that the proposed method significantly reduces the parameter quantity and calculation cost of the model, and achieves high accuracy on the three shoe print data sets of CSS-200, CS Database, and FID-300. Furthermore, it achieves decent accuracy on the top1% of the CSS-200 and CS Database (Dust) datasets, which are 94.5% and 95.45%, respectively.

Key words image processing; deep learning; shoe print retrieval; low-quality shoeprint; feature fusion; EfficientNet

1 引言

在刑事案件的现场勘察过程中,鞋印是一种常见的痕迹物证,它可以为公安部门提供有关作案人的作案过程、人身特点等信息。鞋印检索作为法庭科学的

一个重要研究方向,不仅为案件的串并提供依据,还成为近些年“监控+鞋印”技战法^[1]中锁定嫌疑人的重要一环,即通过鞋印检索获取到现场鞋印所对应的鞋型信息,而后再进一步通过监控排查筛选嫌疑人。随着公安系统信息化建设的飞速发展,公安机关的鞋印样本

收稿日期: 2021-12-22; 修回日期: 2022-01-09; 录用日期: 2022-01-27; 网络首发日期: 2022-02-09

基金项目: 公安部技术研究计划项目(2020JSYJC21)、中央高校基本科研业务费项目(2021JKF203)、上海市现场物证重点实验室开放课题基金(2021XCWZK04)

通信作者: *tangyunqi@ppsuc.edu.cn

数据库中包含的鞋印花纹种类也越来越多,但现场不仅有清晰、完整的高质量鞋印,还有残缺、模糊的低质量鞋印。因此,如何在具有大量混淆样本的数据库中将现场鞋印所对应的正样本排在较为靠前的位置至关重要,好的鞋印检索算法不仅能大大增加通过鞋印找到对应鞋型信息的概率,还能节省筛选鞋印时所花费的时间。

传统鞋印检索算法的特征提取多采用尺度不变特征转换(SIFT)、方向梯度直方图(HOG)等传统特征,亦或通过手工标注特征,较为繁琐且普遍精度不高。而近些年随着深度学习的兴起,越来越多的学者将深度学习应用于鞋印检索当中,提取鞋印图片的深度特征进行检索,深度学习算法相较于传统算法不仅效率高,且泛化性更好。但现有的特征提取网络主要有以下2个问题:1)对于残缺、模糊的低质量鞋印检索效果较差;2)提取的鞋印检索的描述符过大导致计算成本较高。

针对以上2个问题,本文对EfficientNet进行改进,并使其作为鞋印检索算法的特征提取网络,希望通过引入轻量级网络达到降低计算成本的目的,同时融合多尺度的全局特征及足尖、足弓、足跟的局部特征,使得特征提取网络提取到更加丰富的鞋印花纹语义信息,对清晰、完整的高质量鞋印和残缺、模糊的低质量鞋印检索均具有较强的适用性。

2 相关工作

传统鞋印检索算法的特征提取主要依靠人工标注或提取其传统特征。例如:2015年,Wang等^[2]将鞋印划分为2个语义区域,并根据在司法实践中的优先级和可靠信息的数量计算其置信度,对于置信度较高的区域,采用基于小波-傅里叶变换的全局不变描述符计算相似度;2017年,Richetelli等^[3]比较相位相关(POC)、关键点和随机变换一致结合、傅里叶-梅林变换等3种方法,发现POC方法性能更好;2019年,彭飞^[4]将鞋印图像中可识别性较强的区域称为局部语义信息块,并选取其中清晰、重复性高的信息块,通过归一化互相关的方法匹配样本鞋印中相似度较高的区域,计算相似度,并将其与全局特征融合取得较好的效果;2020年,周思越^[5]将待检索鞋印中的标志性语义信息块和周期性语义信息块用于局部语义滤波器组的训练,并计算局部语义滤波器模型与样本鞋印之间的相似度,取得进一步的提升。这些方法多从提取鞋印的传统特征入手,但特征提取的区域多受制于人们对鞋印的认知,提取过程较为繁琐,且难以提取到鞋印丰富的语义特征。

近些年深度学习的发展如火如荼,许多学者将深度学习的方法应用于鞋印检索的过程当中:2019年,Kong等^[6]使用微调的Resnet50提取鞋印特征,并提出多通道归一化互相关的相似性度量方法,该方法在李

生神经网络中取得良好的效果,但其所需计算量过大,难以实际应用;2021年,史文韬等^[7]使用微调的vgg16网络提取鞋印特征,将Conv5-1层的输出张量直接展开作为鞋印的描述子,并引入selective convolutional descriptor aggregation(SCDA)方法^[8]对残缺鞋印的特征图进行筛选,虽然该方法在一定程度上削弱了背景噪声的干扰,但需要将完整与残缺鞋印分开进行检索,且卷积层输出后直接展开维度过高,所需计算量大。

3 基于多尺度特征加权融合的鞋印检索算法

3.1 鞋印检索算法

鞋印检索算法的目的是根据给定的现场鞋印,在样本库中查询具有同种花纹特征的鞋印图像。记 q 为待检索的现场鞋印, $\{g_1, g_2, \dots, g_n\}$ 为样本库鞋印。通过计算,现场鞋印 q 与样本库鞋印 $g_i(i \in [1, n])$ 之间的相似性评分记为 $\{d_1, d_2, \dots, d_n\}$,将 d_i 作为 g_i 的排序得分并依据排序得分输出检索结果。

鞋印检索算法流程如图1所示。首先使用自建鞋印训练集对鞋印特征提取网络进行训练,其次将待检索鞋印、样本鞋印、样本鞋印的水平翻转图共同输入全局与局部特征融合的双分支鞋印特征提取网络中得到3类鞋印图片相应的描述符,分别记为 $D_{\text{Descriptor}}$ 、 $D_{\text{Descriptor}1}$ 、 $D_{\text{Descriptor}1'}$ 。然后分别计算待检索鞋印描述符与样本鞋印描述符、样本鞋印水平翻转图描述符之间的余弦距离,记为 c_{dist} 和 c'_{dist} ,并将 $\min(c_{\text{dist}}, c'_{\text{dist}})$ 记为待检索鞋印与样本鞋印间的相似性评分 d_i ,并依据相似性评分对样本鞋印进行排序输出。 c_{dist} 和 c'_{dist} 的表达式分别为

$$c_{\text{dist}} = \frac{\|D_{\text{Descriptor}}\|^2 \cdot \|D_{\text{Descriptor}1}\|^2 - D_{\text{Descriptor}} \cdot D_{\text{Descriptor}1}}{\|D_{\text{Descriptor}}\|^2 \cdot \|D_{\text{Descriptor}1}\|^2}, \quad (1)$$

$$c'_{\text{dist}} = \frac{\|D_{\text{Descriptor}}\|^2 \cdot \|D_{\text{Descriptor}1'}\|^2 - D_{\text{Descriptor}} \cdot D_{\text{Descriptor}1'}}{\|D_{\text{Descriptor}}\|^2 \cdot \|D_{\text{Descriptor}1'}\|^2}。 \quad (2)$$

3.2 EfficientNet-B3骨干网络

EfficientNet是Tan等^[9]发表在ICML2019的工作,其采用的模型复合缩放方法在ImageNet数据集的分类任务中达到当时最高的84.3%的精度。综合考虑计算成本和精度收益,选择采用的骨干网络为EfficientNet-B3,其结构如表1所示。

从表1可以看出,该网络主要由3个部分组成。第1部分采用1层 3×3 的卷积对图片的浅层特征进行提取。第2部分共7层($P_1 \sim P_7$),每层均由MobileNet^[10]中的倒残差(MBConv)模块重复堆叠形成,MBConv模块的特征提取过程如图2(a)所示。但MBConv模块原本的squeeze-and-excitation(SE)注意力机制^[11]只

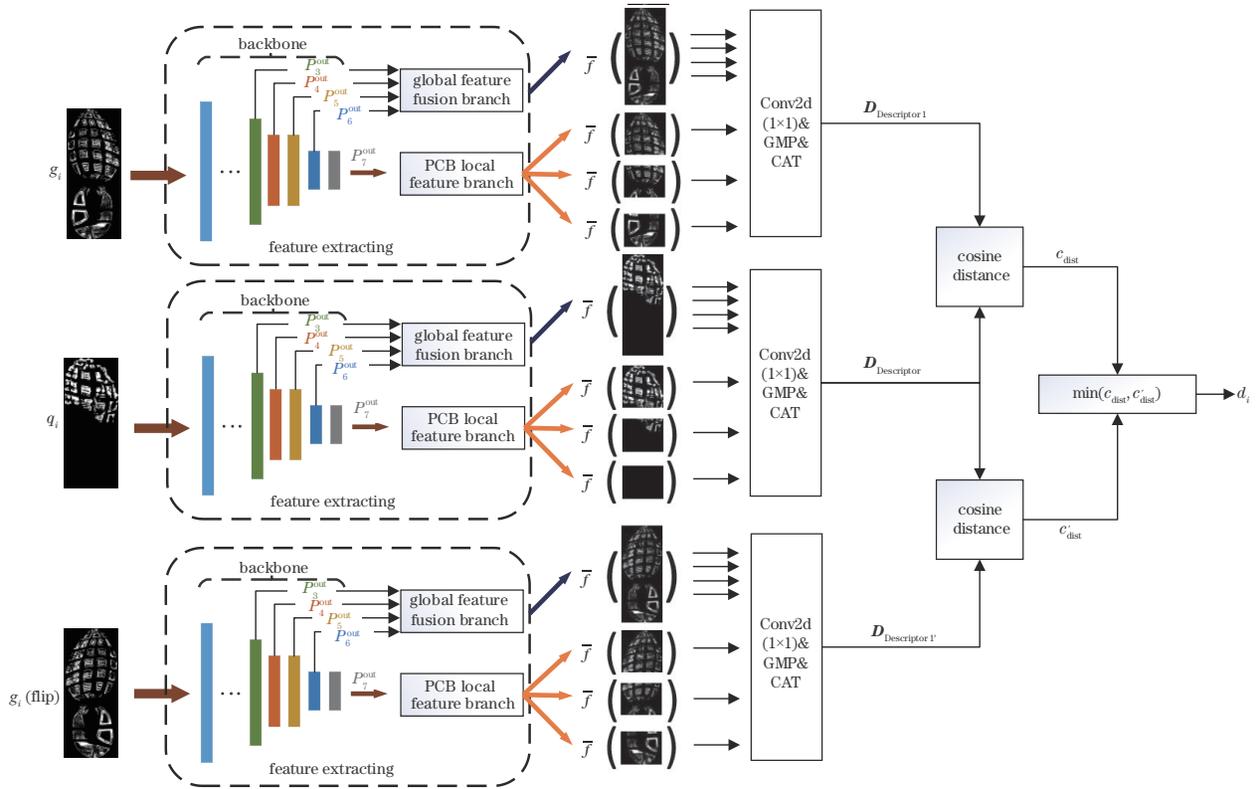


图 1 鞋印检索算法流程图

Fig. 1 Flow chart of shoe print retrieval algorithm

表 1 EfficientNet-B3 结构
Table 1 Structure of EfficientNet-B3

Stage	Operator	Resolution (H×W)	Channels	Layers
1	Conv3×3	300×300	40	1
2	MBCov1, k3×3	150×150	24	2
3	MBCov6, k3×3	75×75	32	3
4	MBCov6, k5×5	38×38	48	3
5	MBCov6, k3×3	19×19	96	5
6	MBCov6, k5×5	19×19	136	5
7	MBCov6, k5×5	10×10	232	6
8	MBCov6, k3×3	10×10	384	2
9	Conv1×1 & GMP & FC	10×10	1536	1

考虑通道间信息的重要性而忽视位置信息的重要性,且常见的关注位置信息的注意力机制如 convolutional block attention module(CBAM)^[12]等带来的计算开销并不符合本实验希望降低计算成本的目的。为了避免产生大量的计算开销的同时使得网络能更多地注意到鞋印特征的位置信息,将 EfficientNet 中的 SE 注意力机制替换为 coordinate attention(CA)注意力机制^[13]对网络结构进行优化,该注意力机制利用 2 个一维全局池化操作分别将垂直和水平方向的输入特征聚合为 2 个独立的方向感知特征图,将 2 个独立方向的注意力图乘以输入特征图来增强特征图的表示能力。改进后的 MBCov 模块如图 2(b)所示。第 3 部分采用 1×1 卷

积、批量归一化、全局最大池化层和全连接层(GMP & FC)对张量进行处理后输出。

3.3 多尺度全局特征加权融合分支

为了使特征提取网络提取到的全局特征能够包含更加丰富的深层语义信息和浅层视觉信息,所提算法在 EfficientNet-B3 骨干网络上采用特征金字塔^[14]的思想对其进行改进,选择 P_3^{in} 、 P_4^{in} 、 P_5^{in} 、 P_6^{in} 进行特征融合。全局鞋印特征提取分支如图 3 所示,融合后得到 P_3^{out} 、 P_4^{out} 、 P_5^{out} 、 P_6^{out} ,并在训练阶段分别计算其损失,得到 Loss 1、Loss 2、Loss 3、Loss 4。

需要注意的是,以往对不同尺度的特征进行融合时最常用的方式是先将特征图上采样调整为同一分辨率,然后直接相加。这种融合方式认为不同尺度的特征图所贡献的信息是平等的:

$$P_6^{\text{out}} = \text{Conv}(P_6^{\text{in}}), \quad (3)$$

$$P_5^{\text{out}} = \text{Conv}[P_5^{\text{in}} + \text{Resize}(P_6^{\text{out}})], \quad (4)$$

$$P_3^{\text{out}} = \text{Conv}[P_3^{\text{in}} + \text{Resize}(P_4^{\text{out}})]. \quad (5)$$

但是在后续的消融实验中发现不同分辨率的图像对最终的输出贡献的信息并不是完全平等的。因此,本研究采用快速归一化融合(fast normalized fusion)的特征加权融合方式^[15],其表达式为

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} \cdot I_i, \quad (6)$$

式中: ω_i 为可学习的权重; I_i 为不同分辨率的特征图; ϵ 是一个值为 0.0001 的小数,它的作用是防止权重过小

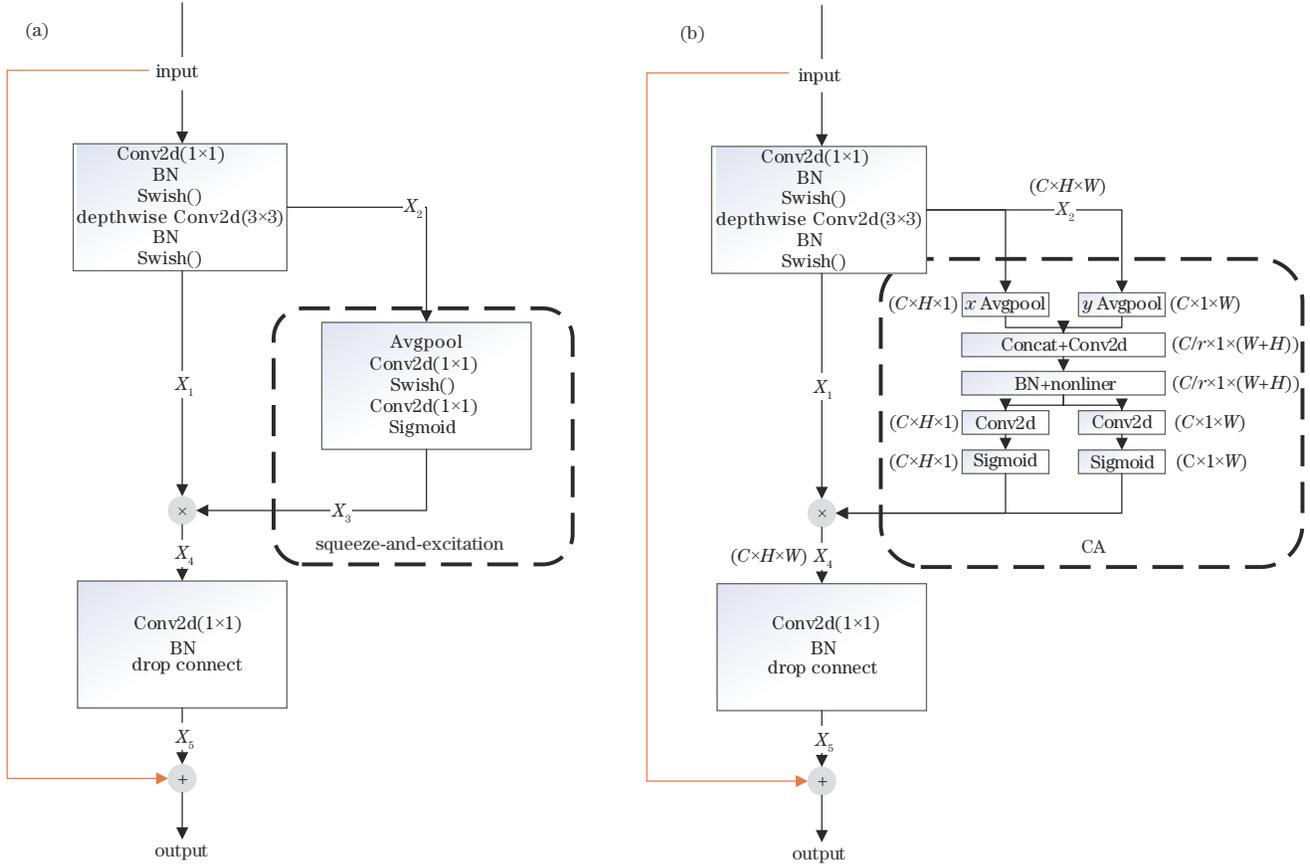


图 2 改进 MBCConv 模块结构图。(a)原始 MBCConv 模块；(b)改进 MBCConv 模块

Fig. 2 Structure of improving MBCConv block. (a) Original MBCConv block; (b) MBCConv block's improvement

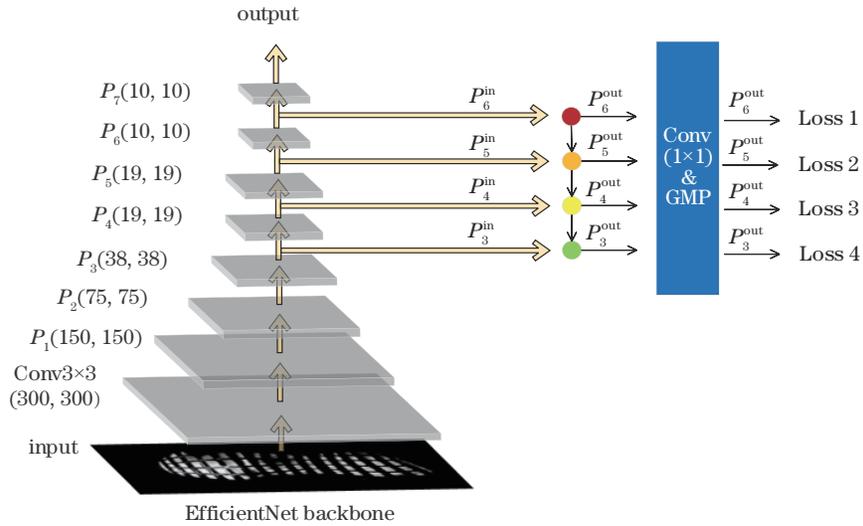


图 3 多尺度全局特征加权融合分支结构图

Fig. 3 Structure of multi-scale feature weighted fusion branch

导致模型不稳定。

在每次使用 ω_i 后,通过 ReLU 激活函数来保证 $\omega_i \geq 0$ 。这种赋值方式不仅能为不同尺寸的特征图赋予不同的重要性,而且还保证权重值保持在 0~1 之间,不会过大或过小导致训练不稳定。采用快速归一化融合后的特征融合可描述为

$$P_6^{\text{out}} = \text{Conv}(P_6^{\text{in}}), \quad (7)$$

$$P_5^{\text{out}} = \text{Conv} \left[\frac{\omega_1 \cdot P_5^{\text{in}} + \omega_2 \cdot \text{Resize}(P_6^{\text{out}})}{\omega_1 + \omega_2 + \epsilon} \right], \quad (8)$$

$$P_3^{\text{out}} = \text{Conv} \left[\frac{\omega_5 \cdot P_3^{\text{in}} + \omega_6 \cdot \text{Resize}(P_4^{\text{out}})}{\omega_5 + \omega_6 + \epsilon} \right]. \quad (9)$$

3.4 PCB分支

为了进一步增强特征提取网络对低质量鞋印的特

征提取能力,采用分部特征提取(PCB)方法^[16]提取鞋印的局部特征,并依据法庭科学对鞋印结构的划分特点将主干网络移除全连接层后得到的大小为 10×10 的特征图按 4:3:3 的比例水平切割为 3 部分,分别代表足尖、

足弓、足跟,而后对所得到的 3 个部分分别进行全局最大池化操作并使用 1×1 卷积降维,经过批量归一化和 dropout 处理后得到 3 个特征图,如图 4 所示。在训练阶段,分别计算其损失,记为 Loss 5、Loss 6、Loss 7。

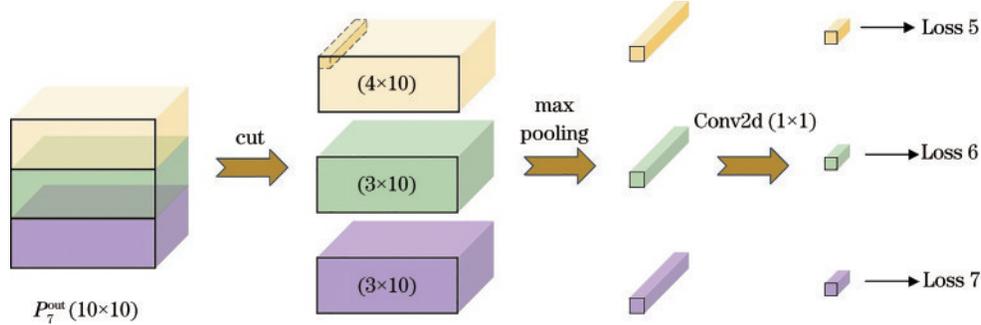


图 4 PCB 分支结构图

Fig. 4 Structure of PCB branch

3.5 度量学习

难样本采样三元组损失^[17]是一种广泛应用于图像检索任务中的度量学习损失:

$$L_{th} = \frac{1}{P \times K} \sum_{a \in \text{batch}} \left(\max_{p \in A} d_{a,p} - \min_{n \in B} d_{a,n} + \alpha \right), \quad (10)$$

式中: P 为训练所用的每一个 batch 中随机挑选的鞋印的类别数;每一类中随机挑选 K 个样本,即每一个 batch 均有 $P \times K$ 个样本;batch 中的每一个样本 a 挑选一个距离最远的正样本 p 及一个距离最近的负样本 n 组成一个三元组,在训练时缩小样本 a 与正样本 p 距离的同时将负样本 n 推远; α 是根据实际需要设定的阈值参数,本实验设置为 0.3。

标签平滑交叉熵损失^[18]在传统的交叉熵损失函数上添加标签平滑正则化,能够使模型拥有更好的泛化性:

$$L_{sm} = \beta \cdot \left[(1 - \epsilon) ce(i) + \epsilon \sum \frac{ce(j)}{N} \right], \quad (11)$$

式中: ce 是交叉熵损失; ϵ 是一个很小的正数; β 是一个超参数,代表标签平滑交叉熵损失的权重系数,本实验将其设置为 2。

由于鞋印花纹的类间差异较小,单独使用难样本采样三元组损失效果并不好,本实验最终在训练阶段使用标签平滑交叉熵损失来辅助难样本采样三元组损失进行训练,并将计算得到的损失相加进行联合表达。

$$L(n) = L_{sm}(n) + L_{th}(n), n = 1, 2, \dots, 7. \quad (12)$$

4 实验设置及数据集组成

4.1 参数设置及评价指标

实验所用操作系统为 Windows 10,显卡型号为 NVIDIA RTX 3060,显示内存为 12 GB。模型训练使用的深度学习框架为 Pytorch,编程语言为 Python,将

所有鞋印图片分辨率调整至 $500 \text{ pixel} \times 1000 \text{ pixel}$,数据增强方法采用 random resized crop、random horizontal flip 等,使用随机梯度下降(SGD)方法进行训练,epoch 设置为 200,初始学习率设置为 0.001,学习率函数采用 cosine 函数,weight decay 设置为 10^{-4} ,momentum 设置为 0.9。

衡量检索算法的常用指标有查全率、准确率和平均准确率等。所提算法主要在案件侦破的过程中为破案提供线索和依据,检索返回的正确结果排名靠前可以为办案人员节省更多的精力。因此,本实验主要采用累计匹配特性(CMC)曲线作为算法的评价指标,该曲线的横轴代表检索返回所有候选图片的前 $k\%$ 位,下文均用 $\text{top}k\%$ 表示,纵轴表示正确率。

4.2 鞋印数据集及前处理工作

为了测试所提算法在公安实战中的检索效果,采用 CS-Database^[19]、CSS-200^[20]、FID-300^[21] 这 3 个鞋印公开数据集对检索算法的有效性进行测试,3 个鞋印数据集均采集自真实的案发现场。其中,CS-Database、CSS-200 数据集均是通过我国公安机关标准采集流程所得的,3 个数据集的鞋印图片均通过公安机关现场勘察系统二值化处理后裁剪去除比例尺,并对极个别具有明显噪声影响的鞋印进行了人工处理,如图 5 所示。

同时,为了进一步贴合公安实际应用场景,3 个测试集的样本库除了具有待检索鞋印所对应的正样本外还包含大量的混淆鞋印样本(negative),具体构成如表 2 所示,其中除 CS-Database(高质量鞋印)外均为低质量现场鞋印数据集。

用于训练特征提取网络的数据集为自主构建的鞋印训练集,训练集共包含 640 类鞋印花纹:439 类采集自公安机关标准采集的样本鞋印数据库中,均为清晰完整的高质量鞋印;201 类为模仿的案发现场的残缺鞋印。为了使特征提取网络模型具有更好的泛化

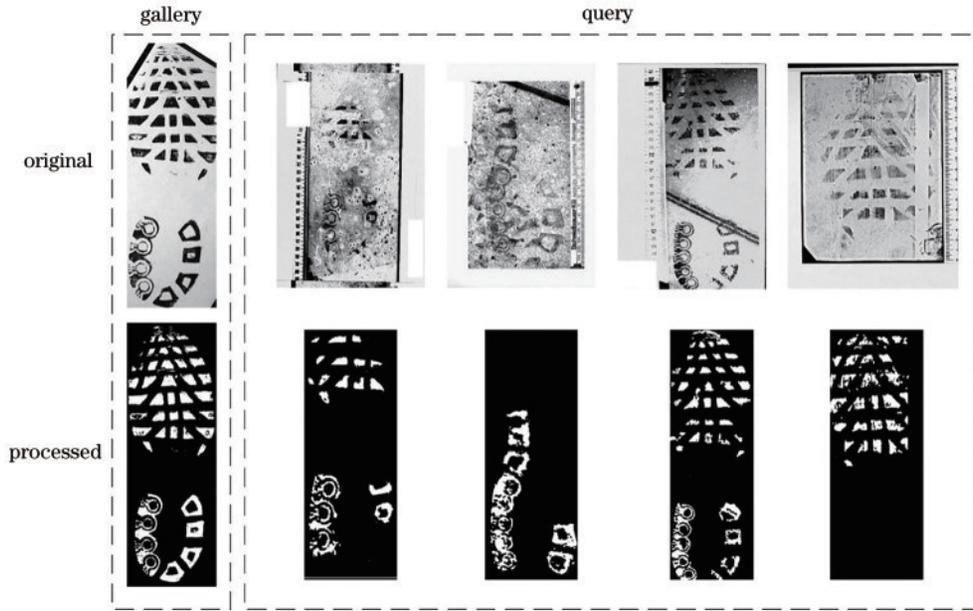


图 5 鞋印数据集前处理的过程展示
Fig. 5 Pretreatment of shoe print datasets

表 2 不同测试集的鞋印图片组成
Table 2 Composition of different data sets

Dataset	Query	Gallery	Negative
CS-Database (High-quality)	100	100	
CS-Database (Blood)	53	9	91
CS-Database (Blood enhancement)	53	9	91
CS-Database (Dust)	66	9	91
FID-300	300	300	875
CSS-200	200	200	4800

性,对原始训练集进行数据增强,包括添加高斯噪声、随机长宽比裁剪、随机旋转等操作,具体展示如图 6 所示。经数据增强后,训练集共有 640 类 26485 张鞋印图片。

5 实验结果及分析

5.1 消融实验

为了验证所提特征提取网络的有效性,分别将 EfficientNet-B3 主干网络最后的全连接层移除后所得到的输出直接展开作为鞋印图片的描述符,与将不同分支的特征融合(无 fast normalized fusion)得到的输出展开作为鞋印图片的描述符在 CSS-200 数据集上进行比较,具体结果如图 7 所示。从图 7 可以看出,相较于传统的 SE 注意力机制,CA 注意力机制显著提升 top1%、top2% 的检索精度,全局与局部双分支特征融合提取网络在 CSS-200 数据集的平均精度、累计匹配得分上均取得较大提升。而且通过图 7 可以看出,全局特征分支在进行特征融合时选择 P_3 、 P_4 、 P_5 、 P_6 这

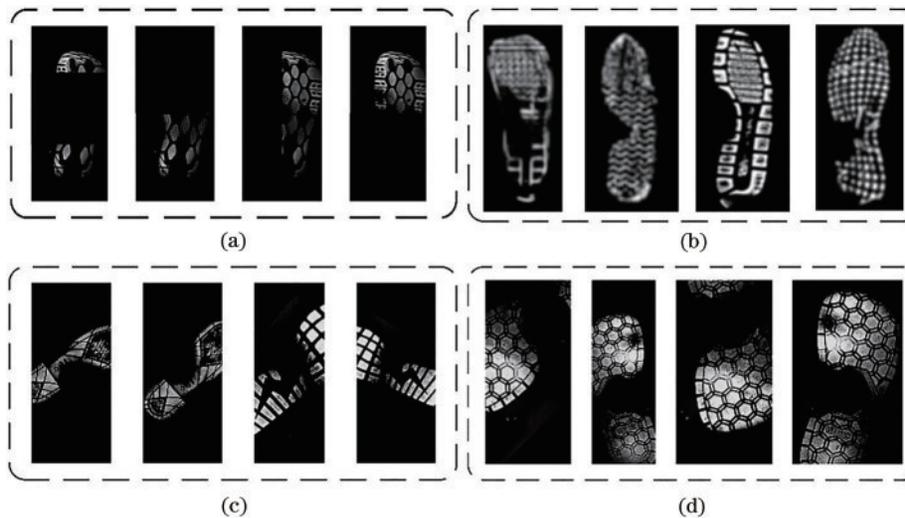


图 6 训练集数据展示。(a)随机擦除;(b)高斯噪声;(c)随机旋转;(d)随机长宽比裁剪
Fig. 6 Presentation of train dataset. (a) Random erase; (b) Gaussian noise; (c) random rotate; (d) random resized crop

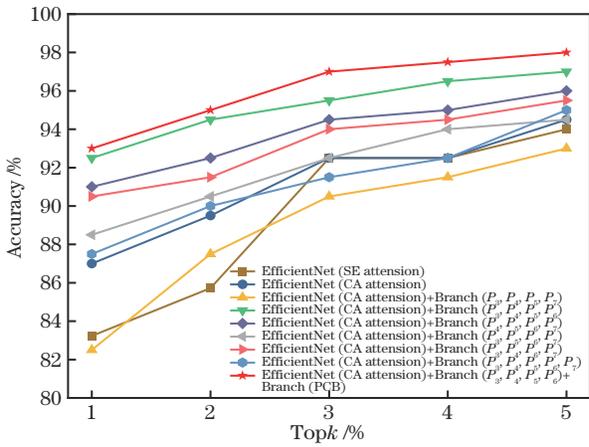


图 7 不同注意力机制、层特征融合输出在 CSS-200 数据集上的 CMC 曲线

Fig. 7 CMC curves of different attention mechanisms and layer feature fusion output on CSS-200

4 个层的输出进行特征融合效果最好,这也侧面印证了在进行特征融合时不同层所贡献的信息的重要性不同,因此简单地在上采样后直接相加并不是最好的

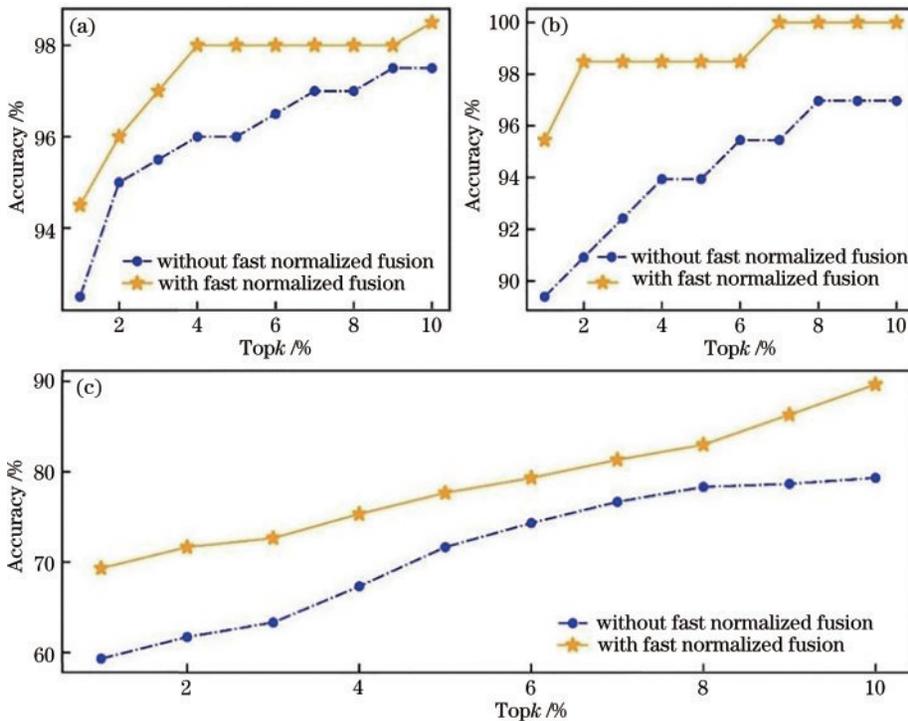


图 8 是否使用 fast normalized fusion 的 CMC 曲线。(a) CSS-200; (b) CS-Database; (c) FID-300

Fig. 8 CMC curves of whether to use fast normalized fusion. (a) CSS-200; (b) CS-Database; (c) FID-300

5.3 部分检索返回结果可视化

为了更直观地展示所提算法在残缺、模糊的低质量鞋印上的检索效果,选取 CSS-200 数据集中部分具有代表性的低质量鞋印进行检索返回结果可视化分析,可视化结果如图 9 所示。图 9 左侧为待检索鞋印,右侧为样本库中相似性评分排名前十位(rank 1~rank 10)的样本鞋印,其中绿色框表示该样本鞋印与待检索

融合方式。

为了验证全局特征融合分支在特征融合时所采用的 fast normalized fusion 的有效性,本实验比对了算法在不同测试集上是否使用 fast normalized fusion 的 CMC 曲线,如图 8 所示。从图 8 可以看出,在特征融合前对不同层赋予可学习的权重参数可以显著提升不同层所包含信息的利用率。

5.2 所提模型与其他方法的比较

为了验证所提特征提取网络的有效性,本实验选择 CS-Database、FID-300、CSS-200 这 3 个鞋印公开数据集作为测试集,选取当下先进的鞋印检索算法及常见的网络模型进行对比,包括 POC、局部语义滤波器组、混合特征和邻域图像^[22]、多通道归一化互相关(MCNCC)、选择性卷积特征描述子融合(SCDA)及一系列针对 Resnet 进行改进的模型^[10,14,23]等,具体如表 3~5 所示。综合考虑实验结果及各测试集特点,本实验所设计的特征提取网络在一定程度上提取到了更加丰富的语义信息,不仅在高质量鞋印的检索上取得了较好的结果,同时也适用于残缺、模糊的低质量鞋印的检索。

鞋印是同一类别,即正确的检索结果,红色框则表示错误的检索结果。可视化结果表明,在拥有大量混淆样本的数据库中,所提算法依然能够将低质量待检索鞋印的正确检索结果排在较为靠前的位置,这将会大大减少侦察人员在筛选正确的鞋印样本时花费的时间和精力,符合设计的初衷。

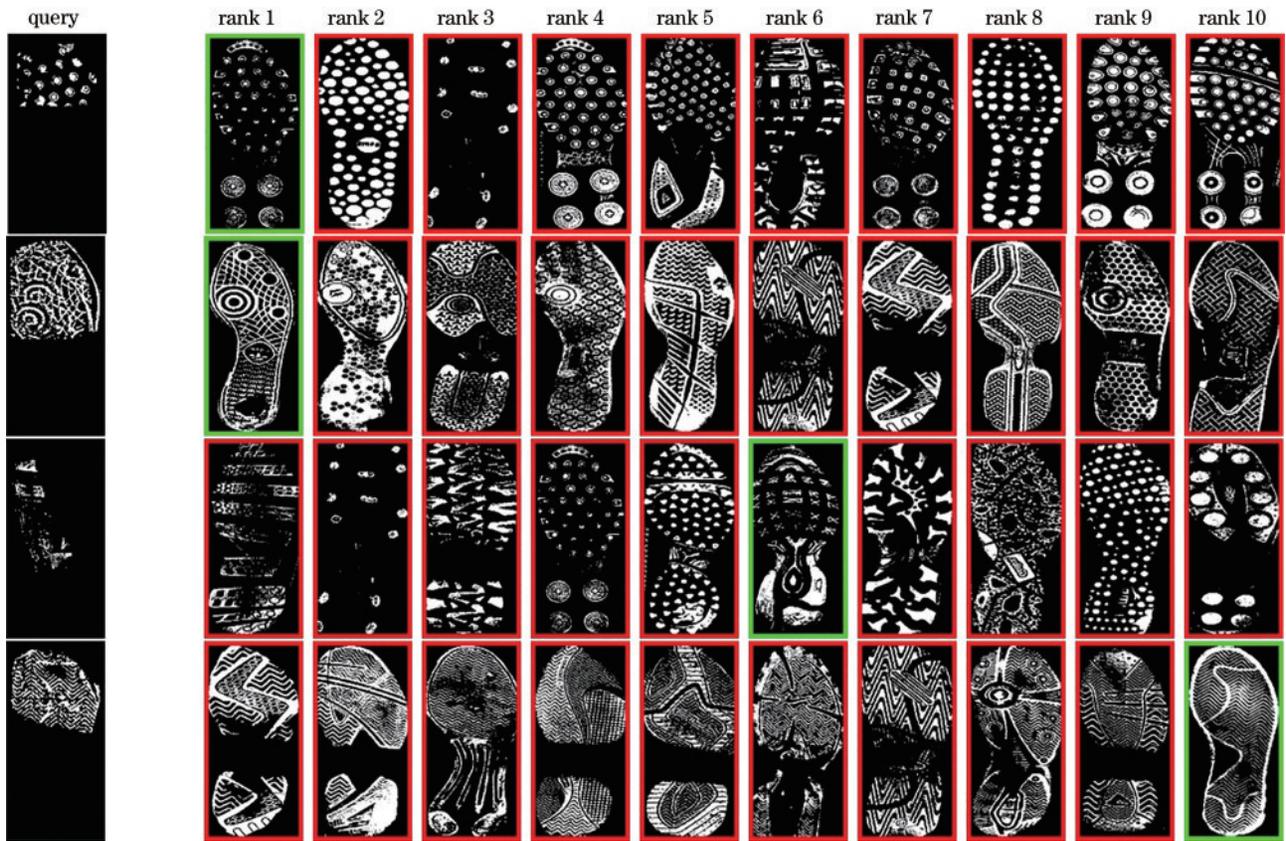


图 9 CSS-200 测试集前十位检索结果可视化

Fig. 9 Visualization of top 10 search results on CSS-200 test set

表 3 所提方法在 CS-Database 数据集上和其他方法比较

Table 3 Comparison of proposed method and other methods on CS-Database dataset

	Method	Accuracy / %				
		top1%	top2%	top3%	top4%	top5%
High-quality	POC	99	99	100	100	100
	MCNCC	99	100	100	100	100
	Local Semantic Patch and Manifold Ranking	99	100	100	100	100
	hybrid features and neighboring images	99	100	100	100	100
	Local Semantic Filter Bank	99	100	100	100	100
	SCDA	100	100	100	100	100
	Proposed method	100	100	100	100	100
Blood	POC	45.3	50.9	54.7	64.2	66
	MCNCC	92.5	98.1	100	100	100
	Local Semantic Patch and Manifold Ranking	79.2	94.3	94.3	94.3	96.2
	hybrid features and neighboring images	92.5	100	100	100	100
	Local Semantic Filter Bank	94.3	100	100	100	100
	SCDA	100	100	100	100	100
	Proposed method	100	100	100	100	100
Dust	POC	47	47	51.5	51.5	54.5
	MCNCC	86.4	89.4	90.9	90.9	95.5
	Local Semantic Patch and Manifold Ranking	83.3	84.8	87.9	90.9	90.9
	hybrid features and neighboring images	89.4	93.9	95.5	95.5	95.5
	Local Semantic Filter Bank	95.45	96.97	98.48	98.48	98.48
	SCDA	87.87	90.9	90.9	90.9	90.9
	Proposed method	95.45	98.48	98.48	98.48	98.48

表 4 所提方法在 CSS-200 数据集上和其他方法比较

Table 4 Comparison of proposed method and other methods on CSS-200 dataset

Method	Accuracy / %				Params	FLOPs
	top1%	top2%	top5%	top10%		
Conv-25088	62.5	85.5	92.5	94	116M	1.6×10^{10}
SCDA	92.5	94.5	95	96.5	138M	1.9×10^{10}
Resnet101	59.5	80.5	82.5	91.5	49M	7.6×10^9
ResneXt	64	87	91	93.5	53M	4.2×10^9
Resnet50-FPN	78	85.5	90	91.5	60M	9.1×10^9
SENet	79.5	85	85	93.5	146M	2.1×10^{10}
Proposed method	94.5	96.5	98	100	23M	2.6×10^9

表 5 所提方法在 FID-300 数据集上和其他方法比较

Table 5 Comparison of proposed method and other methods on FID-300 dataset

Method	Accuracy / %	
	top1%	top10%
Local Semantic Filter Bank	73	93.7
Local Semantic Patch and Manifold Ranking	61.3	88
hybrid features and neighboring images	71.8	87.3
MCNCC	79	89
CABM	58	79
Conv-25088	47.16	76.59
SCDA	50.84	79.93
Proposed method	69.8	90.94

6 结 论

所提算法在 CS-Database 和 CSS-200 数据集上均取得较好的检索结果,CS-Database (Dust、Blood)、CSS-200 数据集的查询库都是由较为模糊、残缺的现场鞋印图像组成的,CS-Database(High-quality)的查询库是高质量的公安机关标准采集鞋印,这说明所提算法在模糊、残缺的低质量鞋印及清晰、完整的高质量鞋印特征提取上均具有良好的泛化性。在 FID-300 数据集上同样取得较好的结果,但与当下最好的算法仍有差距,主要是由于该数据集的鞋印采集标准与我国公安机关采集标准存在差异,并非垂直于鞋印进行拍摄,导致鞋印的尺度差距过大、形变严重,对识别造成了很大困难。综合来看,所提算法通过引入轻量级网络 EfficientNet 大幅减小了模型的参数量,降低了计算成本,且采用全局与局部特征融合的方式进一步提高了模型在不同质量鞋印上的泛化能力。

参 考 文 献

[1] 杨智雄,唐云祁,张家钧,等.基于改进YOLOv4的行人鞋部检测算法[J].激光与光电子学进展,2022,59(8):0810007.
Yang Z X, Tang Y Q, Zhang J J, et al. Detection algorithm of pedestrian shoe area based on improved YOLOv4[J]. Laser & Optoelectronics Progress, 2022,

59(8): 0810007.
[2] Wang X N, Sun H H, Yu Q, et al. Automatic shoeprint retrieval algorithm for real crime scenes[M]//Cremers D, Reid I, Saito H, et al. Computer vision-ACCV 2014. Lecture notes in computer science. Cham: Springer, 2015, 9003: 399-413.
[3] Richetelli N, Lee M C, Lasky C A, et al. Classification of footwear outsole patterns using Fourier transform and local interest points[J]. Forensic Science International, 2017, 275: 102-109.
[4] 彭飞.基于局部语义块和流行排序的鞋印图像检索[D].大连:大连海事大学,2019:11-18.
Peng F. Local semantic patch and manifold ranking based shoeprint retrieval[D]. Dalian: Dalian Maritime University, 2019: 11-18.
[5] 周思越.基于局部语义滤波器组的低质量鞋印图像检索算法[D].大连:大连海事大学,2020:15-39.
Zhou S Y. Local semantic filter bank based low quality shoeprint image retrieval[D]. Dalian: Dalian Maritime University, 2020: 15-39.
[6] Kong B, Iii J S, Ramanan D, et al. Cross-domain image matching with deep feature maps[J]. International Journal of Computer Vision, 2019, 127(11): 1738-1750.
[7] 史文韬,唐云祁.基于选择性卷积特征描述子融合的鞋印检索算法[J].科学技术与工程,2021,21(16):6772-6779.
Shi W T, Tang Y Q. Shoeprints retrieval algorithm based on selective convolutional descriptor aggregation[J]. Science Technology and Engineering, 2021, 21(16): 6772-6779.
[8] Wei X S, Luo J H, Wu J X, et al. Selective convolutional descriptor aggregation for fine-grained image retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881.
[9] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28)[2021-11-25]. <https://arxiv.org/abs/1905.11946>.
[10] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-11-25]. <https://arxiv.org/abs/1704.04861>.
[11] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
[12] Woo S, Park J, Lee J Y, et al. CBAM: convolutional

- block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [13] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [14] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [15] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.
- [16] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and A strong convolutional baseline) [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.
- [17] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-03-22) [2021-11-25]. <https://arxiv.org/abs/1703.07737>.
- [18] Müller R, Kornblith S, Hinton G. When does label smoothing help? [EB/OL]. (2020-06-10) [2021-11-25]. <https://arxiv.org/abs/1906.02629>.
- [19] 韩雨彤. 基于深度学习网络的鞋印检索算法[D]. 北京: 中国人民公安大学, 2022: 21-24.
Han Y T. Shoe printing algorithm based on deep learning network[D]. Beijing: Chinese People's Public Security University, 2022: 21-24.
- [20] 史文韬, 唐云祁. 基于微调 VGG-16 的现场鞋印检索算法[J]. 中国人民公安大学学报(自然科学版), 2020, 26(3): 22-29.
Shi W T, Tang Y Q. Research on forensic shoeprint retrieval algorithm by fine-tuning VGG-16[J]. Journal of People's Public Security University of China (Science and Technology), 2020, 26(3): 22-29.
- [21] Kortylewski A, Albrecht T, Vetter T. Unsupervised footwear impression analysis and retrieval from crime scene data[M]//Jawahar C V, Shan S G. Computer vision-ACCV 2014 workshops. Lecture notes in computer science. Cham: Springer, 2015, 9008: 644-658.
- [22] Wu Y J, Wang X N, Zhang T. Crime scene shoeprint retrieval using hybrid features and neighboring images[J]. Information, 2019, 10(2): 45-60.
- [23] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.