

DOI:10.19651/j.cnki.emt.2107731

# 基于改进残差网络和数据增强的鞋型识别算法\*

张家钧 唐云祁 杨智雄

(中国人民公安大学 侦查学院 北京 100038)

**摘要:** 在监控视频中搜索犯罪现场出现的嫌疑鞋型是目前侦破案件的重要手段,针对该手段自动化程度低、人工筛选易漏查等问题,提出一种基于改进深度残差网络和数据增强的鞋型识别算法。为增强网络特征提取能力,对深度残差网络进行了研究,在不增加任何参数数量的前提下改进瓶颈结构,提升算法精度;针对瓶颈结构中下采样操作存在的问题,改进下采样模块,缓解网络下采样时信息丢失问题;引入 Mixup 和光学变换数据增强算法,建立数据之间的线性关系,丰富数据的多样性,进而增强网络模型的鲁棒性;最后,采取中心损失函数和 Softmax 损失函数联合训练的方法,使训练数据达到更好的聚类效果。为验证所提算法的有效性和实用性,在多背景鞋型数据集上对所提算法进行测试,测试结果表明,所提算法 mAP、Rank-1 精度分别达到 66.83%、86.77%,可以有效提高网络识别精度。

**关键词:** 鞋型识别;残差网络;瓶颈结构;Mixup 算法;中心损失函数

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.2060

## Shoe type recognition algorithm based on improved residual network and data augmentation

Zhang Jiajun Tang Yunqi Yang Zhixiong

(School of Investigation, People's Public Security University of China, Beijing 100038, China)

**Abstract:** It is an important means to search the suspected shoes in the surveillance video at the scene of crime. Aiming at the problems of low automation and manual screening, this paper proposes a shoe recognition algorithm based on improved deep residual network and data augmentation. In order to enhance the ability of network feature extraction, the deep residual network is studied. The bottleneck structure is improved without adding any parameters to improve the accuracy of the algorithm. Aiming at the problem of down sampling operation in bottleneck structure, the down sampling module is improved to alleviate the problem of information loss in network down sampling. Mixup and optical transform data augmentation algorithm are introduced to establish the linear relationship between data, enrich the diversity of data, and enhance the robustness of network model. Finally, the combined training method of center loss function and Softmax loss function is adopted to make the training data achieve better clustering effect. In order to verify the effectiveness and practicability of the proposed algorithm, the proposed algorithm is tested on multi background shoe data sets. The test results show that the accuracy of mAP and Rank-1 of the proposed algorithm is 66.83% and 86.77% respectively, which can effectively improve the accuracy of network recognition.

**Keywords:** shoe type recognition; residual network; bottleneck structure; Mixup algorithm; center loss function

## 0 引言

近年来,根据犯罪现场遗留鞋印查询对应鞋型,进而到视频监控中搜索相同鞋型确定犯罪嫌疑人是公安机关侦破的重要手段。2015年,某市发生两起技术开锁入室盗窃案,在犯罪现场除了足迹办案民警没有发现其他任何痕

迹物证。由于现场物证较少,办案民警从现场足迹着手,根据现场遗留鞋印到全国公安机关鞋样本查询系统中查询得知,现场足迹可能是由鸿星尔克和耐克两种品牌的运动鞋所遗留。根据这一线索,办案民警结合案发时间和鞋子的品牌信息,调取案发现场周围监控视频信息,经过细心排查,发现穿着这两种品牌运动鞋的犯罪嫌疑人,办案民警迅

收稿日期:2021-08-30

\* 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项目(2021XCWZK04)资助

速锁定嫌疑人,打破办案僵局<sup>[1]</sup>。该侦查破案手段虽具有较强的实战性,但是在监控视频中搜索鞋型这一工作完全依赖民警人工完成,不仅会消耗大量人力物力,且会因主观原因造成漏检。因此,研究监控视频中自动识别鞋型算法对公安机关侦破案件具有重要意义<sup>[2]</sup>。

随着深度学习的快速发展,特别是 AlexNet<sup>[3]</sup>、VGGNet<sup>[4]</sup>、GoogLeNet<sup>[5]</sup>、ResNet<sup>[6]</sup>、DenseNet<sup>[7]</sup>等深度卷积神经网络的出现,深度学习在图像分类、目标检测<sup>[8]</sup>、图像检索<sup>[9]</sup>等领域取得了巨大成就。自从2015年 ResNet<sup>[6]</sup>网络的出现,广大研究者将目光聚焦在了改进 ResNet 网络结构工作上。文献[10]对 ResNet 的残差结构进行了分析研究,发现残差结构中的 ReLU 层和 BN 层的位置对网络性能的影响较大,因此文章在不增加参数量的前提下只通过改变 ReLU 层和 BN 层的位置提出了一种预激活的残差结构,成功搭建了 1 001 层的 ResNet,研究得到了更易于训练且精度更高的 ResNetV2 网络。文献[11]在 ResNet 的基础上使用了分组卷积,在不增加模型参数量和运算量的前提下提出了精度更高、效果更好的 ResNext 网络。文献[12]在 ResNet 网络的基础上通过引入注意力机制进而提出 SENet 网络模型,使网络可以自适应选择提取更重要的特征进行学习。在残差结构中引入注意力机制的网络模型除 SENet 外,还有自适应调节合适大小卷积核的 SKNet<sup>[13]</sup>、分组注意力机制的 ResNeSt<sup>[14]</sup>、自适应选择一维卷积核大小的 ECANet<sup>[15]</sup>等。文献[16]在残差结构中构造分层的残差连接取代了单个  $3 \times 3$  的卷积核,增大了网络感受野,提出了精度更高的 Res2Net 网络。文献[17]提出的 RegNet,通过综合手工设计网络和神经架构搜索的优点,设计得到了参数化网络群的网络设计空间,在推理速度远高于 EfficientNet 同时,精度也超过了 EfficientNet。文献[18]在 ResNet 的基础上,提出了分段的残差结构,在 Start ResBlock、Middle ResBlock、End ResBlock 分别采用不同的 BN 和 ReLU 组合结构,并且在残差结构中的  $3 \times 3$  卷积层中引入更多的通道信息,提出了信息传播更强的 IResNet 网络结构。文献[19]指出,随着卷积神经网络层数的加深,梯度在进行反向传播时,并不能保证每一残差模块都能学习到有效特征表达,因此文献[19]从减少网络层数加深网络宽度的角度出发,设计了一种层数较浅宽度更宽的 WRN 网络,在公开数据集上取得了更好的效果。文献[20]针对 ResNet 残差结构存在冗余输出的问题,设计了新的升维结构,增强了网络的梯度传递,经过大量实验表明,新的升维结构在类别较少的分类任务中有较好的效果。在深度学习中,数据增强也一直是研究的热点。丰富的数据是解决计算机视觉问题的重要保障,但是搜集和标注数据却是一个极其费力耗时的过程,因此,越来越多的人对数据增强算法展开研究。基本的图像增强算法主要是图像变换,例如翻转、平移、缩放、模糊、光学变换、噪声等<sup>[21-24]</sup>,但是随着深度学习的快速发展,一些更具针对性的图像增强

算法也层出不穷,例如随机遮挡<sup>[25]</sup>、样本混合<sup>[26]</sup>、虚拟图像生成<sup>[27]</sup>等算法。

虽然针对 ResNet 的研究较多,但是深度残差网络在下采样时存在信息丢失问题,且原始 ResNet 瓶颈结构中的 ReLU 层和 BN 层的位置尚存在一定的不足之处<sup>[28]</sup>。针对上述问题,本文对深度残差网络瓶颈结构和下采样模块进行了改进,提出一种基于改进深度残差网络和数据增强的鞋型识别算法。首先针对原始 ResNet50 网络瓶颈结构 ReLU、BN 层的位置问题,提出两种重组方案,分别用于浅层(conv2\_x、conv3\_x)和深层(conv4\_x、conv5\_x)瓶颈结构中,进而加强特征信息在网络中的有效传递;之后对瓶颈结构中主干和残差支路的下采样操作进行改进,进而缓解原始网络下采样时信息丢失的问题;同时为增强数据样本间的线性关系和丰富数据的多样性,引入 Mixup 和光学变换数据增强算法,增强模型的鲁棒性;最后,为使训练数据达到更好的聚类效果,在增大类间间距的同时缩小类内差距,本文使用中心损失函数(Center Loss)和 Softmax 损失函数联合训练的方法,进一步提升模型识别性能。通过对模型的改进,使得网络瓶颈结构特征表达能力大大加强,帮助网络充分利用所学特征进一步提升鞋型识别精度。

## 1 本文方法

### 1.1 ResNet50 瓶颈结构的改进

图 1 为原始 ResNet50 网络的瓶颈结构。每个瓶颈结构可以由式(1)、(2)表示。

$$y_i = g(x_i) + F(x_i, W_i) \quad (1)$$

$$x_{i+1} = f(y_i) \quad (2)$$

其中,  $x_i$  和  $x_{i+1}$  分别代表第  $i$  个瓶颈结构的输入和输出,  $F$  代表残差函数,  $W_i$  代表权值矩阵,  $f$  代表 ReLU 激活函数,函数  $g$  为:

$$g(x_i) = x_i \quad (3)$$

由图 1 可以看出残差支路由两组 conv—BN—ReLU 层和一组 conv—BN 层组成,在瓶颈结构末端经过 ReLU 激活函数得到输出。其中 BN 层是对神经网络某一层输入信号的归一化处理,使其输出结果满足均值为 0 方差为 1,稳定数据的分布,降低深层网络训练难度。ReLU 激活函数的输出范围为  $(0, x)$ ,信号经过 ReLU 激活函数,所有的负值变为 0,正值不变,使神经元具有了稀疏激活性的特点,且在  $x > 0$  时其梯度一直为 1,缓解了梯度弥散的问题。ResNet 网络中 conv—BN—ReLU 层是目前较为常见的组合,但是其在信息传播过程中存在一定的缺陷。原始残差支路中的层顺序为:conv1—BN1—ReLU—conv2—BN2—ReLU—conv3—BN3,其中信号经过 ReLU 激活函数之后输入到卷积层中,由 ReLU 激活函数的公式可以得知其输出结果范围为  $(0, x)$ ,每层卷积层的输入信号并没有经过 BN 层做归一化处理,其数据不再满足 0 均值和单位方差的分布,且随着层数的加深,网络中传递的信号波动会逐渐增

大,造成网络难以训练的问题;另外,在瓶颈结构末端,信号经过 ReLU 激活函数,其输出结果永远非负,前向传播时输入信号会递增,从而影响网络特征表达能力,在训练前期,对网络提取特征能力影响较大。

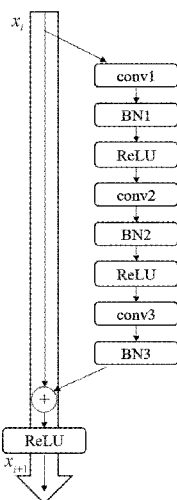


图 1 原始 ResNet50 瓶颈结构

针对上述问题,本文提出两种重组方案,如图 2(a)、(b)所示,分别用于浅层(conv2\_x、conv3\_x)和深层(conv4\_x、conv5\_x)瓶颈结构中。在新提出的重组方案中,针对 BN 层没有有效发挥作用问题,本文在浅层瓶颈结构的残差支路中改变了 conv、BN、ReLU 层的位置顺序,并将瓶颈结构末端的 ReLU 激活函数转移到残差支路中,提出 conv1—ReLU—BN1—conv2—ReLU—BN2—conv3—ReLU—BN3 的组合。在此组合中,BN 层位于 ReLU 激活函数之后,BN 层会对 ReLU 层的输出进行归一化处理,对其输出形成限制作用,并使数据满足 0 均值和单位方差分布,易于网络训练;在浅层中将瓶颈结构末端的 ReLU 激活函数转移到残差支路中,使瓶颈结构的输出范围从  $(0, x)$  变成  $(z, x)$ ,  $z$  为小于 0 的数,消除 ReLU 层对负权值清零对传播信息造成的影响,进而增强瓶颈结构的特征表达能力。同样,在深层(conv4\_x、conv5\_x)瓶颈结构的残差支路中本文做出了类似改进,在残差支路中提出 conv1—ReLU—BN1—conv2—ReLU—BN2—conv3 的组合,在瓶颈结构的末端经过 ReLU、BN 层得到输出,之前在 conv2\_x、conv3\_x 瓶颈结构中,为了增强网络的特征表达能力,将瓶颈结构末端的 ReLU 激活函数层转移到了残差支路中,消除了对负权值清零对信息传播造成的影响,但是随着网络层数的增加,由于瓶颈结构末端没有 BN 层,没有对输出信号进行归一化处理,增加了网络训练的困难,因此,在深层瓶颈结构中我们将残差支路末端的 ReLU—BN3 组合层转移至瓶颈结构末端,降低网络训练难度。

1.2 ResNet50 瓶颈结构中下采样的改进

在原始 ResNet50 瓶颈结构中,主干和残差支路的下采样操作均是在卷积核大小为  $1 \times 1$ 、步长为 2 的卷积层中进

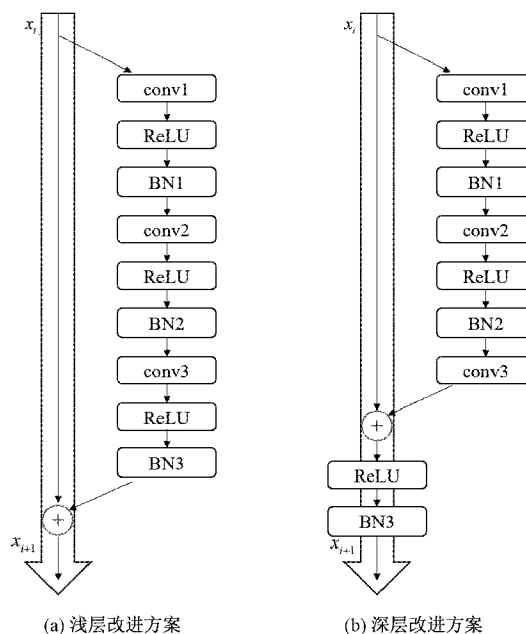


图 2 ResNet50 瓶颈结构重组方案

行的,其下采样操作如图 3 所示。使用步长为 2 的  $1 \times 1$  卷积核进行下采样操作会丢失输入特征图的大量重要信息,对网络中信息的传递造成较大的影响。因此,针对上述问题,本文在瓶颈结构主干上的下采样模块中引入  $3 \times 3$ 、步距为 2 的平均池化层,在得到相同尺寸大小特征图的前提下保证了传播信息的完整性,在平均池化之后连接卷积核大小为  $1 \times 1$ 、步距为 1 的卷积层和 BN 层,进一步挑选有效特征信息;在残差支路中,将原本在卷积核大小为  $1 \times 1$  的卷积层中进行的下采样操作转移至卷积核大小为  $3 \times 3$  的卷积层中进行,可以有效避免在残差支路中进行下采样操作会丢失输入特征图的大量重要信息问题,改进后的下采样模块如图 4 所示。

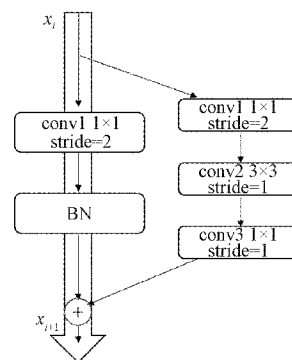


图 3 原始 ResNet50 下采样模块

1.3 Mixup 和光学变换数据增强算法

1) Mixup 数据增强算法

丰富的数据是深度学习解决计算机视觉问题的根本,数据增强技术是深度学习中常用的数据处理方法,可以提高数据的多样性,进而增强训练模型的性能。Mixup<sup>[29]</sup>数

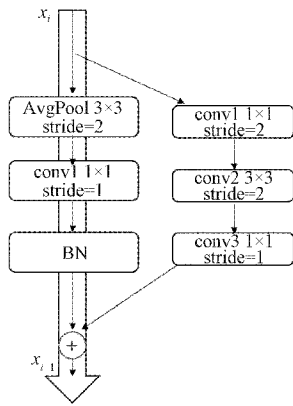


图 4 改进下采样模块

据增强算法不同于传统的数据增强算法,其通过逐像素线性相加将两类图像混合,并将数据标签单热向量编码通过不同的权重相加进而得到多热向量编码,相当于起到了标签平滑的作用,可以进一步增强训练模型的鲁棒性。Mixup 算法原理如式(4)~(6)所示。

$$\tilde{x} = \lambda x_i + (1 - \lambda x_j) \tag{4}$$

$$\tilde{y} = \lambda y_i + (1 - \lambda y_j) \tag{5}$$

$$\lambda = \text{Beta}(\alpha, \alpha), \alpha \in (0, \infty) \tag{6}$$

其中,  $x_i$  和  $x_j$  是随机选择的两个图像样本,  $y_i$  和  $y_j$  是样本  $x_i$  和  $x_j$  对应的独热标签,  $\tilde{x}$  是经 Mixup 算法混合后的图像样本,  $\tilde{y}$  是  $\tilde{x}$  对应的标签。Mixup 算法图像生成结果如图 5 所示,图 5 左侧是两类训练样本,右侧是通过不同的权重将两类样本逐像素线性相加得到的不同效果展示图。

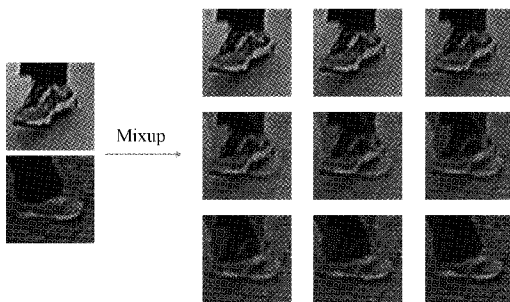


图 5 Mixup 数据增强

2) 光学变换数据增强算法

在公安实战中,受到天气和监控摄像头等设备的影响,鞋子的明暗程度和色彩会呈现出不同的效果,因此本文采取光学变换数据增强方法<sup>[30]</sup>进一步使数据贴近公安实战。实验选取饱和度(saturation)、对比度(contrast)、亮度(brightness)3种光学变换数据增强方法,通过给定上述3种光学变换方法一定的变换区间范围,随机选取区间内的变换值进行光学变换得到最终数据,算法图像生成结果如图 6 所示。图 6 分别展示了 brightness、contrast、saturation 变换和最终的增强数据效果图。

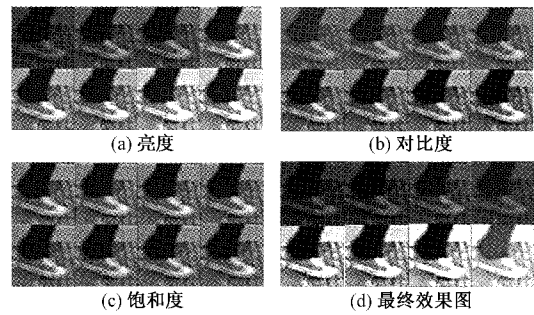


图 6 光学变化数据增强

1.4 损失函数

Softmax 损失函数在分类任务中具有较好的效果,但是其只强调了类间分离而没有强调类内紧凑。本文采用的数据集均为鞋类,类间相似度较高,因此采用中心损失函数<sup>[31]</sup>和 Softmax 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,使数据达到更好的聚类效果,中心损失函数公式如式(7)所示。

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \tag{7}$$

其中,  $m$  表示批次大小,  $x_i$  表示 FC 层之前的特征,  $c_{y_i}$  表示第  $y_i$  个类别的特征中心。最终,实验采用的损失函数  $L$  如式(8)所示。

$$L = L_s + \lambda L_c \tag{8}$$

其中,  $L_s$  代表 Softmax 损失函数,  $L_c$  代表 Center Loss 函数,  $\lambda$  为  $L_c$  的权重,本文将其设置为 0.000 1。

2 实验及结果分析

2.1 实验设置

本文实验均是在 Linux 3.10.0 操作系统下、采用 PyTorch 深度学习框架进行的。处理器为 Intel(R) Xeon (R) CPU E5-2650 v4 2.20 GHz, GPU 型号为 NVIDIA TITAN X (Pascal), 显存为 12 GHz, PyTorch 版本为 1.2.0, 编译环境为 Python3.5.6。数据输入尺寸均为  $120 \times 120$ , 批处理大小为 64, 优化器为 Adam, 初始学习率为 0.000 1, 采用迁移学习方法训练网络模型。

2.2 数据集

本实验所采用的多背景鞋型数据集<sup>L32</sup>是在中国人民公安大学足迹实验室由 3 个不同角度的监控摄像头采集得到的,此数据集包含 300 类鞋共 35 600 张低分辨率鞋子图像。其中训练集包含 150 类鞋共 30 000 张鞋图;测试集包含 150 类鞋 300 张鞋图;另外挑选 5 000 张混淆鞋样本和测试集共同构成鞋型数据库。多背景鞋型数据集部分数据样式如图 7 所示。

2.3 评价指标

本文选取 Rank-1 和 mAP 精度作为评估算法性能的评价指标。Rank-1 表示模型算法返回的鞋子排序中,第 1 位正确匹配的概率; mAP 反映了所提算法整体性能的好



图7 数据集部分样式图

坏,其可用式(9)~(11)表示。

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$AP = \frac{\sum Precision}{images} \tag{10}$$

$$mAP = \frac{\sum_{i=0}^N AP_i}{N} \tag{11}$$

其中,  $N$  是数据的类别数,  $FP$  代表预测错误的正样本数,  $TP$  代表预测正确的正样本数。

### 2.4 实验及结果分析

#### 1)改进瓶颈结构对模型性能的影响

为验证改进瓶颈结构对模型识别精度的影响和各个改进部分的有效性,本文选用 ResNet50 网络结构在多背景鞋型数据集上进行了消融实验,各组实验的 mAP 和 Rank-1 精度如表 1 所示,其中,avg 代表在瓶颈结构主干上的下采样模块中引入  $3 \times 3$ 、步距为 2 的平均池化层, recombination 代表本文在瓶颈结构残差支路上提出的两种重组方案, stride 代表将原本在卷积核大小为  $1 \times 1$  的卷积层中进行的下采样操作转移至卷积核大小为  $3 \times 3$  的卷积层中进行, e\_ResNet50(effective ResNet50)代表本文提出的有效的 ResNet50 网络。

表1 消融实验结果

方法	avg	recombination	stride	Rank-1/%	mAP/%
ResNet50				71.21	52.23
	✓			72.76	54.49
		✓		75.10	54.78
			✓	74.32	56.77
	✓	✓		75.88	55.91
	✓		✓	74.71	56.99
		✓	✓	75.10	56.97
e_ResNet50	✓	✓	✓	<b>77.04</b>	<b>57.04</b>

由表 1 可以看出,本文对 ResNet50 网络各部分的改进在识别精度上都具有较好的提升效果,与原始 ResNet50 网络相比,本文提出的 e\_ResNet50 网络在 Rank-1 和 mAP 上分别提升了 5.83% 和 4.81%。这是因为通过重组瓶颈结构中的卷积、BN 和 ReLU 层,加强了特征信息在网络中的有效传递,进而增强了瓶颈结构的特征表达能力,且降低了网络训练难度;通过改进瓶颈结构中下采样结构,有效缓解了残差支路中下采样操作丢失输入特征图的大量重要信息问题,使网络能够充分利用学习得到的特征。

同时,本文对提出的 e\_ResNet50 网络的复杂性与原始 ResNet50 网络进行了比较,结果如表 2 所示。由表 2 可以看出本文提出的 e\_ResNet50 在参数量上没有任何增加,仅浮点运算量增加了 0.07 G,实验结果表明,本文提出的 e\_ResNet50 具有较强的实用性。

表2 模型复杂性比对

方法	FLOPs/G	Params/M
ResNet50	1.20	25.56
e_ResNet50	1.27	25.56

#### 2)验证改进瓶颈结构的有效性实验

为了验证本文改进的瓶颈结构的有效性,选取了 SENet、ResNext、ResNet101、Res2Net 在多背景鞋型数据集上进行了测试,实验按照本文提出的改进方案对上述网络进行改进测试,进而验证本文提出的改进方案的有效性,(其中由于 ResNet101 网络层数过深,实验将图 2(a)的结构应用于 ResNet101 中的 conv2\_x、conv3\_x 和 conv4\_x 中,将图 2(b)的结构用于 conv5\_x 中),实验结果如表 3 所示。实验结果表明,在 ResNet 各变体网络中,本文提出的改进方案均能够加强特征信息在网络中的传递,降低网络训练难度,可以有效缓解残差支路中下采样操作丢失特征信息问题,结果证明了本文改进的瓶颈结构具有良好的识别效果且具有一定的泛化性能。

表3 有效性实验

方法	Rank-1/%	mAP/%
ResNet50	71.21	52.23
e_ResNet50	<b>77.04</b>	<b>57.04</b>
SENet	72.76	55.05
e_SENet	<b>76.65</b>	<b>55.36</b>
ResNext	70.82	53.36
e_ResNext	<b>78.21</b>	<b>57.22</b>
ResNet101	71.60	55.09
e_ResNet101	<b>75.49</b>	<b>57.47</b>
Res2Net	71.98	56.89
e_Res2Net	<b>78.21</b>	<b>58.39</b>

3) Mixup 和光学变换数据增强算法对模型性能的影响  
 为验证 Mixup 数据增强算法对模型性能的影响, 本文使用 e\_ResNet50 网络在多背景鞋型数据集上进行了测试, 探究了超参数  $\alpha$  对识别精度的影响, 实验结果如表 4 所示。由表 4 可以看出, 当超参数  $\alpha=0.4$  时, 网络识别精度最高, Rank-1 和 mAP 精度分别为 80.93% 和 60.77%。实验结果表明, Mixup 数据增强算法通过逐像素线性相加将两类图像混合, 将数据标签单热向量编码通过不同的权重相加进而得到多热向量编码, 建立了样本之间的线性关系, 可以有效增强网络模型的鲁棒性, 进而提升识别精度。

表 4 Mixup 数据增强算法对模型性能的影响

$\alpha$	Rank-1/%	mAP/%
e_ResNet50	77.04	57.04
加入 Mixup, $\alpha=0.2$	78.99	58.23
加入 Mixup, $\alpha=0.4$	<b>80.93</b>	<b>60.77</b>
加入 Mixup, $\alpha=0.6$	79.38	59.50
加入 Mixup, $\alpha=0.8$	79.38	60.15
加入 Mixup, $\alpha=1.0$	79.77	59.57

为进一步突出 Mixup 数据增强算法的有效性, 实验选取了另外两种混合数据增强方法: Cutmix<sup>[33]</sup> 和 Cutout<sup>[34]</sup> 算法进行实验比对。Cutmix 数据增强算法是对 A 样本进行随机裁剪, 将裁剪部分填充到 B 样本对应位置中, 同时将数据标签单热向量编码通过不同的权重相加进而得到多热向量编码; Cutout 数据增强算法是在样本中随机选取一个固定大小的正方形区域, 将其全 0 填充, Mixup、Cutmix、Cutout 数据增强算法效果如图 8 所示。实验将 Cutmix 中的超参数  $\beta$  设置为 1, 使其加权系数(0,1)符合 beta 分布; 将 Cutout 中的裁剪区域分别设置为 30、35、40、45 四个数值进行实验比对, 实验结果如表 5 所示。实验结果表明, Cutmix 和 Cutout 数据增强算法均没有 Mixup 数据增强算法提升效果明显, 这可能因为 Cutmix 数据增强算法直接将 A 样本随机裁剪的部分填充到 B 样本对应位置中, 裁剪部分可能是背景等干扰部分, 因此识别效果相较于 Mixup 数据增强算法稍差; 而 Cutout 数据增强算法直接在样本中随机选取一个区域将其全 0 填充, 可能会将鞋子重要特征区域遮挡掉, 因此两者对识别精度的提升效果均没有 Mixup 明显, 实验结果进一步证明了 Mixup 数据增强算法的有效性。

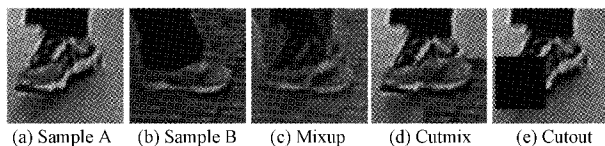


图 8 3 种数据增强方法

为验证光学变换数据增强算法对模型性能的影响, 本文使用 e\_ResNet50 网络在多背景鞋型数据集上进行了测

表 5 3 种数据增强算法比较

方法	Rank-1/%	mAP/%
Mixup	<b>80.93</b>	<b>60.77</b>
Cutmix	78.60	59.42
Cutout(size=30)	76.26	56.58
Cutout(size=35)	77.04	55.58
Cutout(size=40)	77.82	55.84
Cutout(size=45)	77.82	55.96

试, 实验探究了不同的变换区间范围对识别精度的影响, 实验结果如表 6 所示。由表 6 可以看出, 当变换区间范围为(0.6,1.4)时具有较好的识别精度, Rank-1 和 mAP 精度分别为 79.38% 和 60.27%。实验结果表明, 光学变换数据增强算法能够使多背景鞋子数据的明暗程度和色彩会呈现出不同的效果, 使数据更加贴近公安实战。

表 6 光学变换数据增强实验结果

变换区间范围	Rank-1/%	mAP/%
(0.5,1.5)	78.60	60.79
(0.6,1.4)	<b>79.38</b>	60.27
(0.7,1.3)	77.82	60.00

在随机裁剪、水平翻转数据增强算法的基础上, 实验最终另外选取 Mixup 和光学变换两种数据增强算法, 其识别精度如表 7 所示。实验结果表明, 加入 Mixup 和光学变换数据增强算法之后, 识别精度具有明显提升, Rank-1 和 mAP 精度分别为 84.05% 和 64.41%。

表 7 数据增强算法最终实验结果

方法	Rank-1/%	mAP/%
e_ResNet50	77.04	57.04
加入 Mixup	80.93	60.77
加入 optical transformation	79.38	60.27
加入 Mixup 和 optical transformation	<b>84.05</b>	<b>64.41</b>

#### 4) 损失函数对模型性能的影响

为验证 Center Loss 函数对网络识别精度的影响, 本文使用加入数据增强算法之后的 e\_ResNet50 网络在多背景鞋型数据集上进行了测试, 实验结果如表 8 所示, 加入 Center Loss 函数之后, Rank-1 和 mAP 精度分别提升了 2.33% 和 1.63%, 实验结果表明, Center Loss 函数能够使样本实现更好的聚类效果, 在 Softmax 损失函数的基础上联合 Center Loss 函数训练网络, 能够实现增大类间距离的同时缩小类内间距, 进一步增强模型的泛化性能, 提高识别精度。

表 8 损失函数对识别精度的影响

方法	Rank-1/%	mAP/%
本文模型	84.05	64.41
本文模型加入 Center Loss	<b>86.38</b>	<b>65.77</b>

Re-ranking 通过对识别结果进行重新排序,能够进一步提升网络识别精度,实验在测试模型最终性能时加入 Re-ranking 算法,对识别结果重新排序,实验结果如表 9 所示。结果表明,本文提出的最终算法 Rank-1 和 mAP 精度分别达到了 86.77%和 66.83%。

表 9 Re-ranking 实验结果

方法	Rank-1/%	mAP/%
本文方法	86.38	65.77
本文方法加入 Rc-ranking	<b>86.77</b>	<b>66.83</b>

### 5)关注区域可视化

为验证本文提出算法的有效性,本文使用原始 ResNet50 网络、e\_ResNet50 网络和加入数据增强之后的 e\_ResNet50 网络进行特征提取,将最后一层卷积层关注的特征区域进行可视化<sup>[35]</sup>,结果如图 9 所示。从图 9 中可以看出,原始 ResNet50 网络关注的特征区域范围不够准确,没有学习到鞋子的一些重要特征信息,并且网络关注了部分干扰背景信息,而本文提出的 e\_ResNet50 网络和加入数据增强之后的 e\_ResNet50 网络则能够精准定位有效特征区域,能够学习到更加准确的特征信息,并且能够排除复杂背景的干扰,更加直观地证明了本文所提算法的有效性。

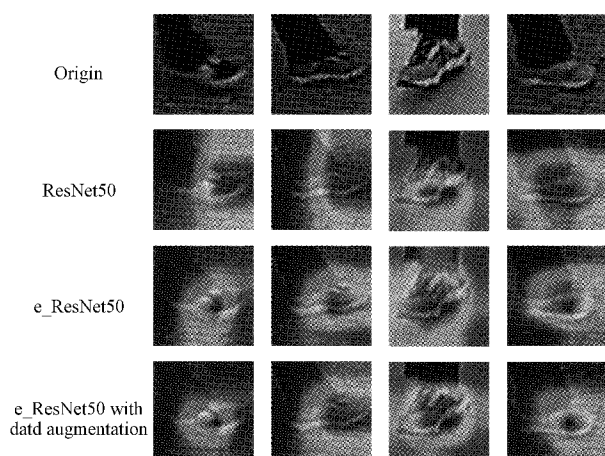


图 9 可视化效果图

### 6)识别结果可视化

为更加直接展示本文提出算法的有效性,本文将原始 ResNet50 网络、e\_ResNet50 网络和加入数据增强之后的 e\_ResNet50 网络的 Rank-5 识别结果进行可视化,可视化结果如图 10 所示,左侧图像是待识别图像,右侧 5 张图像代表查询返回结果,其中带有√标志的代表正确的查询结果。

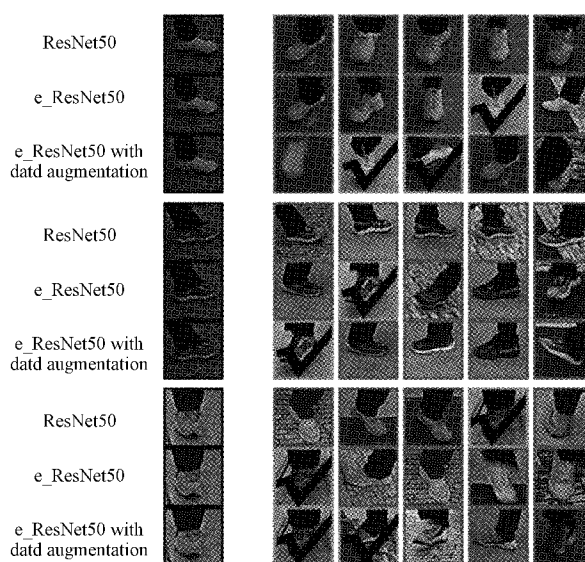


图 10 查询结果可视化

由图 10 可以看出,原始 ResNet50 网络识别效果较差,Rank-5 查询结果里不能准确查询到正确结果或者正确结果排序较靠后。相比于原始 ResNet50 网络,本文提出的 e\_ResNet50 网络和加入数据增强之后的 e\_ResNet50 网络则具有较好的识别效果,在 Rank-5 查询结果里能够较为准确的查询到正确结果,图 10 识别结果可视化进一步证明了本文所提算法的有效性。

## 3 结 论

将计算机视觉技术应用到公安实战中不仅可以加快侦破案件速度,而且可以避免民警受到主观因素的影响。本文对监控视频中鞋型识别问题展开研究,将图像识别技术运用到公安实战中,提出了一种基于改进深度残差网络和数据增强的鞋型识别算法。针对 ResNet50 网络瓶颈结构中存在的缺陷,提出两种新的重组方案,分别用于浅层和深层瓶颈结构中,在没有增加任何参数量的前提下进一步加强网络特征表达能力,提高识别精度;为避免网络在下采样步骤丢失过多有效特征信息,本文对下采样模块进行了改进,有效避免了下采样丢失过多信息问题;为增强网络模型的鲁棒性,建立数据之间的线性关系,丰富数据的多样性,本文引入 Mixup 和光学变换数据增强算法,有效提升了网络模型性能;同时为了使数据样本实现更好的聚类效果,将中心损失函数和 Softmax 函数联合训练,进一步提升网络识别精度。在多背景鞋型数据集上测试所提算法,Rank-1 和 mAP 精度可以达到 86.77%和 66.83%,为解决公安实战问题提供了一种可行方案。

但是针对监控视频中行人所穿鞋型的自动识别工作尚处于起步阶段,在实际案件中,鞋子影像会受分辨率和运动等影响而出现极度模糊的情况,因此后续工作将针对数据集开展相关研究,探究超分辨率重建等方法对识别精度的

影响,进一步使算法模型更加贴近公安实战工作。

### 参考文献

- [1] 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] 许磊, 晏于文, 黎智辉, 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] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]. NIPS, Curran Associates Inc, 2012: 1097-1105.
- [4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint, 2014, Arxiv:1409.1556.
- [5] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. Processing of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [6] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [7] HUANG G, LIU Z, DER MAATEN L V, et al. Densely connected convolutional networks [C]. Processing of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] GHODRATI A, DIBA A, PEDERSOLI M, et al. DeepProposals: Hunting objects by cascading deep convolutional layers [J]. International Journal of Computer Vision, 2017, 124(2): 115-131.
- [10] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]. European Conference on Computer Vision, Springer, Cham, 2016: 630-645.
- [11] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1492-1500.
- [12] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 99: 1-1.
- [13] LI X, WANG W, HU X, et al. Selective kernel networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.
- [14] ZHANG H, WU C, ZHANG Z, et al. ResNeSt: Split-attention networks [EB/OL]. (2020-04-19) [2021-5-24]. <https://arxiv.org/pdf/2004.08955.pdf>.
- [15] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020: 11534-11542.
- [16] GAO S, CHENG M M, ZHAO K, et al. Res2net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019(99): 1.
- [17] RADOSAVOVIC I, KOSARAJU R P, GIRSHICK R, et al. Designing network design spaces [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10428-10436.
- [18] DUTA I C, LIU L, ZHU F, et al. Improved residual networks for image and video recognition [C]. 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021: 9415-9422.
- [19] ZAGORUYKO S, KOMODAKIS N. Wide residual networks [J]. Computer Vision and Pattern Recognition, 2016, DOI: 10.5244/C.30.87.
- [20] 储岳中, 汪佳庆, 张学锋, 等. 基于改进深度残差网络的图像分类算法[J]. 电子科技大学学报, 2021, 50(2): 243-248.
- [21] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: Learning augmentation strategies from data [C]. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019, Piscataway: IEEE, 2019: 113-123.
- [22] YU X, WU X, LUO C, et al. Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework[J]. GIScience & Remote Sensing, 2017, 54(5): 741-758.
- [23] HUSSAIN Z, GIMENEZ F, YI D, et al. Differential data augmentation techniques for medical imaging classification tasks [C]. Proceedings of the American Medical Informatics Association Annual Symposium, Washington, Nov 4-8, 2017: 979.
- [24] GARCIA GARCIA A, ORTS ESCOLANO S, OPREA S, et al. A review on deep learning techniques applied to semantic segmentation[J]. ArXiv Preprint, 2017, ArXiv: 1704.06857.
- [25] ZHONG Z, ZHENG L, KANG G, et al. Random erasing data augmentation [C]. Proceedings of the



- AAAI Conference on Artificial Intelligence, Menlo Park: AAAI, 2020: 13001-13008.
- [26] GUO H, MAO Y, ZHANG R. Mixup as locally linear out-of-manifold regularization[C]. Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Honolulu, Jan 27-Feb 1, 2019, Menlo Park: AAAI, 2019: 3714-3722.
- [27] FRID-ADAR M, DIAMANT I, KLANG E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification [J]. Neurocomputing, 2018, 321: 321-331.
- [28] CHEN G, CHEN P, SHI Y, et al. Rethinking the usage of batch normalization and dropout in the training of deep neural networks[EB/OL]. (2019-03-15)[2021-5-24]. <https://arxiv.org/pdf/1905.05928v1.pdf>.
- [29] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[EB/OL]. (2018-04-27)[2021-5-24]. <https://arxiv.org/pdf/1710.09412.pdf>.
- [30] 林成创,单纯,赵淦森,等. 机器视觉应用中的图像数据增广综述[J]. 计算机科学与探索, 2021, 15(4): 583-611.
- [31] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition[C]. European Conference on Computer Vision, Springer, Cham, 2016:499-515.
- [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):0215004.
- [33] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features [C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 6023-6032.
- [34] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-11-29)[2021-5-24]. <https://arxiv.org/pdf/1708.04552.pdf>.
- [35] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-Cam: Visualexplanations from deep networks via gradient-based localization[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 618-626.

### 作者简介

张家钧,硕士研究生,主要研究方向为刑事智能技术、模式识别等。

E-mail: ppsuczjj@163.com

唐云祁,博士,副教授,主要研究方向为模式识别、刑事智能技术等。

E-mail: tangyunqi@ppsuc.edu.cn

杨智雄,硕士研究生,主要研究方向为刑事智能技术、模式识别等。

E-mail: 1298457116@qq.com