

融合自适应感受野与多支路特征的鞋型识别算法

张家钧, 唐云祁, 杨智雄

(中国人民公安大学 侦查学院, 北京 100032)

摘要: 随着监控摄像头的普及和图侦技术的快速发展,“鞋印+监控”技战法成为公安机关侦破案件的重要手段。该技战法根据现场嫌疑鞋印推断出嫌疑鞋型,进而在犯罪现场周围监控视频中查找对应鞋型,锁定犯罪嫌疑人。然而现有鞋型识别算法无法充分提取嫌疑鞋印的重要特征,导致识别准确率降低。针对该问题,提出一种融合自适应感受野模块与多支路特征的鞋型识别算法。通过设计一种自适应感受野模块,使网络自适应选择合适大小的感受野特征,增强网络的特征提取能力,同时构建多支路特征融合模型,融合网络的深层和浅层特征,以充分利用有效特征进行鞋型识别,从而提高识别精度。在此基础上,采用中心损失函数和标签平滑损失函数联合训练的方法,在增大类间差距的同时缩小类内差距,增强模型的泛化能力。在多背景鞋型数据集上进行实验,结果表明,该算法 Rank-1 和 mAP 精度分别为 79.77% 和 62.18%,具有较优的识别效果,为公安刑侦实战提供了一种可行方案。

关键词: 鞋型识别;自适应感受野;特征融合;中心损失函数;标签平滑

开放科学(资源服务)标志码(OSID):



中文引用格式: 张家钧,唐云祁,杨智雄.融合自适应感受野与多支路特征的鞋型识别算法[J].计算机工程,2022,48(6):295-303.

英文引用格式: ZHANG J J, TANG Y Q, YANG Z X. Shoe type recognition algorithm with adaptive receptive field and multi-branch feature[J]. Computer Engineering, 2022, 48(6): 295-303.

Shoe Type Recognition Algorithm with Adaptive Receptive Field and Multi-Branch Feature

ZHANG Jiajun, TANG Yunqi, YANG Zhixiong

(School of Investigation, People's Public Security University of China, Beijing 100032, China)

[Abstract] With the popularization of surveillance cameras and the rapid development of graphic investigation technology, the "shoe print and surveillance" technique has become an essential means for public security organs to detect cases. This technique infers the shoe type of the suspect from the shoe print on the scene, then search the corresponding shoe type from the surveillance video around the crime scene, and then locates the suspect. A shoe type recognition algorithm based on adaptive receptive field module and multi-branch feature fusion is proposed to address the low automation problems and insufficient extraction of important shoe features by existing algorithms. An adaptive receptive field module is designed to help the network adaptively select suitable receptive field features, and then enhance the network feature extraction ability. A multi-branch feature fusion model is constructed to fuse the deep and shallow features in the network, using effective features to improve the recognition accuracy. Based on this, the Center Loss function and Label Smoothing (LS) function are used to jointly train the square method, which can effectively reduce the intra-class spacing while increasing the inter-class spacing and enhancing the generalization performance of the model. Experiments are carried out on the established multi-background shoe data set, and the results show that the algorithm is improved by rank-1 and mAP reaches 79.77% and 62.18%, respectively, which is more accurate recognition effect and provides a feasible scheme for combating public security criminal investigation.

[Key words] shoe type recognition; adaptive receptive field; feature fusion; Center Loss function; Label Smoothing (LS)

DOI: 10. 19678/j. issn. 1000-3428. 0061942

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)。

作者简介: 张家钧(1996—),男,硕士研究生,主研方向为刑事智能技术;唐云祁(通信作者),副教授、博士生导师;杨智雄,硕士研究生。

收稿日期: 2021-06-16 **修回日期:** 2021-07-20 **E-mail:** ppsuczjj@163.com

0 概述

通过作案人遗留在犯罪现场的嫌疑鞋印推断嫌疑鞋型,进而在监控视频中追踪犯罪嫌疑人已经成为公安机关侦破案件的重要技战法。随着监控摄像头的普及,该技战法在侦查破案中具有重要作用。文献[1]介绍2016年某超市内发生的一起命案,现场唯一有价值的物证是数枚来自同一人的带血鞋印。在该案件中,侦查人员通过特征标示、拼接比较、重合比较等方法确定监控视频中一男子所穿鞋型与案发现场鞋型为同类鞋型,在监控视频中成功锁定犯罪嫌疑人,实现了从案发现场遗留鞋印到监控的跨模态追踪溯源。

“鞋印+监控”技战法虽然有较强的实战价值,但是在监控视频中比对嫌疑鞋型的工作全部由人工完成。我国警力尚处于不足的状态,无法确保在规定时间内精准搜索到犯罪嫌疑人。因此,一种面向监控视频的鞋型自动识别算法成为研究热点。相比人脸识别问题,面向监控视频的鞋型识别问题更复杂,其原因受运动、光照、分辨率较低等因素的影响,在监控视频中的鞋子区域大多是模糊不清的,导致可利用有效特征较少。因此,通过提取低分辨率鞋子影像有效特征进行鞋型自动识别是现阶段亟须解决的难题。

针对上述问题,本文提出一种基于自适应感受野模块与多支路特征融合的鞋型识别算法。通过构建自适应感受野模块(Adaptive Receptive Field Module, ARFM),在模块末端连接通道注意力机制,使网络自动选择合适大小的感受野特征,设计三支路特征融合模型,充分利用有效特征进行鞋型识别,采用Center Loss^[2]和标签平滑损失^[3]联合函数对网络进行训练,使样本实现更好的聚类效果。

1 相关工作

基于内容的图像检索(Content-Based Image Retrieval, CBIR)是计算机视觉领域的研究热点。在深度学习发展之前,CBIR主要基于手工标注的特征,如尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)^[4],因手工特征描述受限,导致检索效率较低。自深度学习得到迅速发展之后,特别是AlexNet^[5]、VGGNet^[6]、GoogLeNet^[7]、ResNet^[8]、DenseNet^[9]等深度卷积神经网络的出现,得益于其强大的特征提取能力,深度学习广泛应用在图像分类、目标检测^[10]、图像检索^[11]等领域。

文献[12]提出全连接层具有较强的语义特征描述能力,并且其本身是向量形式,因此在早期工作中,直接选择全连接层作为图像的特征表示向量,但是仅选择一层全连接层会限制网络的检索性能。因此,文献[13]提出融合多层全连接层特征进行图像检索,取得更优的检索结果。文献[14]对GoogLeNet

的3组inception结构进行平均池化、 1×1 卷积和全连接层操作,并将这3组提取到的1 024维特征进行拼接,并对得到3 072维特征向量进行检索,其检索结果优于传统的特征提取算法。文献[15]指出,全连接层对图像分类的贡献突出,但是缺乏图像细节特征和局部几何不变特性,对图像的尺寸、位置等变化较敏感。因此,研究人员将卷积层的输出作为图像特征表示向量,卷积层中神经元仅连接特征图中的部分区域并且参数共享,具有对图像几何变换的有效性。将卷积层的输出作为图像特征表示向量需要选择合适的卷积特征聚合方法,如SPoC^[16]、CroW^[17]、R-MAC^[18]、SCDA^[19]、PWA^[20]、GEM^[21]等方法。文献[22]基于CroW和SCDA方法提出深度卷积特征聚合(DFW)方法,综合考虑深度卷积特征的位置、区域和通道的重要性,并对特征进行加权聚合,以获得更优的检索效果。文献[23]基于CroW方法提出一种新的空间和通道特征加权聚合方法,生成差异性加权向量,并对权重矩阵进行滤波处理,具有较优的检索结果。

文献[24]提出在卷积神经网络中每层提取的特征具有层次性,低层特征主要包含纹理、边缘等细节特征,随着层数的加深,网络感受野逐渐增大,高层特征主要包含高级语义特征。基于此,文献[25]融合低层细节特征和高层语义特征,提高网络检索性能。文献[26]提出融合全局特征和局部特征的两阶段的图像检索方法,利用全局特征进行检索得到top30结果,采用局部特征进行重新检索排序,在Google Landmarks dataset v2数据集上达到最高的检索精度。文献[15]提出多层特征融合的检索精度相比于单独使用低层特征或者高层特征的检索精度高。文献[27]提出一种三支路特征层融合模型,使网络充分利用有效特征进行行人重识别,实验结果表明,多层特征融合模型具有更优的检索性能。针对传统神经网络在学习目标高层语义特征时存在目标边缘、纹理等浅层特征丢失的问题,文献[28]采用Haar-like提取目标边缘、纹理特征,同时利用Adaboost进行特征降维,以提取目标的边缘、纹理浅层特征,将目标浅层特征与神经网络提取的深层特征相融合,能够有效提升检测精度。文献[29]通过Gabor滤波器获取目标不同方位的特征,利用MS-CLBP获取目标的局部纹理、空间和轮廓信息,并将浅层特征与神经网络学习到的高层语义特征相融合输入到SVM分类器中,得到优于当时最先进方法的分类精度。

针对监控视频中行人所穿鞋型自动识别研究较少的问题,本文将深度学习运用到鞋型识别工作中,不仅取代人工盯查监控工作,大幅加快鞋型匹配速度,还可以避免人工盯查监控时受主观因素的影响,提升公安机关侦破案件的效率。

2 本文算法

2.1 自适应感受野

感受野是指神经网络中每层输出特征图的像素点在原始图像上映射的区域大小,网络感受野越大表示其接触到原图像范围越大,所包含的特征信息更加全面。增大感受野的主要方式是增加网络深度

和采用下采样操作,在深度学习初期,网络层数越深,训练得到的模型效果越好,但是通过池化层增大网络感受野的弊端是随着图像分辨率的降低,图像的细节信息也会随之丢失。因此,本文提出一种自适应感受野模块,其骨干网络采用 ResNet50,模块结构如图 1 所示。

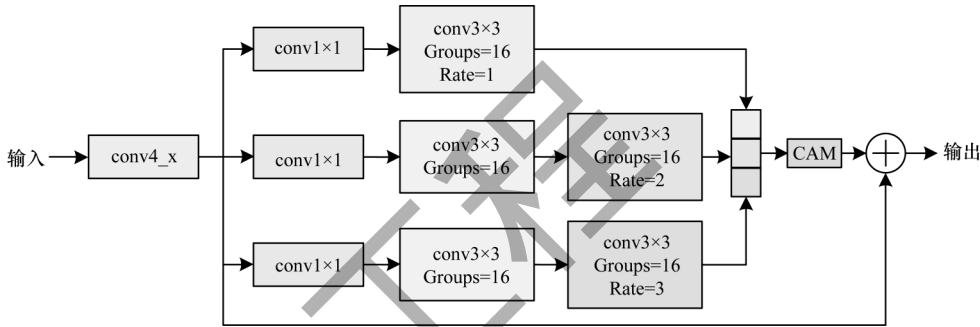


图 1 自适应感受野模块
Fig.1 Adaptive receptive field module

为获得不同大小的网络感受野,本文设计 3 条卷积支路,每条支路采用不同空洞率的空洞卷积^[30]。本文在 3×3 卷积层中采用分组卷积^[31],Groups 设置为 16,减少模块参数量,使自适应感受野模块轻量

化,在模块末端连接通道注意力机制^[32],实现自适应选择合适大小的感受野特征^[33],从而提升鞋型识别性能。图 2 为引入自适应感受野模块之后的网络架构。

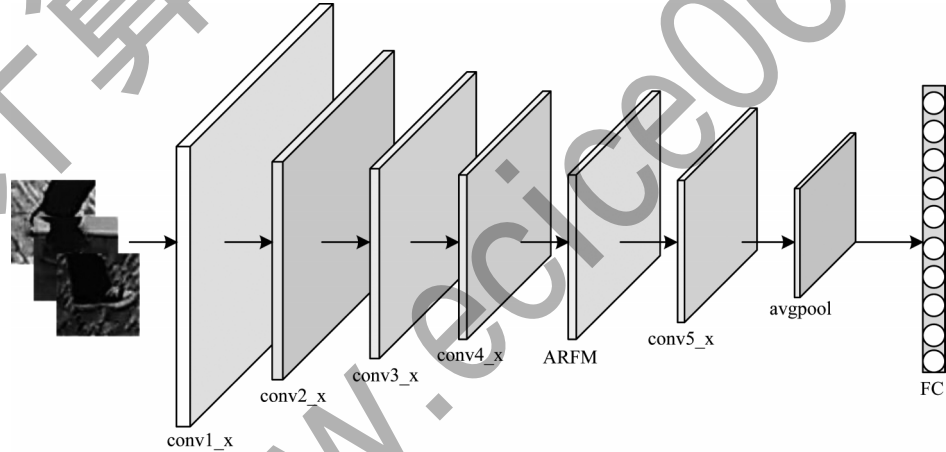


图 2 本文网络架构
Fig.2 Architecture of the proposed network

2.1.1 空洞卷积

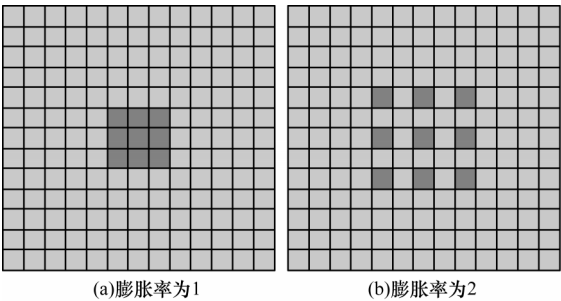
空洞卷积是指在传统卷积的基础上增加零填充,通过设置不同的膨胀率,在不增加额外参数的情况下扩大网络感受野。感受野的计算如式(1)所示:

$$r_n=r_{n-1}+(k_n-1)\prod_{i=1}^{n-1}s_i$$
(1)

其中: r_n 为该层网络感受野大小; r_{n-1} 为前层网络感受野大小; k_n 为该层卷积核大小; s_i 为*i*层步长。

不同膨胀率空洞卷积示意图如图 3 所示,图 3(a)为传统的标准卷积操作,感受野大小为 3×3。图 3(b)表示空洞卷积的膨胀率大小为 2,通过增加零填充,此时感受野大小为 5×5。空洞卷积在不增加

额外计算量的同时仅使用不同的膨胀率获得不同大小的感受野,以捕获更加全面的特征信息。



(a)膨胀率为1 (b)膨胀率为2

图 3 不同膨胀率的空洞卷积示意图

Fig.3 Schematic diagram of cavity convolution with different dilation rates

空洞卷积卷积核的计算如式(2)所示:

$$K_n = k_n + (k_n - 1) \times (d - 1) \quad (2)$$

其中: K_n 为空洞卷积卷积核大小; k_n 为真实卷积核大小; d 为空洞卷积使用的膨胀率参数。

2.1.2 分组卷积

本文在 3×3 卷积层中采用分组卷积减少参数量,使自适应感受野模块更加轻量化。分组卷积首先对输入的特征图进行分组,在每组特征图中再进行卷积操作。假设某层输入特征图的通道数、宽、高、输出通道数分别为 C 、 W 、 H 、 K ,采用传统标准卷积方式的参数量 P_1 如式(3)所示:

$$P_1 = C \times W \times H \times K \quad (3)$$

分组卷积将传统标准卷积分为 G 组,其参数量

P_2 为 $G \times \frac{C}{G} \times W \times H \times \frac{K}{G}$,如式(4)所示:

$$P_2 = \frac{C \times W \times H \times K}{G} \quad (4)$$

采用分组卷积的参数量是传统标准卷积参数量的 $\frac{1}{G}$,能够有效减少网络参数量,提高网络训练效率。本文分组数 G 设置为16。

2.1.3 通道注意力机制

注意力机制使得网络自适应选择对当前任务更关键的特征信息。在本文提出的自适应感受野模块中,3条支路分别代表不同大小的感受野特征。本文在模块末端连接通道注意力机制,使每条支路具有不同重要性的权重,从而实现网络自适应选择合适大小感受野特征的目的。通道注意力机制模型如图4所示。

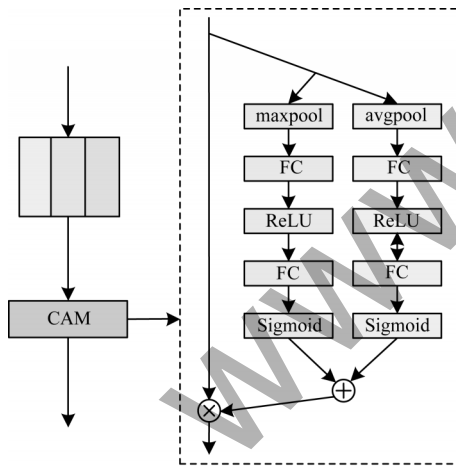


图4 通道注意力机制模型

Fig.4 Channel attention mechanism module

首先对输入特征进行压缩,即通过平均池化和最大池化整合各通道的信息,之后通过两层全连接建模通道间的相关性,获取每条支路的重要性权重,通过Sigmoid激活函数获取0~1之间的归一化权重,将两条支路的归一化权重系数相加后加权到各通道特征中,实现网络自适应选择合适大小的感受

野特征。注意力权重系数 Q 如式(5)所示:

$$Q = \sigma \left(Q_1 \left(\delta \left(Q_0 \left(\text{avgpool}(X) \right) \right) \right) \right) + \sigma \left(Q_1 \left(\delta \left(Q_0 \left(\text{maxpool}(X) \right) \right) \right) \right) \quad (5)$$

其中: $X \in \mathbb{R}^{c \times h \times w}$ 为上层输出特征; $Q_0 \in \mathbb{R}^{\frac{c}{r} \times 1 \times 1}$ 、 $Q_1 \in \mathbb{R}^{c \times 1 \times 1}$ 分别为经过第1层和第2层FC层的特征; δ 为ReLU激活函数; σ 为Sigmoid激活函数。

2.2 实例与批量标准化

批量标准化(Batch Normalization, BN)^[34]是卷积神经网络中数据归一化常用的方法,BN对输入每批次的数据进行归一化处理,使每层的输出归一化至均值为0和方差为1的分布,确保数据分布的一致性。实例标准化(Instance Normalization, IN)与BN相反,IN仅作用于单张图片,通过对单张图片的所有像素求均值和标准差,可以降低场景迁移(如背景、光照等变化)时对识别效果产生的影响。本文使用多背景数据集,且光照、角度等条件均不同,利用IN减少外观差异产生的影响,但是IN在减小同类个体差异的同时也会损失一些有效特征信息。BN保留较多区分不同个体的特征信息,但是该特征信息受外观影响较大。因此,结合IN和BN两者的优点,本文采用文献[35]提出的IBN-Net,一半通道使用IN,剩余通道使用BN。在卷积神经网络中,由外观因素带来的特征差异主要存在于低层中,高层受外观差异产生的影响较小。本文在conv2_x中每组卷积块的第1层卷积层之后使用IBN结构,增强网络识别能力。实例与批量标准化结构如图5所示。

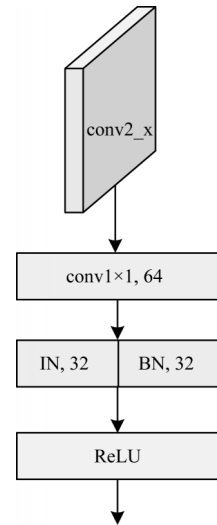


图5 实例与批量标准化结构

Fig.5 Structure of instance and batch normalization

2.3 多支路特征融合模型

本文提出的多支路特征融合模型能够充分利用浅层特征和深层特征,通过特征融合方式弥补神经网络无法充分利用有效特征进行识别的不足。在卷积神经网络中,卷积层提取的特征各不相同,浅层特

征一般是纹理、边缘等细节特征信息,深层特征通常是高级语义特征。深层特征虽然具有语义表达能力,但是如果网络单独使用深层特征会损失图像细节信息,从而影响鞋型识别性能。本文选择 ResNet50 网络中的 conv2_x、conv3_x、conv5_x 这 3 个模块的输出特征,以充分利用有效特征,以 conv5_x 输出特征为主,conv2_x、conv3_x 输出特征为辅进行特征融合,舍弃 conv4_x 模块的原因是 conv4_x 的语义特征没有 conv5_x 模块明显,其细节等特征信息没有 conv2_x 和 conv3_x 模块显著。首先 Branch 1 引出 conv2_x 模块的输出,获得与

conv5_x 相同大小的输出特征图,经过全局平均池化得到大小为 $4 \times 4 \times 256$ 的特征 F_{Branch1} ;同理,Branch 2 将 conv3_x 模块的输出引出,经过全局平均池化得到大小为 $4 \times 4 \times 512$ 的特征 F_{Branch2} ;Branch 3 直接将 conv5_x 的输出引出,得到大小为 $4 \times 4 \times 2048$ 的特征 F_{Branch3} 。将经过统一尺寸的特征 F_{Branch1} 、 F_{Branch2} 、 F_{Branch3} 融合之后得到大小为 $4 \times 4 \times 2816$ 的特征 F 。本文在特征 F 之后加入降维模块 (Dimensionality Reduction Module, DRM),以降低维度和增加特征 F 的非线性特性,从而增强网络表达能力。多支路特征融合模型如图 6 所示。

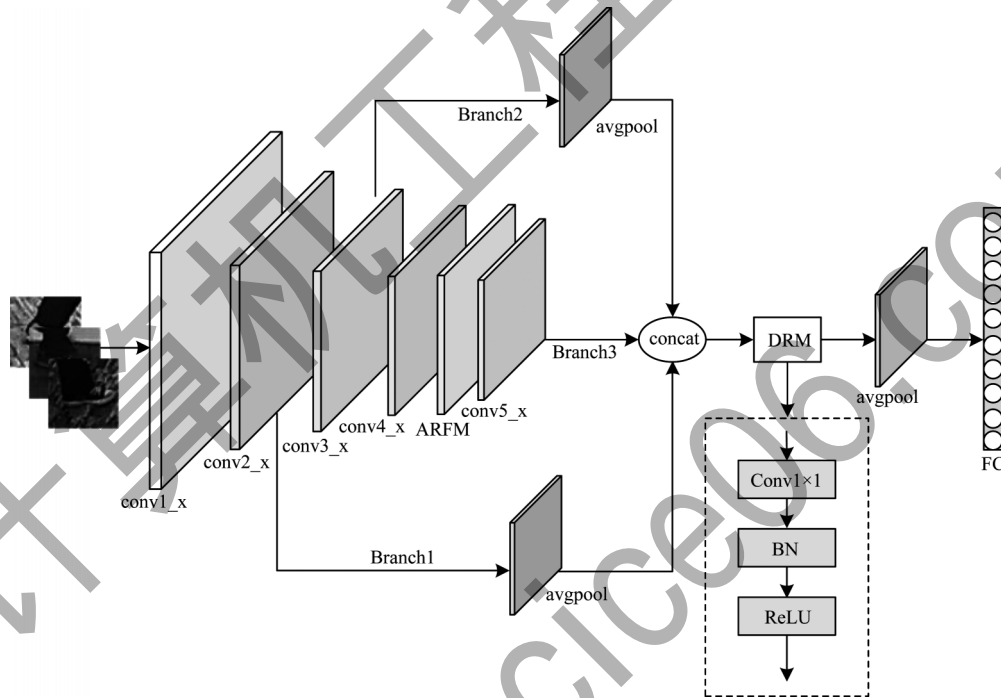


图6 多支路特征融合模型

Fig.6 Multi-branch feature fusion model

2.4 损失函数

本文数据集中鞋类较多,且相似度较高,在训练过程中存在类间差距不明显、类内差距较大等问题。因此,在增大类间差距的同时缩小类内差距,提高模型的识别精度,使样本达到更好的聚类效果。

传统的 Softmax 损失函数将整个空间按照类别个数进行划分,在分类任务中具有较优的效果,但是 Softmax 函数并没有强调类间分离和类内紧凑。基于此,本文提出联合 Center Loss 训练的方法。Center Loss 作为一种辅助训练损失函数,能够有效缩小类内差距并保持类间差异,Center Loss 函数如式(6)所示:

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (6)$$

其中: m 为批次大小; x_i 为全连接层之前的特征; c_{y_i} 为第 y_i 个类别的特征中心。本文在 Softmax 损失函数中加入 LS (Label Smoothing),通过在输出中引入噪

声,降低网络训练过拟合现象的发生及网络对真实标签的依赖性,以提高模型的泛化能力。本文最终采用的损失函数 L 如式(7)所示:

$$L = L_{\text{Label}} + \lambda L_c \quad (7)$$

其中: L_{Label} 为加入 Label Smoothing 的 Softmax 损失函数; L_c 为 Center Loss 函数; λ 为 Center Loss 函数的权重,本文 λ 取值 0.000 1。

3 实验与结果分析

3.1 实验设置

本文实验操作系统为 Linux 3.10.0,是基于 PyTorch 深度学习框架展开的算法研究。CPU 为 Intel® Xeon® CPU E5-2650 v4 2.20 GHz,显卡设置为 NVIDIA TITAN X (Pascal),显存 12 GHz,深度学习平台为 PyTorch1.2.0,编译环境为 Python3.5.6。输入图像尺寸为 120×120 像素,采用随机裁剪、水平翻转方式进行数据增强处理,采用 Adam 作为优化器,初

始学习率设置为0.000 1。

3.2 数据集

为验证本文方法的有效性,本文在构建的多背景数据集上进行测试。数据集是由3个不同视角的监控摄像头在中国人民公安大学足迹实验室采集的图像,300类鞋共35 300张低分辨率鞋子图像。鞋子图像由人工手动标注完成,训练集包含150双鞋30 000张鞋子图像。测试集包含150双鞋300张鞋子图像。鞋样数据库包含5 000张混淆样本和300张样本鞋子图像。多背景数据集的部分数据样式如图7所示。



图7 多背景数据集的部分数据
Fig.7 Partial data of multi-background datasets

3.3 评价指标

为评估模型性能,本文将均值平均精度(mean Average Precision, mAP)和Rank-1作为评价指标。Rank-1反映识别排序结果第一位的匹配正确率;mAP反映模型的整体性能,如下:

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}} \tag{8}$$

$$A_{AP} = \frac{\sum P}{\text{images}} \tag{9}$$

$$m_{mAP} = \frac{\sum_{i=0}^N A_i^{AP}}{N} \tag{10}$$

其中: T_{TP} 为预测正确的正样本数; F_{FP} 为预测错误的正样本数; N 为类别总数。

3.4 实验结果分析

3.4.1 自适应感受野模块对识别精度的影响

为验证自适应感受野模块(ARFM)对网络识别精度的影响,在多背景鞋子数据集上,本文将对引入自适应感受野模块的ResNet50网络进行测试,实验结果如表1所示。

表1 分组卷积和标准卷积对识别精度的影响

Table 1 Influence of group convolution and standard convolution on recognition accuracy %		
网络	Rank-1	mAP
ResNet50	71.98	55.05
ResNet50+ARFM(标准卷积)	73.93	55.33
ResNet50+ARFM(分组卷积)	74.32	55.93

从表1可以看出,自适应感受野模块能够有效提高网络识别性能,本文采用标准卷积的精度相比于分组卷积的精度略微下降,在ResNet50网络中加入ARFM之后,相比ResNet50,ResNet50+ARFM(分组卷积)的Rank-1和mAP精度分别提高2.34和0.88个百分点,能够有效提高识别精度。ARFM在不降低目标分辨率的同时,使网络自适应选择合适大小的感受野特征进行学习,从而提高鞋型识别精度,实验结果充分验证自适应感受野模块的有效性。

分组卷积和标准卷积对网络复杂性的影响如表2所示。本文将ARFM模块中的分组卷积替换成标准卷积。从表2可以看出,相比ResNet50,当使用标准卷积时,自适应感受野模块参数量增大 8.66×10^6 ,且网络浮点运算量增加了 0.61×10^9 ;当使用分组卷积时,自适应感受野模块参数量仅增大了 2.14×10^6 ,浮点运算量增大 0.2×10^9 。实验结果表明,在自适应感受野模块中采用分组卷积能够大幅降低参数量和运算量,从而提升网络训练效率。

表2 分组卷积和标准卷积对网络复杂性的影响

Table 2 Influence of group convolution and standard convolution on network complexity		
网络	FLOPs/ 10^9	参数量/ 10^6
ResNet50	1.20	25.56
ResNet50 + ARFM(标准卷积)	1.81	34.42
ResNet50 + ARFM(分组卷积)	1.40	27.70

3.4.2 实例与批量标准化对识别精度的影响

在多背景鞋子数据集上,本文对加入实例与批量标准化IBN的ResNet50网络进行测试,自适应感受野模块与IBN整体对网络识别精度的影响如表3所示。

表3 IBN和自适应感受野模块对识别精度的影响

Table 3 Influence of instance and batch normalization and adaptive receptive field module on recognition accuracy %		
网络	Rank-1	mAP
ResNet50	71.98	55.05
ResNet50 + IBN	72.37	56.41
ResNet50 + ARFM	74.32	55.93
ResNet50 + IBN + ARFM	74.71	56.76

从表3可以看出,相比ResNet50网络,在ResNet50网络中加入IBN的Rank-1和mAP精度分别提高了0.39和1.36个百分点,在ResNet50网络中同时使用IBN和ARFM,Rank-1和mAP精度分别提高2.73和1.71个百分点。因此,IBN能够有效缩小目标差异,在ResNet50网络conv2_x中,每组卷积块的第一层卷积层之后使用IBN结构,能够提高网络的识别能力,并验证IBN结构的有效性。

3.4.3 多支路特征融合模型对识别精度的影响

本文算法是在IBN和ARFM的ResNet50网络基线上引入Branch 1+Branch 2+Branch 3融合特征进行识别。为验证多支路特征融合模型的有效性,本文按照相同的融合方式测试Branch 1+Branch 3、Branch 2+Branch 3、Branch 4+Branch 3、Branch 2+Branch 4+Branch 3、Branch 1+Branch 4+Branch 3、Branch 1+Branch 2+Branch 3这6种方法的识别精度,其中Branch 4是conv4_x引出的分支。多支路特征融合模型的精度对比如表4所示。多支路特征融合模型的Rank-1、Rank-5、Rank-10对比如图8所示。

表4 多支路特征融合模型的精度对比

Table 4 Accuracy comparison among multi-branch feature fusion models

融合模型	Rank-1	mAP
基线	74.71	56.76
基线+Branch 1+Branch 3	75.88	58.10
基线+Branch 2+Branch 3	75.88	57.42
基线+Branch 4+Branch 3	71.60	55.30
基线+Branch 1+Branch 4+Branch 3	73.93	57.55
基线+Branch 2+Branch 4+Branch 3	73.93	56.81
基线+Branch 1+Branch 2+Branch 3	76.65	58.11

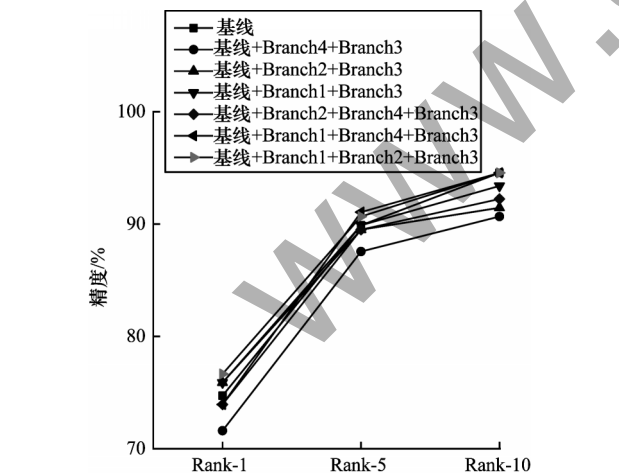


图8 多支路特征融合模型的Rank-1、Rank-5、Rank-10对比

Fig.8 Rank-1,Rank-5,Rank-10 comparison among multi-branch feature fusion models

从表4和图8可以看出,多支路特征融合模型对网络识别性能的提升具有显著效果,在加入IBN和

ARFM的ResNet50基线上,将Branch 1、Branch 2和Branch 3特征融合之后Rank-1精度相比基线提高了1.94个百分点,mAP精度提高1.35个百分点,识别精度具有显著提升。在卷积神经网络中,单独使用深层高级语义特征会丢失图像细节信息,从而影响鞋型识别性能,将神经网络中浅层纹理、边缘等细节特征与深层高级语义特征相融合,使得网络利用鲁棒性较优的鞋型特征进行鞋型识别,从而提升识别精度,进一步验证本文特征融合模型的有效性。

3.4.4 损失函数对识别精度的影响

在多背景鞋子数据集上,本文验证Center Loss函数和LS(Label Smoothing)函数对识别精度的影响,实验结果如表5所示。本文算法加入Center Loss函数和LS函数的Rank-1和mAP分别为78.21%和60.98%。在Softmax损失函数的基础上,本文联合Center Loss函数训练网络在增大类间距离的同时缩小类内差距,使样本实现更优的聚类效果;在Softmax损失函数中加入LS函数能够有效增强模型泛化性能,避免出现训练过拟合现象,提高鞋型识别精度。

表5 损失函数对识别精度的影响

Table 5 Influence of loss function on recognition accuracy

算法	Rank-1	mAP
本文算法	76.65	58.11
本文算法+Center Loss 函数	77.43	57.45
本文算法+Center Loss 函数+LS 函数	78.21	60.98

Re-ranking是图像检索领域常用的测试技巧,通过对检索结果重新排序,提升模型识别性能。本文在测试最终模型性能时加入Re-ranking,实验结果如表6所示。

表6 Re-ranking测试实验结果

Table 6 Experimental results of Re-ranking test

算法	Rank-1	mAP
本文算法+ Center Loss 函数+LS 函数	78.21	60.98
本文算法+ Center Loss 函数+LS 函数+re-ranking	79.77	62.18

从表6可以看出,本文算法具有较优的识别性能,引入Re-ranking的Rank-1和mAP精度分别达到79.77%和62.18%,相比ResNet50基础网络,其Rank-1和mAP精度分别提高7.79和7.13个百分点。

3.4.5 结果可视化

为更加直接展现ARFM、特征融合的有效性,本文在ResNet50、ARFM、特征融合模型上进行部分数据测试,并对Rank-5结果进行可视化,左侧图像是待查询图像,右侧5张图像是从库中返回的查询结果。其中带有√标志的代表正确的查询结果。不同算法的识别结果如图9所示。从图9可以看出,原始

ResNet50 网络识别效果较差,错误结果较多,但是在 ResNet50 网络基础上融合 ARFM 和多层特征后,其识别效果显著提升,同时验证了本文提出的自适应感受野模块和多层特征融合模型的有效性。

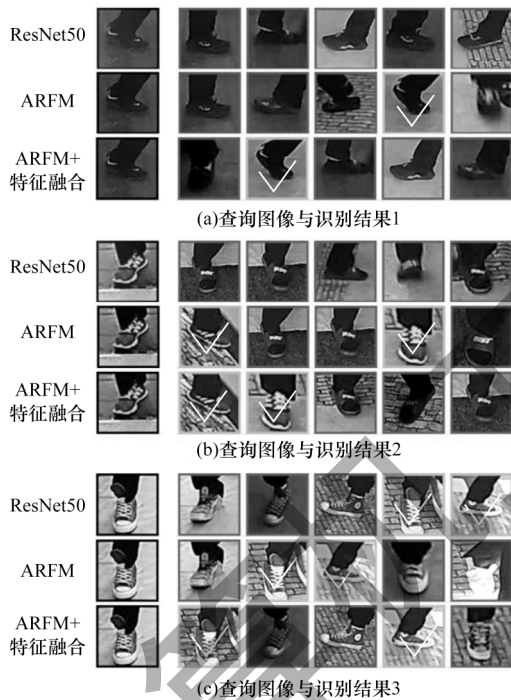


图9 不同算法的识别结果

Fig.9 Recognition results comparison among different algorithms

4 结束语

本文提出基于自适应感受野与多支路特征融合的鞋型识别算法。设计一种轻量级自适应感受野模块,实现自适应选择合适大小感受野特征,提升识别精度,同时融合神经网络浅层特征和深层特征,在 Softmax 损失函数中加入 Label Smoothing 并联合 Center Loss 函数对网络进行训练。实验结果表明,本文算法具有较高的识别精度和较强的实用性。在公安实战中,受监控摄像头分辨率、天气等因素的影响,鞋子影像存在极度模糊和色彩、亮度发生变化的情况。因此,下一步将对数据集进行研究,探究超分辨率重建、数据增强等方法对鞋型识别效果的影响,使算法适用于公安实战工作。

参考文献

- [1] 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.
- [2] WEN Y D, ZHANG K P, LI Z F, et al. A discriminative feature learning approach for deep face recognition[C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 499-515.
- [3] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 2818-2826.
- [4] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2021-05-10]. <https://arxiv.org/abs/1409.1556>.
- [7] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 1-9.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [9] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 2261-2269.
- [10] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [11] GHODRATI A, DIBA A L, PEDERSOLI M, et al. DeepProposal: hunting objects by cascading deep convolutional layers[C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 2578-2586.
- [12] WAN J, WANG D Y, HOI S C H, et al. Deep learning for content-based image retrieval: a comprehensive study [C]// Proceedings of the 22nd ACM International Conference on Multimedia. New York, USA: ACM Press, 2014: 157-166.
- [13] LI Y, KONG X W, ZHENG L, et al. Exploiting hierarchical activations of neural network for image retrieval [C]// Proceedings of the 24th ACM International Conference on Multimedia. New York, USA: ACM Press, 2016: 132-136.
- [14] 侯媛媛,何儒汉,李敏,等. 结合卷积神经网络多层特征融合和K-Means聚类的服装图像检索方法[J]. 计算机科学, 2019, 46(S1): 215-221.
- [15] HOU Y Y, HE R H, LI M, et al. Clothing image retrieval

- method combining convolutional neural network multi-layer feature fusion and K-Means clustering[J]. Computer Science, 2019, 46(S1): 215-221. (in Chinese)
- [15] GONG Y C, WANG L W, GUO R Q, et al. Multi-scale orderless pooling of deep convolutional activation features[C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 392-407.
- [16] YANDEX A B, LEMPITSKY V. Aggregating local deep features for image retrieval [C]// Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2015: 1269-1277.
- [17] KALANTIDIS Y, MELLINA C, OSINDERO S. Cross-dimensional weighting for aggregated deep convolutional features [C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 685-701.
- [18] TOLIAS G, SICRE R, JÉGOU H. Particular object retrieval with integral max-pooling of CNN activations[EB/OL]. [2020-05-15]. <https://arxiv.org/pdf/1511.05879.pdf>.
- [19] 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.
- [20] XU J, SHI C Z, QI C Z, et al. Unsupervised part-based weighting aggregation of deep convolutional features for image retrieval[EB/OL]. [2020-05-15]. <https://arxiv.org/pdf/1705.01247.pdf>.
- [21] RADENOVIC F, TOLIAS G, CHUM O. Fine-tuning CNN image retrieval with no human annotation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1655-1668.
- [22] 朱杰,赵相坤,谢博鋈,等. 基于深度特征加权的图像表示方法[J]. 郑州大学学报(理学版), 2020, 52(1): 47-53.
- ZHU J, ZHAO X K, XIE B J, et al. Deep feature weighting based image representation [J]. Journal of Zhengzhou University (Natural Science Edition), 2020, 52(1): 47-53. (in Chinese)
- [23] 李恒,赵广社,王鼎衡,等. 加权聚合深度卷积特征的图像检索方法[J]. 信息与控制, 2020, 49(1): 55-61.
- LI H, ZHAO G S, WANG D H, et al. An image retrieval method based on weighting aggregation of deep convolutional features[J]. Information and Control, 2020, 49(1): 55-61. (in Chinese)
- [24] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [25] CHEN W, LIU Y, WANG W P, et al. Deep learning for instance retrieval: a survey[EB/OL]. [2020-05-15]. <https://arxiv.org/abs/2101.11282>.
- [26] DING Z Y, SONG L, ZHANG X T, et al. Selective deep ensemble for instance retrieval[J]. Multimedia Tools and Applications, 2019, 78(5): 5751-5767.
- [27] 熊炜,杨荻椿,艾美慧,等. 基于多支路特征融合的行人重识别研究[J]. 计算机工程与科学, 2021, 43(2): 329-339.
- XIONG W, YANG D C, AI M H, et al. Person re-identification based on multi-branch feature fusion[J]. Computer Engineering & Science, 2021, 43(2): 329-339. (in Chinese)
- [28] AI J Q, TIAN R T, LUO Q W, et al. Multi-scale rotation-invariant haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12): 10070-10087.
- [29] SHI Q Q, LI W, ZHANG F, et al. Deep CNN with multi-scale rotation invariance features for ship classification[J]. IEEE Access, 2018, 6: 38656-38668.
- [30] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 636-644.
- [31] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 5987-5995.
- [32] 张家钧,唐云祁,杨智雄,等. 基于注意力机制的鞋型识别算法[J]. 激光与光电子学进展, 2022, 59(2): 365-373.
- ZHANG J J, TANG Y Q, YANG Z X, et al. Shoe type recognition algorithm based on attention mechanism[J]. Laser & Optoelectronics Progress, 2022, 59(2): 365-373. (in Chinese)
- [33] 刘航,汪西莉. 自适应感受野机制遥感图像分割模型[J]. 中国图象图形学报, 2021, 26(2): 464-474.
- LIU H, WANG X L. Remote sensing image segmentation model based on an adaptive receptive field mechanism[J]. Journal of Image and Graphics, 2021, 26(2): 464-474. (in Chinese)
- [34] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]// Proceedings of the 32nd International Conference on International Conference on Machine Learning. New York, USA: ACM Press, 2015: 448-456.
- [35] PAN X G, LUO P, SHI J P, et al. Two at once: enhancing learning and generalization capacities via IBN-Net [C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 484-500.