

文章编号 1004-924X(2018)12-3040-11

## 基于改进深度卷积网络的铁路入侵行人分类算法

郭保青<sup>1,2\*</sup>, 王 宁<sup>1,2</sup>

(1. 北京交通大学 机械与电子控制工程学院, 北京 100044;

2. 北京交通大学 载运工具先进制造与测控技术教育部重点实验室, 北京 100044)

**摘要:** 异物侵入铁路限界严重影响行车安全, 识别铁路限界内的人员侵入对保证铁路运营安全具有重要意义。由于既有铁路图像异物侵入检测系统只能检测报警图像, 无法区分是人员侵入的正确报警还是光线干扰导致的误报警, 为了降低上述误报警, 本文建立了铁路异物侵限报警样本的训练集和测试集, 提出了将改进的深度卷积网络提取的高层 Alex 特征和 HOG 特征相结合并用于深度卷积神经网络模型训练的分类算法。首先引入了改进的 AlexNet 深度卷积神经网络模型, 提取了自动学习的 Alex 高层特征, 然后将其与 HOG 特征相结合形成 Alex-HOG 组合特征, 最后利用组合特征对分类网络进行训练。铁路异物侵限报警测试样本库的实验表明, 该方法对 1 498 张测试样本图像的识别准确率高达 98.46%, 时间为 3.78 s, 实时性和准确率均有较大提高, 对降低系统误报率具有重大意义。

**关键词:** 铁路异物分类识别; 行人检测; 深度卷积网络; HOG 组合特征

**中图分类号:** TP391; U216.3 **文献标识码:** A **doi:** 10.3788/OPE.20182612.3040

## Pedestrian intruding railway clearance classification algorithm based on improved deep convolutional network

GUO Bao-qing<sup>1,2\*</sup>, WANG Ning<sup>1,2</sup>

(1. School of Mechanical, Electronic and Control Engineering,  
Beijing Jiaotong University, Beijing 100044, China;

2. Beijing Jiaotong University, Key Laboratory of Vehicle Advanced Manufacturing,  
Measuring and Control Technology, Ministry of Education, Beijing 100044, China)

\* Corresponding author, E-mail: bqguo@bjtu.edu.cn

**Abstract:** Objects intruding railway clearance pose great threat to normal railway operations. Identifying intruding pedestrians within the railway clearance limit was of great significance to ensure the safety of railway operations. The existing railway intrusion detection system only detected the intrusion, but did not distinguish whether it was a true alarm of pedestrian intrusion or false alarm caused by light interferences. To reduce false alarms, a training and test set of the alarm image samples were established. A pedestrian classification algorithm based on improved deep convolutional network, trained with combined features of HOG and high-level Alex was then proposed. First, an improved AlexNet deep convolutional neural network model was introduced to extract high-level Alex features

**收稿日期:** 2018-04-27; **修订日期:** 2018-06-17.

**基金项目:** 国家重点研发计划资助项目(No. 2016YFB1200402); 铁路总公司重点研发计划资助项目(No. 2017T001-B); 国家留学基金委员会资助项目(No. 201707095075)

by automatic learning; the extracted features were then combined with HOG features to form the combined features of Alex-HOG. Finally, the combined features were used to train the classification network. Experiments on the test set show that the proposed method has a high recognition accuracy of 98.46% in 3.78 s for 1 498 test image samples. The improvements in both accuracy and real-time performance will greatly reduce the false alarm rate of the railway intrusion detection system.

**Key words:** railway objects classification and identification; pedestrian detection; deep convolutional network; HOG combined features

## 1 引 言

随着我国高速铁路的快速发展,铁路运输在人们生活中的重要性日益提高,高铁成网运营后的安全问题已成当务之急。异物入侵高速铁路限界将会引发行车事故,严重危害国家财产和人民生命安全。针对既有检测系统基于图像背景差分方法获取的报警图像的人工统计显示,报警图像的大部分为夜间光线干扰造成的误报警(约占报警总数的 70%),剩余报警中的绝大部分(约占报警总数的 30%)均为夜间施工人员上道作业导致的正确报警,除此之外的其他报警信息非常少。若能准确地区分这些报警图片是行人还是光线干扰造成的误报警,将能极大提高侵入检测的准确率,降低由于夜间光线干扰造成的误报警,对进一步提高既有系统报警准确率具有重要意义。

与防护电网为代表的接触式检测方法相比,视频分析方法因具有安装方便且结果直观的优点而被广泛应用于安防监控领域。在铁路监测方面,董宏辉等人<sup>[1]</sup>提出了智能视频监控图像序列的分析方法,实现了对物体入侵行为的识别及危险行为报警;魏巍<sup>[2]</sup>研究了基于双目立体视觉实现重点区域三维重建的方法,实现了异物侵入检测;史红梅等人<sup>[3]</sup>提出了基于机器视觉与嵌入式技术相结合的异物侵限自动检测方法,实现了对异物目标的快速识别;宁滨等人<sup>[4]</sup>提出了铁路远程瞭望系统的框架,并通过 Markov 随机场的异物识别算法实现了异物识别。上述方法在检测效果和实时性方面均存在不足,杨柳旭等人<sup>[5]</sup>研究了基于快速背景差分的高速铁路异物侵入检测算法,基于该算法研发的高速铁路异物侵入检测系统已在沪宁、杭甬高速铁路示范应用。该系统能在图像中标出异物的位置,但无法识别是何种异物,尤其在夜间存在光线干扰时常会发生误报警。

由于人员进入铁路限界对运营安全威胁更大,而且根据对报警结果的人工统计,真实报警的绝大部分均为人员入侵报警,误报警的绝大部分是由于光线干扰造成的,进一步区分报警图片是否包含行人可以极大程度剔除光线干扰造成的误报警,对降低既有检测系统的误报率具有重要意义,因此本文将入侵目标是否为行人作为分类目标。

目前国内外在行人检测方面的研究大多采取特征提取结合机器学习的方法。在特征提取方面,HOG 特征是行人检测中广泛应用的特征之一<sup>[6-9]</sup>,一些学者还提出了以 HOG 特征为基础的组合特征,如 Jiao Jichao 等<sup>[7]</sup>提出一种 HOG-PCA (HOGP)混合特征描述符来抑制了背景干扰的方法;黄炎等<sup>[8]</sup>提出将 HOG 与 LBP 特征结合的行人检测方法,排除了部分梯度干扰信息造成的误检;Sangeetha 等<sup>[9]</sup>提出一种基于边缘梯度(HEOG)的多尺度高效直方图人体检测方法,在各种具有挑战的人类数据集中达到了较好的检测效果。

上述研究都是基于手工特征实现的,特征适应性较差,调节过程费时费力,而近年来深度卷积网络已经在图像分类识别中表现出独特优势<sup>[10-13]</sup>。Paul 等<sup>[10]</sup>提出了融合均匀卷积神经网络分类器的混合深度神经网络模型来识别视频中的人类行为,在 UCF50 数据集中达到了 99.68% 的高准确率;Tian Y<sup>[11]</sup>针对小物体提出改进的 R-CNN 快速深度卷积神经网络模型,在交通场景车道标记数据集上取得较高的准确率且用时很短;刘智<sup>[12]</sup>等提出基于深度卷积网络和多尺度信息的人体行为识别方法,对 MSRDailyActivity3D 数据集的 11~16 种人体行为的平均识别准确率为 98%;熊昌镇等<sup>[13]</sup>提出了一种结合深度卷积神经网络模型与主体检测的图像检索方法,1 000 幅测试图片正确率为 96.5%,比现有方法提升了 6.6 个百分点。除了上述主要用于行人分类的研究外,深度卷积网络也开始应用于复杂铁路场景

的分类上,刘文棋<sup>[14]</sup>提出了基于五层深度信念网络的铁路异物检测算法,在对实际场景视频的测试中取得了较好的效果;王洋<sup>[15]</sup>针对铁路场景提出基于卷积神经网络特征和稀疏性参数的快速特征提取算法,能够准确识别图像中的列车。

以上利用深度卷积网络进行分类的研究都需要大样本的训练数据,在小样本情况下,其表现不如传统的特征稳定,并且单一特征存在一定不足,将传统手工特征与深度卷积网络结合也是目前的主要研究方向之一。这方面的研究主要集中在人脸和行人数据集的分类上,Rui Ting 等人<sup>[16]</sup>从训练的 CNN 中提取特征图,然后基于 HOG+SVM 的特征映射选择方法,有效构建高维特征图以提高行人检测性能;丁文秀<sup>[17]</sup>提出将图像与其 SIFT 描述符的稀疏表示进行融合的方法,该分层特征学习方法在行人分类方面性能有明显提升;DaeSeong<sup>[18]</sup>提出基于局部梯度模式(LGP)并使用 CNN 应用海马结构的人脸检测方法,算法准确性提高了 3%~9%。

针对前期沪宁、杭甬高速铁路异物侵入检测系统夜间光线干扰造成的误报率高的问题,本文在前期检测结果的基础上,建立了铁路侵限异物报警训练和测试样本库,并在 pytorch 框架下提出了基于 Alex-HOG 组合特征利用改进的深度卷积网络进行铁路场景报警图片背景与行人分类识别的算法。铁路现场实验结果表明,本文算法能够应对铁路场景中复杂的光照变化,更好地提取夜间条件下物体的特征信息,能够在 3.78 s 内完成测试样本库中 1 498 张图像的分类且准确率达到 98.46%,在大幅提高检测准确性的同时能够满足实时性要求。

## 2 研究基础及方法

本课题组基于快速背景差分算法设计的高速铁路异物侵入监测系统已经在沪宁和杭甬高速铁路示范应用<sup>[5]</sup>。该系统总体结构如图 1 所示,包含了线路沿线相机、视频存储服务器、分析处理服务器、报警服务器和光纤传输网。铁路沿线的相机通过光纤网将视频传输至视频存储服务器,分析处理服务器从中读取实时视频进行异物侵限判断并将结果传输至报警服务器,报警服务器可以借助网络实现报警信息的发布。

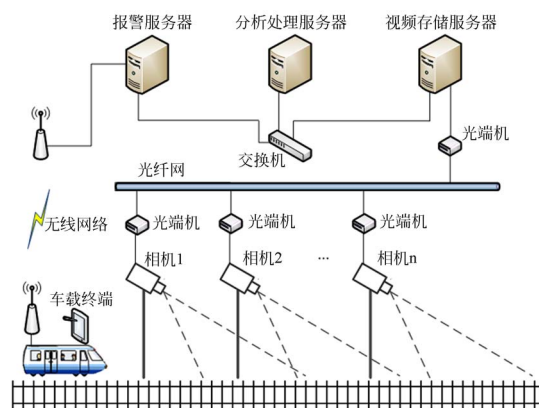


图 1 铁路异物侵限监测系统结构

Fig. 1 Structure of railway objects intrusion detection system

该系统自 2016 年 8 月安装完毕后至今一直稳定运行,积累了大量的现场报警图片。由于该系统主要通过改进的背景差分方法快速获取报警图像,现场的高实时性要求决定了不能应用过于复杂的算法,因此这些报警图像中既包括了如图 2(a)所示的正确报警,也包括了如图 2(b)所示的由于光线干扰造成的错误报警(光线变化时将背景误认为是入侵异物)。通过对大量报警结果的人工分析,正确报警中的绝大多数均为夜间人员侵入报警,错误报警中的绝大部分是由于光线变化或目标遮挡将背景信息判断为入侵物体,若能将这些错误报警和正确报警信息区分开来,就能极大降低系统的误报率,提高系统的可用性。

针对以上问题,本文主要利用改进的深度卷积网络提取特征并结合传统 HOG 特征对现有铁路异物侵入检测系统的报警信息进一步分类确认,区分报警图片是否为行人入侵。若为行人,则



(a) 检测到行人的正确报警  
(a) True alarm of intruding pedestrian



(b)光线干扰引起的误报警

(b) False alarm of disturbance light

图 2 既有异物侵限检测系统报警图像

Fig. 2 Alarm images of existing intrusion detection system

认为是正确报警;若是由光线干扰造成的背景图片,则认为是误报警。本文基于对现有系统报警图片的人工标注建立铁路异物侵限系统正确报警与错误报警样本库,在提取 HOG 特征的基础上,引入了改进的深度卷积网络,将训练好的网络模型作为一个特征提取器提取高层特征,将 Alex-HOG 组合特征输入到设计的多层分类网络结构里实现对异物侵入检测系统报警信息的再分类,算法的总体思路如图 3 所示。

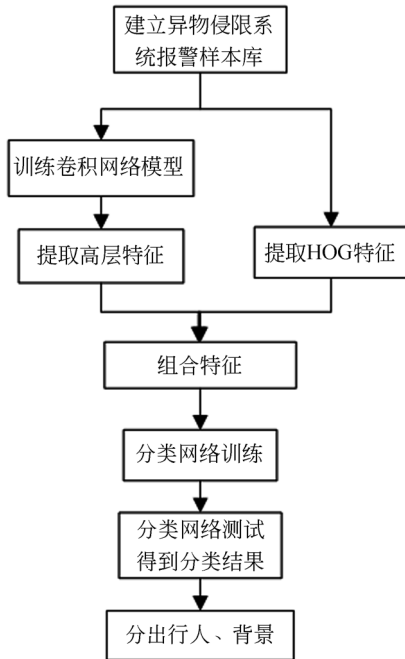


图 3 侵限异物分类算法流程

Fig. 3 Flow chart of intruding objects classification algorithm

### 3 基于 Alex-HOG 组合特征的改进深度卷积分类网络

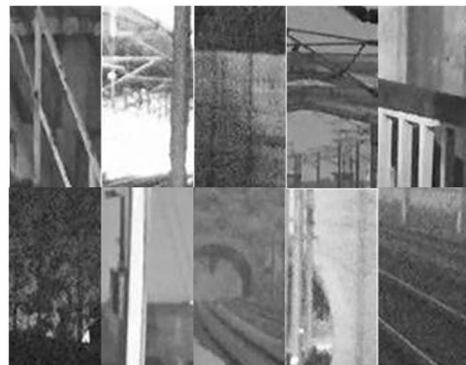
#### 3.1 构建样本库

本文是对沪宁与杭甬高铁既有异物侵限报警系统的报警图片进行再分类研究,为了后续深度卷积网络的训练和测试,首先要建立相应的报警图片样本库。样本库的建立根据图 2 所示报警图片上自动标注的红色矩形框(检测到的异物)对目标进行截取,并人为对截取的图像进行分类标注,构建行人(正确检测样本)和背景(错误检测样本)样本库,样本库中部分样本本图像如图 4 所示(彩图见期刊电子版)。



(a) 正确报警样本库

(a) True alarm sample library



(b) 错误报警样本库

(b) False alarm sample library

图 4 样本库图片

Fig. 4 Images of sample library

正确报警样本库图 4(a)中的行人样本大多为单人,也包括因既有系统自动判别异物过程中识别不准确导致的部分人像情况,如图 4(a)中第 1 行的第 2~5 张图;错误报警样本库图 4(b)中一般为杂乱的铁路背景,包括电线杆、铁轨、树、台阶、栏杆以及单纯光线干扰造成的误报警等。最后整理得到的训练样本库包括 4 000 张错误背景样本、1 500 张行人正确样本,测试样本库包括 1 000 张错误背景样本、498 张行人正确样本。由于原始报警图片中红色矩形框是利用快速背景差分算法自动捕获的,标出的异物在原始图像中分辨率的范围为  $13 \times 32$  到  $159 \times 289$  之间,为确保样本一致性,通过拉伸压缩的简单处理将样本大小统一调整为  $64 \times 128$ ,图 4 为调整后的样本。

### 3.2 HOG 特征提取

首先采用 Gamma 校正法对输入图像进行颜色空间的标准化,然后利用公式(1)和(2)计算图像每个像素的水平和垂直方向梯度

$$G_x(x, y) = H(x+1, y) - H(x-1, y), \quad (1)$$

$$G_y(x, y) = H(x, y+1) - H(x, y-1). \quad (2)$$

根据公式(3)和(4)计算像素点  $(x, y)$  处梯度幅值和梯度方向。

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}, \quad (3)$$

$$\theta(x, y) = \tan^{-1} \left[ \frac{G_y(x, y)}{G_x(x, y)} \right]. \quad (4)$$

将  $64 \times 128$  的样本图像划分成  $8 \times 8$  像素的细胞,可以产生  $8 \times 16 = 128$  个细胞,然后每 4 个细胞组成一个块,以块作为模板沿水平和垂直方向进行扫描,则水平方向将有 7 个扫描窗口,垂直方向将有 15 个扫描窗口,块的个数为  $7 \times 15 = 105$ 。将每个细胞在  $[0, \pi]$  的梯度方向平均划分为 9 个区间。然后对细胞内每个像素用梯度方向进行加权投影,得到 9 维的 HOG 特征,每相邻 4 个细胞为一个块,本文在块内利用公式(5)二范数(L2)归一化直方图,使得图像的局部变化产生的影响减小。Dalal<sup>[6]</sup>通过相应的实验证明了行人检测中使用该归一化函数能够取得较好的效果。

$$\|x\|_2 = \left( \sum_{k=1}^n |x_k|^2 \right)^{1/2}. \quad (5)$$

之后把 4 个细胞的特征向量联起来得到块的 36 维特征向量。最后将图像的所有块的 HOG 特征串联起来得到该样本的 HOG 特征,  $64 \times 128$  的样本图像总共有  $36 \times 105 = 3\,780$  个特征。这就是最终可供分类使用的特征向量。

### 3.3 改进的 AlexNet 深度卷积神经网络

#### 3.3.1 深度卷积神经网络

深度卷积神经网络通过前向传播输出计算值,再利用反向传播调整权重和偏置,从而完成了一个完整的学习过程。其中,在前向传播中,通过自底向上层层计算得到模型的计算结果输出值,如式(6)和式(7)所示:

$$x_l = f(u_l), \quad (6)$$

$$u_l = \omega_l x_{l-1} + b_l, \quad (7)$$

其中,  $x_l$  是  $l$  层的输出,  $x_{l-1}$  是  $l-1$  层的输出,也是  $l$  层的输入,  $f$  是激活函数,  $\omega_l$  是  $l$  层的权重值,  $b_l$  为偏置。

在反向传播中,根据损失值来计算梯度值从而进行学习,整个网络的梯度值从上向下通过自动求导并反向组合每一层的梯度值得到。

在激活函数的选择上,本文摒弃了传统的 Sigmoid 和 tanh 激活函数,采用了 ReLu 非饱和线性函数,它在随机梯度下降阶段的收敛速度比其他函数快很多,能够大大缩短学习周期,有效改善梯度消失及收敛波动。同时,ReLu 只需一个阈值即可得到激活值,因此计算方法更加简单。ReLu 函数的操作原理见式(8):

$$\text{relu}(x) = \max(x, 0). \quad (8)$$

该分段线性函数将正值保留,负值置为 0,实现了单侧抑制,从而使得神经元具有稀疏激活性,提高深度卷积神经网络模型的速度和效率。

由于神经网络学习过程的本质就是学习数据分布,一旦训练数据与测试数据的分布不同,那么网络的泛化能力也大大降低,网络要在每次迭代都去学习适应不同的分布,这样将会大大降低网络的训练速度。所以首先需要对输入数据做一个归一化处理,本文引入了 Ioffe 提出的 BN(Batch Normalization)方法,该方法能很大程度上加速训练过程并且放宽网络初始化的条件,同时可以降低权重衰减,并在一定程度上提升网络的识别效果。BN 算法在 GooleNet、ResNet 等新网络中有广泛地应用,其流程如下:

BN 算法在每一次迭代中的每一层输入都进行了归一化,将输入数据的分布归一化为均值为 0,方差为 1 的分布,如式(9)~(11)所示:

$$x^{(k)} = \frac{x^k - E[x^k]}{\sqrt{\text{Var}[x^k] + \epsilon}}, \quad (9)$$

$$E[x^k] = \frac{1}{m} \sum_{i=1}^m x^k, \quad (10)$$

$$\sqrt{\text{Var}[x^k]} = \frac{1}{m} \sum_{i=1}^m (x^k - E[x^k])^2, \quad (11)$$

其中:  $x^k$  表示输入数据的第  $k$  维,  $E[x^k]$  表示该维的平均值,  $\sqrt{\text{Var}[x^k]} + \epsilon$  表示标准差。

该算法一个致命的缺点是把每层的数据分布都固定了,强行归一化会破坏掉刚学习到的特征。BN 算法在第二步中设置了两个可学习变量  $\gamma$  和  $\beta$ ,去还原出上一层需要学习的数据分布,解决了以上问题。这样 BN 就把原来不固定的数据分布全部转换为固定的数据分布,而这种数据分布就是要学习到的分布,从而加速了网络的训练,如式(12)所示:

$$y^{(k)} = \gamma^k \hat{x}^{(k)} + \beta^k. \quad (12)$$

本文在样本数据学习过程中采用权值共享来达到减少参数的目的,并采用梯度下降法来更新权值,在整个网络中使用 0.000 1 的权值衰减,这个非常小的权值衰减对模型学习非常重要,它减少了模型的训练误差,权值的更新规则如式(13)和式(14)所示:

$$v_{i+1} = 0.9 v_i - 0.000 5 \epsilon \omega_i - \epsilon \left\langle \frac{\partial L}{\partial \omega} \middle| \omega_i \right\rangle_{D_i}, \quad (13)$$

$$\omega_{i+1} = \omega_i + v_{i+1}, \quad (14)$$

其中:  $i$  为迭代次数,  $v$  为动量值,  $D_i$  为第  $i$  批数据,  $\epsilon$  为学习率,手动调整学习率的值为 0.01,之后在模型基础上,再以 10 的倍数调整学习率,且所有层的学习率为同一个值,  $\omega$  为权重值,每一层的初始权重值都是通过均值为 0、方差为 0.01 的高斯分布随机得到的。

### 3.3.2 改进的 AlexNet 深度卷积网络

目前常用的深度卷积网络模型有 LeNet、ResNet、Vgg、AlexNet 以及 GooleNet 等。相比而言,Vgg、GooleNet 以及 ResNet 的模型更为复杂,层数更多,而 AlexNet 网络结构较为简单,对样本二分类也能达到较好效果,但是计算量仍然较大。本文综合考虑深度卷积网络模型的复杂程度、运行时间等因素影响,在 AlexNet 经典网络模型的基础上通过对卷积层数和卷积核大小进行优化,在保证分类效果的基础上尽量减小计算量。

经典 AlexNet 模型由 5 个卷积层和 3 个全连接层组成,并且引入 LRN(局部响应归一化)来提高网络模型的泛化能力和训练速度,但 Karen<sup>[19]</sup>

指出这个方法作用不大,所以在此对归一化方法进行改进,采用 BN 方法替代 AlexNet 网络原本的 LRN,将 BN 层置于网络激活函数层的前面,并且在全连接层不使用 Dropout 层。

另外,经典的 AlexNet 模型用到了非常大的卷积核,比如第一卷积层的  $11 \times 11$ ,之前人们的观念是卷积核越大,感受野越大,看到的图片信息越多,获得的特征越好,但是大的卷积核会导致计算量的暴增,不利于模型深度的增加,计算性能也会降低。为了解决这一问题,本文对经典 AlexNet 网络结构进行改进,用两个小卷积核的卷积层替代经典网络的大卷积层,即引入了两个  $7 \times 7$  的卷积核代替第一个卷积层 Cov 的  $11 \times 11$ ,降低卷积核个数的同时加入 BN 方法,从而实现异物侵限报警图像进行特征提取和分类识别,具体结构如图 5 所示。

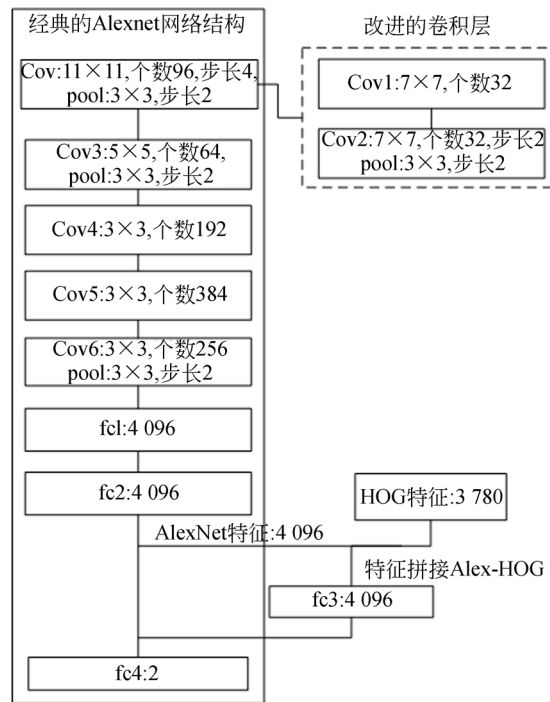


图 5 改进的 AlexNet 网络结构  
Fig. 5 Improved structure of AlexNet

改进的 AlexNet 深度卷积网络结构包括 6 个卷积层和 4 个全连接层,具体如下:

(1)第一卷积层 Cov1 的输入数据为样本库中  $64 \times 128 \times 1$  的灰度图像,卷积层由 32 个大小  $k$  为  $7 \times 7$  的卷积核构成,其包括卷积部分、BN 层和 ReLu 激励函数三部分。

(2)第二卷积层 Cov2 输入数据不变,由 32

个大小  $k$  为  $7 \times 7$ , 扩充边缘  $p$  为 2 的卷积核构成。其包括卷积部分、BN 层、ReLU 激励函数和降采样过程四部分, 降采样层中卷积核大小  $k_2$  为  $3 \times 3$ , 步长  $s$  为 2。

(3) 第三卷积层 Cov3 的输入数据为  $27 \times 59 \times 32$  的像素, 卷积层由 64 个大小  $k$  为  $5 \times 5$ , 扩充边缘  $p$  为 2 的卷积核构成, 该层结构与第二层相同。

(4) 第四卷积层 Cov4 输入数据为  $13 \times 29 \times 64$  的像素, 卷积层由 192 个大小  $k$  为  $3 \times 3$ , 扩充边缘  $p$  为 1 的卷积核构成, 该层无降采样层, 结构与第一层相同。

(5) 第五卷积层 Cov5 输入数据为  $13 \times 29 \times 192$  的像素, 卷积层由 384 个大小  $k$  为  $3 \times 3$ , 扩充边缘  $p$  为 1 的卷积核构成, 该层结构与第四层完全相同。

(6) 第六卷积层 Cov6 输入数据为  $13 \times 29 \times 384$ , 卷积层由 256 个大小  $k$  为  $3 \times 3$ , 扩充边缘  $p$  为 1 的卷积核构成, 该层无 BN 层。降采样中卷积核大小  $k_2$  为  $3 \times 3$ , 步长  $s$  为 2。

(7) 第一层全连接层 fc1 输入数据的尺寸是  $6 \times 14 \times 256$ , 神经元数目为 4 096, 包括全连接层、BN 方法、ReLU 激励函数, 该层有 4 096 个  $6 \times 14 \times 256$  尺寸的滤波器对输入数据进行卷积运算, 通过 4 096 个神经元输出运算结果; 对 4 096 个运算结果进行 BN 处理, 然后通过 ReLU 激活函数生成的 4 096 个值为本层的输出结果。

(8) 第二层全连接层 fc2 与 fc1 大体相同。只是没有用 BN 层, 该层的输出作为特征与 HOG 特征进行拼接, 得到 7 876 维的 Alex-HOG 组合特征。

(9) 第三层全连接层 fc3 为了减少计算量, 将 7 876 维的 Alex-HOG 组合特征作为输入, 转化为 4 096 维。

(10) 第四层全连接层 fc4 为分类层, 为了对沪宁与杭甬高铁的异物侵限报警的背景和行人图像的样本库进行二分类, 将样本库中行人图像的标签置为 1, 背景图像的标签置为 0, 输出为 2。

在完成样本库构建的基础上, 本文首先利用改进的 AlexNet 网络对训练样本进行训练, 并保存训练好的模型, 之后利用了迁移学习的思想, 将训练好的 AlexNet 深度卷积网络当做特征提取器, 提取了 AlexNet 高层特征, 并将该特征与

HOG 特征组合得到 Alex-HOG 组合特征, 然后利用训练样本的组合特征对分类网络进行训练, 再利用训练好的分类网络对测试样本的组合特征进行测试, 得到分类结果。

## 4 实验对比与分析

为验证本文算法的有效性, 分别进行了 3 组实验。第一组将经典行人检测方法 HOG+SVM 分别应用于常用的 INRIAPerson 行人样本库和本文的铁路行人样本库, 测试经典方法对铁路场景行人检测的适用性; 第二组利用经典 AlexNet 网络分别对 INRIAPerson 行人样本库和本文铁路场景行人样本库进行对比测试; 第三组对改进的 AlexNet 网络进行训练, 将多次随机初始化中准确率最高的模型保存下来, 对改进的 AlexNet 网络模型进行前向传播提取 Alex 特征, 并与 HOG 特征形成 Alex-HOG 组合特征, 再利用该网络进行测试, 得到本方法在铁路行人测试样本库的准确率和时间。实验均在 Windows 操作系统的 Python 环境中实现, 后两组采用了 Python 的 PyTorch 框架及 CUDA-GPU 加速方案, 采用 NVIDIA Tesla K80 双 GPU (24 GB 显存) 显卡进行 GPU 加速, CPU 为英特尔 Xeon E5-1620 四核处理器。

INRIAPerson 行人样本库是目前使用非常广泛的静态行人检测样本库, 其正样本为各场景下不同姿态的人或人群, 负样本包括房屋、动物、花草、山水、街道及车辆等各种背景图片, 该样本库包含了不同光照条件下的图片, 其训练集包含了 2 416 个正样本行人图片, 1 218 个负样本背景图片; 测试集包括 1 126 个正样本行人图片, 661 个负样本背景图片。正样本中人体大部分为站立姿态, 图片大小统一为  $64 \times 128$ 。铁路行人训练样本库如 3.1 所述包括 4 000 张背景、1 500 张行人, 测试样本库包括 1 000 张背景、498 张行人, 图片大小也为  $64 \times 128$ 。

### 4.1 HOG+SVM 行人检测方法分类效果对比

HOG+SVM 是行人检测的经典算法之一, 该方法原理简单, 在静态图像行人检测实验中取得了非常好的效果。本实验将分别测试 HOG+SVM 方法对经典行人样本库 INRIAPerson 和本文铁路行人样本数据库的检测准确率, 以检验该

方法对铁路行人样本库的适用性。

INRIAPerson 和本文的铁路行人样本库的特征及 HOG+SVM 方法应用于两个样本库的测试结果如表 1 所示。针对 INRIAPerson 样本库,HOG+SVM 的正检率为 94.35%,测试时间为 15.79 s;但该方法在铁路行人样本库中的正检率仅为 85.62%,检测时间为 12.11 s。由此可见,经典的 HOG+SVM 行人检测方法对 INRIAPerson 数据库能够取得较高的准确率,但对本文建立的铁路行人样本库检测率偏低,究其原因主要是 INRIAPerson 数据库虽然包含了光照变化的样本图像,但仍然是较为清晰的图片,图像质量和清晰度都较高,而铁路场景下获取的报警图像多为夜间光线干扰图片,图片清晰度不高,场景更复杂,给再分类带来了难度。两个样本库中两张典型低照度图片如图 6 所示,明显可以看出铁路场景图片的清晰度更差,因此,需要选择更好的算法以进一步提高铁路行人样本库的检测准确率。

表 1 经典 HOG+SVM 对不同数据库的检测结果

Tab. 1 Classical HOG+SVM Results for Different Dataset

	训练集正/ 负样本量	测试集正/ 负样本量	测试准 确率/%	测试 时间/s
INRIAPerson 库	2 416/1 218	1 126/661	94.35	15.79
铁路行人库	1 500/4 000	498/1 000	85.62	12.11



(a) INRIAPerson 图 (b)铁路场景图  
(a) INRIAPerson image (b)Railway image

图 6 INRIAPerson 与铁路行人样本库低照度图对比  
Fig. 6 Low-light images in INRIAPerson and Railway Dataset

### 4.2 经典 AlexNet 分类效果对比

考虑到近年经典 AlexNet 深度卷积网络在图像分类识别应用中表现出的独特优势,本文将经典 AlexNet 分类方法分别用于 INRIAPerson 数据库和铁路行人样本数据库作为第二个对比实验。

AlexNet 网络分类中至关重要的一环就是训练过程中的权值更新方法。为了更好地对样本图像进行分类,训练过程中需要不断更新权值,目前常用的权重更新方法有:SGD、Momentum、Adam、RMSProp 等,各方法的对比结果如图 7 所示,其中 SGD 和 Momentum 方法波动都较小,但是相比之下 Momentum 方法获得的损失值均值更低,因此本文将其作为梯度更新的方法,以下实验均采用该方法。

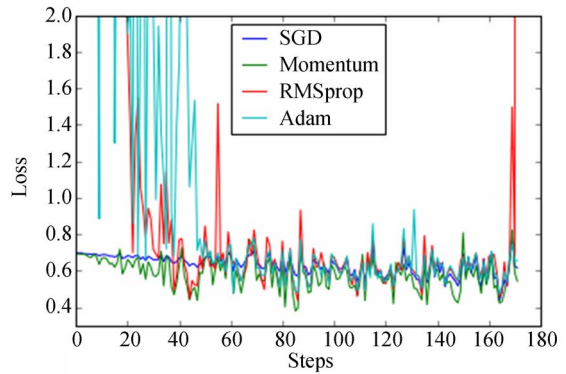


图 7 AlexNet 参数选取

Fig. 7 Parameter selection of AlexNet network

利用经典 AlexNet 对两个样本库的训练采用相同的过程,一次训练过程迭代 30 次,损失和准确率随迭代次数的变化趋势如图 8 所示,虽然在一定的范围的波动,但会逐渐收敛,没有出现严重的过拟合现象,参数设置如下:迭代次数为 30,学习率 lr=0.01,动量 momentum=0.93,衰减系数为 0.000 1。

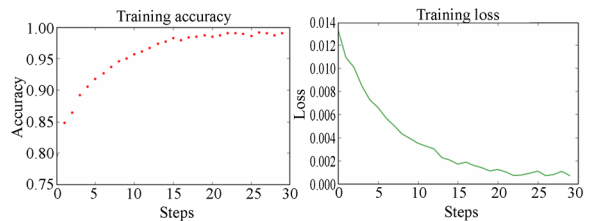


图 8 AlexNet 训练过程

Fig. 8 Training processing of AlexNet

由于训练参数可能不是最优,所以通过采取多次随机初始化的方法来提高准确率,取测试准确率最高的参数为最终结果。图 9 为两个数据库分别采用 30 次随机初始化所获得的测试集准确率,选取其中最高准确率对应的参数模型最为最终测试模型。其中 INRIAPerson 样本库最高准确率为 99.76%,测试时间为 5.32 s;铁路行人样本库最高准确率为 97.15%,测试时间为 4.42 s。相比经典 HOG+SVM 方法, AlexNet 方法针对两个数据集的准确性都有较大提高,但针对铁路行人样本集的准确率仍然低于 INRIAPerson 样本集,有近 3% 的误检率。

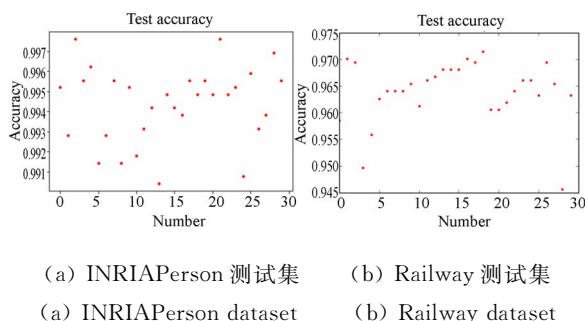


图 9 不同测试集的准确率

Fig. 9 Accuracy of different test dataset

#### 4.3 基于 Alex-HOG 组合特征的改进 AlexNet 分类效果对比

由于经典 AlexNet 网络结构计算量较大,而且对铁路行人检测准确性有待进一步提高,本文对 AlexNet 网络结构进行了改进,引入了两个  $7 \times 7$  的卷积核代替第一个卷积层的  $11 \times 11$ ,并改变卷积核的个数。为进一步提高分类的准确率,在改进的 AlexNet 网络结构基础上,引入 BN 方法,在每一个卷积层的激励函数前加入 BN 层,取代 LRN 方法,取消全连接层的 Dropout 层,并采用动态学习率以及随机初始化的方法对改进的 AlexNet 网络进行训练。

利用 4.2 所述方法对改进的 AlexNet 网络结构进行多次随机初始化训练,将准确率最高的模型保存下来作为应用模型,参数设置如下:学习率  $lr=0.001$ ,动量  $momentum=0.93$ ,衰减系数为  $0.0005$ 。图 10 为铁路行人样本库 30 次随机初始化所获得的测试集准确率。

改进 AlexNet 模型训练完成后,将准确率最高的网络模型作为特征提取器,经过一次前向传

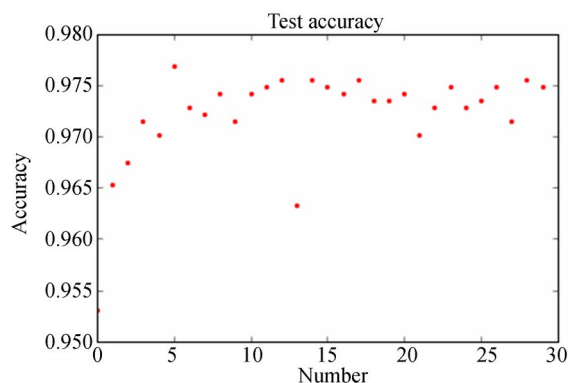


图 10 改进 AlexNet 的测试准确率

Fig. 10 Test accuracy of improved AlexNet

播得到 fc2 层的输出特征,即 Alex 特征以备后用。由于 fc2 层的特征非常不明显,图 12 列出了铁路行人样本库中部分样本原图及第一层卷积层 Cov1 的特征地图。



(a) 场景中背景与人的图像

(a) Original images of background and pedestrian



(b) Cov1 层中背景与人的特征地图

(b) Feature maps of background and pedestrian in layer Cov1

图 11 背景与人的原始图像及第一卷积层的特征地图  
Fig. 11 Background and pedestrian original images and feature maps in layer Cov1

为了进一步提高分类的准确率,将训练样本集提取的 AlexNet 特征与 HOG 特征进行串联拼接,得到 Alex-HOG 组合特征,将其作为两个全连接分类网络的输入,再次对训练样本进行训练,不断迭代并更新权值,将本次训练好的分类网络作为最终采用的网络结构,将测试集的 Alex-HOG 组合特征输入训练好的分类网络中进行

测试。

利用上述方法获得的基于 Alex-HOG 组合特征的改进深度卷积网络对铁路行人测试样本库进行测试,准确率为 98.46%,测试时间为 3.78 s。3 种算法针对铁路行人样本库的测试准确率及时间如表 2 所示。由此可见,相比于 HOG+SVM 和经典 AlexNet 分类方法,本文提出的基于 Alex-HOG 组合特征的改进 AlexNet 对铁路行人样本库分类准确率更高,耗时更短。

表 2 铁路行人样本库 3 种算法准确率及耗时

Tab.2 Accuracy and elapsed time of three algorithm for railway pedestrian dataset

	HOG+SVM	经典 AlexNet	本文算法
准确率/%	85.62	97.15	98.46
时间/s	12.11	4.42	3.78
单张耗时/ms	8.08	2.95	2.52

## 5 结 论

本论文提出了基于 Alex-HOG 组合特征利

用改进的 AlexNet 深度卷积网络进行铁路异物侵限报警图片再分类的算法,并在沪宁与杭甬高铁的实际异物侵限报警图片样本库中进行了对比实验。论文首先构建了铁路异物侵限报警样本库,而后在 python 的 pytorch 框架下对 AlexNet 网络进行改进,加入两个卷积层和 BN 方法,并改进卷积核的个数,在保证准确率的基础上减少了计算量。然后将改进的深度卷积神经网络 Alex-Net 作为特征提取器,提取分类器前一层的特征与 HOG 特征相结合,将 Alex-HOG 组合特征作为输入对分类器网络进行测试,得到分类结果。三种不同算法针对不同数据集的对比实验表明:传统 HOG+SVM 和 AlexNet 方法直接用于铁路行人检测达不到在 INRIAPerson 样本集中的准确率,而本文提出的基于 Alex-HOG 组合特征利用改进的 AlexNet 深度卷积网络进行铁路异物侵限报警图片再分类的算法的测试准确率能够达到 98.46%,测试时间为 3.78 s,在提高准确率的同时也降低了分类时间,可快速准确地完成铁路侵入行人的再分类,提高了报警的准确性。

## 参考文献:

- [1] 董宏辉,葛大伟,秦勇等. 基于智能视频分析的铁路入侵检测技术研究[J]. 中国铁道科学, 2010, 31(2): 121-125.  
DONG H W, GE D W, QIN Y, et al.. Research on railway invasion detection technology based on intelligent video analysis[J]. *China Railway Science*, 2010, 31(2): 121-125. (in Chinese)
- [2] 魏巍. 基于立体视觉的异物侵界检测方法研究[D]. 北京: 北京交通大学, 2014, 12-36.  
WEI W. *Research on Method of Intrusion Detection Based on Stereo Vision System*[D]. Beijing: Beijing Jiaotong University, 2014, 12-36. (in Chinese)
- [3] 史红梅,柴华,王尧等. 基于目标识别与跟踪的嵌入式铁路异物侵限检测算法研究[J]. 铁道学报, 2015, 37(7): 59-65.  
SHI H M, CHAI H, WANG Y, et al.. Study on railway embedded detection algorithm for railway intrusion based on object recognition and tracking [J]. *Journal Of The China Railway Society*, 2015, 37(7): 59-65. (in Chinese)
- [4] 宁滨,余祖俊,朱力强,等. 铁路远程瞭望系统研究与应用[J]. 铁道学报, 2014, 36(12): 62-69.  
NING B, YU Z J, ZHU L Q, et al.. Remote observation system of railway and its application [J]. *Journal of The China Railway Society*, 2014. 36(12): 62-69. (in Chinese)
- [5] 郭保青,杨柳旭,史红梅,等. 基于快速背景差分的高速铁路异物侵入检测算法[J]. 仪器仪表学报, 2016, 37(6): 1371-1378.  
Guo B Q, Yang L X, Shi H M, et al.. High-speed railway clearance intrusion detection algorithm with fast background subtraction[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(6): 1371-1378. (in Chinese)
- [6] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, USA; IEEE. 2005: 886-893.
- [7] JIAO JC, WANG X, DENG ZL. Build a robust learning feature descriptor by using a new image visualization method for indoor scenario recognition [J]. *Sensors*, 2017, 17(7): 1569-1585.
- [8] 黄炎,范赐恩,朱秋平,等. 联合梯度直方图和局部

- 二值模式特征的人体检测[J]. 光学 精密工程, 2013, 21(4):1047-1053.
- HUANG Y, FAN C E, ZHU Q P, *et al.*. HOG-LBP pedestrian detection[J]. *Opt. Precision Eng.*, 2013, 21(4):1047-1053. (in Chinese)
- [9] SANGEETHA D, DEEPA P. A low-cost and high-performance architecture for robust human detection using histogram of edge oriented gradients [J]. *Microprocessors And Microsystems*, 2017, 53(7): 106-119.
- [10] IJJINA, E P, MOHAN C K. Hybrid deep neural network model for human action recognition [J]. *Applied Soft Computing*, 2016, 46(1):936-952.
- [11] TIAN Y, GELEMETER J, WANG X, *et al.*. Lane marking detection via deep convolutional neural network [J]. *Neurocomputing*. 2018, 46:46-55.
- [12] 刘智, 黄江涛, 冯欣. 构建多尺度深度卷积神经网络行为识别模型[J]. 光学 精密工程, 2017, 25(3): 799-805.
- LIU ZH, HUANG J T, FENG X. Action recognition model construction based on multiscale deep convolution neural network [J]. *Opt. Precision Eng.*, 2017, 25(3): 799-805. (in Chinese)
- [13] 熊昌镇, 单艳梅, 郭芬红. 结合主体检测的图像检索方法[J]. 光学 精密工程, 2017 25(3): 792-798.
- XIONG CH ZH, SHAN Y M, GUO F H. Accurate image locating combining deep convolution network with SURF registering [J]. *Opt. Precision Eng.*, 2017, 25(3): 792-798. (in Chinese)
- [14] 刘文祺. 基于深度神经网络的铁路异物检测算法 [D]. 北京交通大学, 2016.
- LIU W Q. *Railway Obstacle Detection Algorithm Based On Deep Neural Network* [D]. Beijing: Beijing Jiaotong University, 2014, 12-36. (in Chinese)
- [15] 王洋, 余祖俊, 朱力强, 等. 基于 CNN 的高速铁路侵限异物特征快速提取算法[J]. 仪器仪表学报, 2017, 5(38): 1267-1275.
- WANG Y, YU Z Y, ZHU L Q, *et al.*. Fast feature extraction algorithm for high-speed railway clearance intruding objects based on CNN [J]. *Chinese Journal of Scientific Instrument*, 2017, 5(38): 1267-1275. (in Chinese)
- [16] RUI T, ZOU J H, ZHOU Y. Pedestrian detection based on multi-convolutional features by feature maps pruning [J]. *Multimedia Tools And Applications*, 2017, 76(23): 25079-25089.
- [17] 丁文秀, 孙锐, 闫晓星. 基于分层深度学习的鲁棒行人分类[J]. 光电工程. 2015, 42(9): 21-27.
- DING W X, SUN R, RAN X X. Robust pedestrian classification based on hierarchical deep learning [J]. *Opto Electronic Engineering*, 2015, 42(9): 21-27. (in Chinese)
- [18] DAE S K. A study of face detection algorithm using CNN with mixed-LGP and hippocampus structure [J]. *Journal of Korean Institute of Information Technology*. 2018, 16(1):11-17.
- [19] Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition [C]. *International Conference on Learning Representations (ICLR)*, 2015:1-14.

#### 作者简介:



郭保青(1978—),男,河北涿水人,副教授,博士,2004,2009年于北京交通大学分别获硕士和博士学位,主要从事铁路基础设施检测,机器视觉检测技术方面的研究工作。E-mail: bqguo@bjtu.edu.cn



王宁(1993—),女,山东烟台人,硕士研究生,2016年于北京交通大学获得学士学位,主要研究方向为轨道交通检测图像处理和机器视觉检测。E-mail: 16121280@bjtu.edu.cn