

# 基于深度残差神经网络的地铁站台门与列车门间异物自动检测方法研究<sup>\*</sup>

孔德龙 蒲 凡

(中南民族大学计算机科学学院, 430074, 武汉 // 第一作者, 讲师)

**摘 要** 针对地铁站台门与列车门间现有异物检测方法的精度差、误报率高的问题, 提出基于深度残差神经网络图像识别原理, 利用地铁站台发车指示器图像数据实现站台门与列车门间异物实时检测。首先, 搭建基于深度残差神经网络 ResNet50 模型的自动异物检测系统; 然后, 采集站台发车指示器视频帧信息建立数据集并完成系统训练; 最后, 分析自动异物检测系统对验证信息集的处理效果, 并将该系统应用于实际地铁车站中。处理效果表明: 实际应用验证中最低准确率为 98.7%, 单张视频帧处理总耗时不超过 65 ms, 满足地铁实际运营的要求。

**关键词** 地铁; 站台门异物; 残差神经网络; 图像识别; 自动检测

**中图分类号** U231.4

**DOI:**10.16037/j.1007-869x.2021.12.013

## Research on Automatic Detection Method of Foreign Objects Between Platform Screen Door and Train Door Based on Deep Residual Neural Network

KONG Delong, PU Fan

**Abstract** In view of the problems of poor accuracy and high false alarm rate of the existing detection method of foreign object between metro station platform screen door and train door, image recognition principle based on deep residual neural network is proposed, and real-time detection is realized using metro station departure indicator image. First, automatic foreign object detection system based on deep residual neural network ResNet50 model is built; then, platform departure indicator video frame information is collected to establish a data set and to complete system training; finally, the system processing effect of verification information set is analyzed, and the system is applied in actual metro stations. The processing results show that the minimum accuracy rate in actual application verification is 98.7%, and the processing time for single video

frame does not exceed 65 ms, which meets the requirements of actual metro operation.

**Key words** metro; foreign objects at platform screen door; residual neural network; image identification; automatic detection

**Author's address** School of Computer Science, South-Central University for Nationalities, 430074, Wuhan, China

为防止乘客不慎跌落站台造成安全事故, 大部分地铁在车站站台都设置了站台门。但由于站台门与列车门之间存在宽 20~30 cm 的间隙, 因而在客运高峰期容易发生乘客被夹情况<sup>[1]</sup>。国内地铁线路已发生过数起因乘客夹在空隙中而产生的运营安全事故。现有对该空隙的异物检测主要依赖司机人工目视, 或采用红外、激光等辅助方法。但人工目视容易因疏忽产生漏判, 而红外、激光等方法由于站台环境复杂会出现准确度差、误报率高的问题, 不利于列车的行车安全<sup>[2-3]</sup>。随着地铁无人驾驶技术的迅速发展, 针对无人值守列车更加需要一种准确度高、实时性好的站台自动异物检测方法。

随着人工智能技术的迅速发展, 以深度学习为代表的的人工智能图像识别技术已广泛应用于医疗、军事等领域中, 并取得了良好的应用效果<sup>[4]</sup>。残差神经网络模型是一种性能优异的神经网络模型, 相比于传统的神经网络模型具有训练时间短、图像识别准确率高、实时性好等优点<sup>[5]</sup>。因此, 本文提出构建基于深度残差神经网络 ResNet50 模型的空隙自动异物检测系统, 通过获取并处理站台闭路电视监控系统传输给发车指示器的站台门与列车门视频帧信息, 完成对站台门与列车门之间异物的自动检测。

<sup>\*</sup> 国家自然科学基金资助项目(61803314); 中央高校基本科研业务费项目(CZQ19006)

## 1 深度残差神经网络图像识别

### 1.1 基于深度学习的图像识别原理

深度残差神经网络是卷积神经网络的一种,其结构主要由卷积层、池化层和全连接层组成<sup>[6]</sup>。其典型结构是由卷积层和池化层交替排列组成多组特征提取网络,并在网络尾部设置全连接层。当系统中输入一个大小为  $W$ (长) $\times H$ (宽) $\times C$ (通道数)的张量  $T$ ,首先在网络卷积层中利用卷积函数计算系统输入张量与卷积核的卷积结果,再使用激活函数以张量的形式输出卷积层的特征图。设张量  $T_c$  为张量  $T$  在通道  $c$  上的分量,定义卷积函数为  $\text{Conv}$ ,则可以得到张量  $T$  与第  $n$  个卷积核张量  $K_n$  的卷积结果  $T_o$  为:

$$T_o = \sum_{c=1}^C \text{Conv}(T_c) = \sum_{c=1}^C M_c \cdot K_n = \sum_{c=1}^C \sum_{M_c} \sum_{K_n} (m_{w,h,c} \times k_{w,h}),$$

$$M_c \in T_c, T_c \in T, m_{w,h,c} \in M_c, k_{w,h} \in K_n \quad (1)$$

其中,张量  $M_c$  为张量  $T$  在卷积窗口内的子张量  $M$  在通道  $c$  上的分量。 $T_c$  的卷积计算原理如图 1 所示。

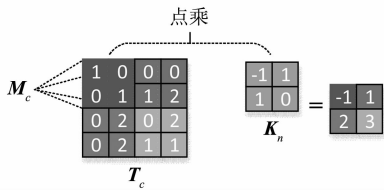


图 1 卷积的计算原理

Fig. 1 Calculation principle of convolution

将卷积结果  $T_o$  作为激活函数的输入量,遍历张量  $T_o$  的每一个像素点  $T_{o,w',h',c'}$ 。令  $b$  为偏置变量,可得到非线性激活函数在卷积核  $K_n$  上的输出张量  $T_1$  为:

$$T_1 = \sum_{T_o} \text{ReLU}(T_{o,w',h',c'} + b) = \sum_{T_o} \max\{0, T_{o,w',h',c'} + b\} \quad (2)$$

将卷积层的输出张量  $T_1$  输入池化层进行降采样处理,从而降低输出张量的维度并在减少计算量的同时不丢失图像特征。池化层通过利用  $M_c$  中的最大像素点取代卷积计算中的  $M_c \cdot K_c$ ,其中,卷积核张量  $K_c$  为在通道  $c$  上的分量。设最大池化函数为  $\text{Pool}$ ,则  $T$  的最大池化结果为:

$$T_o = \sum_{c=1}^C \text{Pool}(T_c) =$$

$$\sum_{c=1}^C \sum_{M_c} \max(M_c), \text{for } M_c \in T_c \quad (3)$$

将  $T_o$  输入到非线性激活函数中,可得到池化层的输出结果。系统输入在经过多个由卷积层与池化层交替排列的特征提取网络后进入全连接层中进行处理。设全连接函数为  $\text{Full}$ ,则  $T$  的全连接计算结果可以表示为:

$$T_o = \sum_{c=1}^C \text{Full}(T_c) = \sum_{c=1}^C T_c \cdot K = \sum_{c=1}^C \sum_{T_c} (t_{c_1,w,h,c} \times k_{w,h}),$$

$$T_c \in T, t_{c_1,w,h,c} \in T_c, k_{w,h} \in K \quad (4)$$

将所得结果  $T^o$  输入到非线性激活函数中,可得到全连接层的二维张量输出结果。

### 1.2 深度残差神经网络

提出深度残差神经网络是为了解决传统卷积神经网络模型在结构向深层化发展时,由于发生梯度弥散而出现测试与训练误差比浅层网络模型大的问题<sup>[7]</sup>。通过在网络模型中引入残差学习的概念,可以有效地解决传统多层卷积神经网络中梯度弥散的问题。一个基本的残差学习模块如图 2 所示。

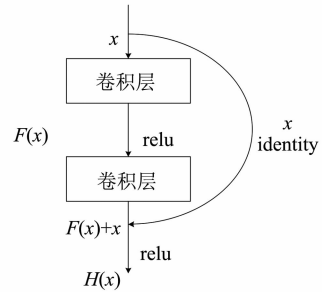


图 2 残差学习模块

Fig. 2 Residual learning module

图 2 中, $x$  为系统输入, $H(x)$  为系统目标映射, $F(x)$  为求和前的网络映射。其核心思想为:由于系统很难直接实现由  $x$  至  $H(x)$  的映射,因此不直接通过堆叠卷积网络来实现一个理想目标映射  $H(x)$  的学习,而是使网络去学习残差函数  $F(x) = H(x) - x$ 。深度残差神经网络就是由多个残差块堆叠组成的。由于残差学习突出学习过程中的微小变化,自动消除了隐含层中的部分冗余层,相对于传统的映射方式,残差映射更容易实现优化,从而解决了深层网络性能退化的问题。为了进一步提升系统性能,本文采用 bottleneck 结构建立残差学习模型,其基本原理如图 3 所示。

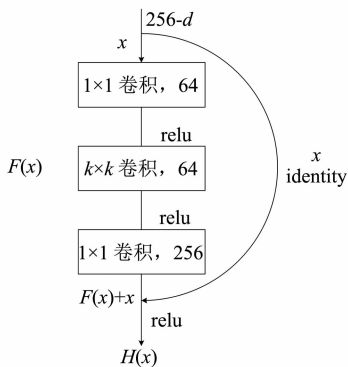


图3 Bottleneck 残差学习模块

Fig. 3 Bottleneck residual learning module

该残差学习模块包括了 3 个卷积层。其中,第一个  $1 \times 1$  的卷积层主要用于降维处理,降维后再由第二个  $3 \times 3$  卷积层处理,最后由第三个  $1 \times 1$  卷积层实现升维恢复。通过这两个  $1 \times 1$  的卷积层的设置减少了网络中的参数数量,同时增加了网络深度,相比基本模型有更好的性能。

## 2 基于深度残差神经网络图像识别的异物检测系统

### 2.1 异物检测系统总体设计

#### 2.1.1 站台空隙异物定义

由于地铁站台门与列车门都包含机械式防夹设备,当列车门或站台门由于异物存在而无法关闭时,会向列车控制系统进行报警。因此,本文的主要研究对象为站台门与列车门间隙处的异物检测。该处异物的定义为:当站台门与列车门完全关闭时,间隙处存在威胁行车与乘客安全的人体或物品。通过查阅相关文献并实际试验,该处异物的外形尺寸不小于  $30 \text{ mm}$  (长)  $\times 12 \text{ mm}$  (宽)  $\times 20 \text{ mm}$  (高)。

#### 2.1.2 异物检测系统总体设计与流程

利用深度残差神经网络来实现地铁站台车门与站台门间的异物自动检测。首先,要建立图像数据集,通过在训练站站台模拟多种异物入侵与乘客被夹方式,并对所模拟图像进行处理,建立残差神经网络训练集与测试集;其次,利用所建网络训练集对深度残差神经网络 ResNet50 模型进行迭代训练并完成对测试集的识别;最后将真实站台图像数据输入已训练完成的残差神经网络中,给出检测结果以实现自动异物检测。其流程如图 4 所示。

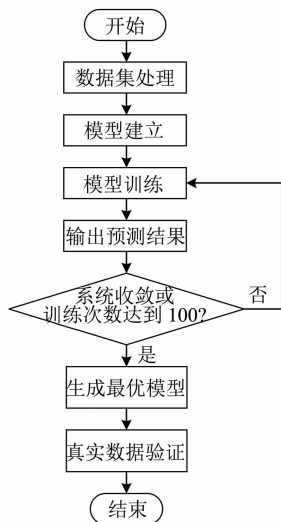


图4 自动异物检测系统检测流程图

Fig. 4 Detection flow chart of automatic foreign object monitoring system

### 2.2 基于残差神经网络图像识别的异物检测方法

#### 2.2.1 系统特征集的建立

由于安全原因,无法直接在真实地铁车站中获取站台门与列车门间存在的异物图像数据。因此,采用在地铁训练站模拟多种实际上、下客状况并采集图像信息,共得大小为  $224 \text{ mm} \times 224 \text{ mm}$  的 1 906 张图像数据。通过样本扩充处理,最终得到 12 245 张图像数据。并按 8:2 的比例划分为训练集与测试集,共得到训练集数据 9 796 张,验证集数据 2 449 张。

#### 2.2.2 模型验证与结果分析

本文采用目前较为先进的 ResNet50 (残差 50) 模型。该模型是由多个残差网络堆叠而成的深层卷积神经网络,总计有 152 层。在图像识别领域有较为成熟的应用,共进行了 1 000 类常见物体的训练,取得了不错的成绩<sup>[8]</sup>。因此,基于迁移学习原理借助已成熟的训练模型,针对本文图像分类问题的特殊情况,只对该模型中的 140~152 层进行针对性训练。

模型采用 2 个节点的 Softmax 层,使用非线性 Sigmoid 激活函数,Dropout 概率  $p$  设置为 0.25,学习率  $a$  设置为 0.1;系统训练最大迭代次数设置为 100 次。通过训练集数据训练使该模型最终实现优化。对该模型进行 100 次迭代训练后,训练集与验证集准确率与数据集损失值如图 5~6 所示。

由图 5~6 可得,在初始阶段训练集准确率为 98.38%,损失值为 0.046 6;而验证集的准确率为

98.98%, 损失值为  $0.051\ 3$ 。随着训练次数的增加, 模型趋于收敛。最终训练集准确率达到 100%、损失值为  $4.39\times10^{-7}$ , 验证集准确率达到 99.9%, 损失值为  $6.04\times10^{-6}$ , 证明该模型具有较强的泛化能力, 满足检测需要。重复多次试验该模型皆能收敛, 且相对误差很小, 并在容许范围内。

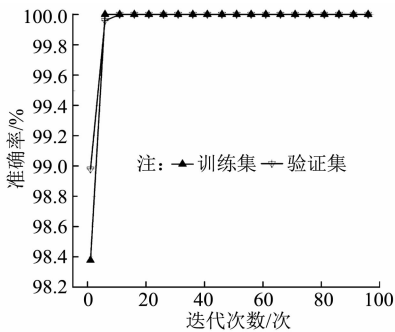


图 5 系统准确率  
Fig. 5 Systematic accuracy

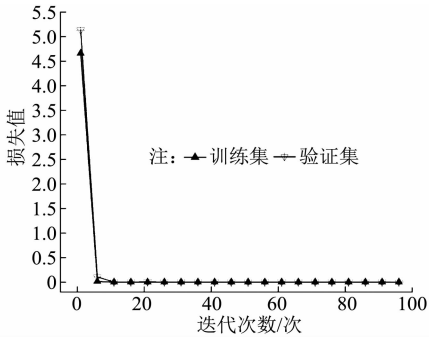


图 6 系统数据损失值  
Fig. 6 System data loss value

特征提取的能力决定了系统对验证集数据的识别准确率。因此, 对系统输入层、浅层卷积层、深层卷积层的输出进行可视化处理。通过对输出通道进行遍历得出 Channel 个大小为  $\text{Size}\times\text{Size}$  的图像, 并将其拼接为一张图片。输入层的输出维度为  $224\times224\times3$ , 对第三维遍历得到 3 张分辨率为 224 像素  $\times$  224 像素的特征图, 将其拼接后如图 7 所示。



图 7 输入层特征图  
Fig. 7 System data loss value

浅层卷积层输出维度为  $112\times112\times64$ , 对第三维遍历得到 64 张分辨率为 112 像素  $\times$  112 像素的特征图, 将其拼接后如图 8 所示。

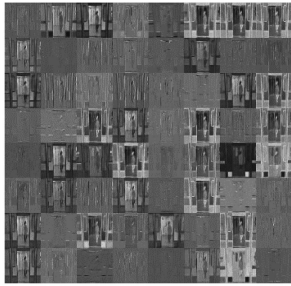


图 8 浅层卷积层特征图  
Fig. 8 Shallow convolution layer characteristics

深层卷积层输出维度为  $14\times14\times256$ , 对第三维遍历得到 256 张分辨率为 14 像素  $\times$  14 像素的特征图, 将其拼接后如图 9 所示。

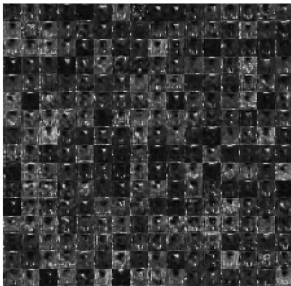


图 9 深层卷积层特征图  
Fig. 9 Deep convolution layer characteristic

从图 7~9 可以看出, 系统准确的提取出了输入数据中的特征, 达到了预期的识别目的。由于地铁安全运营要求该异物识别系统应具有较强的实时处理能力, 因此对 1 000 张数据集的读取和预测耗时进行统计。其中, 处理平台环境: 操作系统为 Windows 10 系统, 处理器为 Core i7-6700, 内存为双通道 16 GB, 显卡为 NVIDIA GTX 1050Ti (4 GB 显存)。系统总耗时统计结果如图 10 所示。

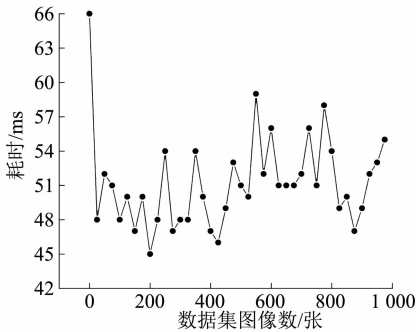


图 10 系统总耗时  
Fig. 10 System total time consumption

从图 10 中可以得出,该模型对单张数据集的预测总耗时不超过 47 ms。由于帧率是 1 s 可以处理的图片数量,系统处理一张图不超过 47 ms。因此系统的处理帧率不低于 21 帧/s,具有良好的实时处理能力,能够满足实际应用需求。利用该自动异物检测系统对武汉地铁某车站中间位置站台门与列车门 1 周视频数据进行处理验证。以线路运行过程中每小时模型预测准确性与单张视频帧总耗时的最低值为统计目标,取 1 周内的平均值进行统计,可得表 1。

表 1 实际预测准确率与单张视频帧总耗时统计  
Tab.1 Statistics of actual predictive accuracy and single video frame total time consumption

时间段	最低准确率/%	最高总耗时/ms	最低帧率/( 帧/s)
06:00—07:00	99.69	42.13	49.14
07:01—08:00	98.84	60.37	16.56
08:01—09:00	98.76	64.48	15.51
09:01—10:00	98.93	61.52	16.25
10:01—11:00	99.42	50.46	19.82
11:01—12:00	99.37	53.31	18.76
12:01—13:00	99.54	52.34	19.11
13:01—14:00	99.42	51.58	19.39
14:01—15:00	99.23	56.87	17.58
15:01—16:00	99.65	48.72	20.52
16:01—17:00	99.23	56.66	17.65
17:01—18:00	98.84	62.35	16.04
18:01—19:00	98.91	61.33	16.31
19:01—20:00	99.26	55.47	18.03
20:01—21:00	99.45	51.52	19.01
21:01—22:00	99.36	52.83	18.93
22:01—23:00	99.74	40.94	24.43

由表 1 可以得出,在线路早、晚高峰时段,系统的预测准确率与处理总耗时有部分下降。产生该现象的原因主要为在早、晚高峰时段车站上、下客人数增多,乘客在站台门前的候车与拥挤行为导致测试环境变为复杂,导致系统预测性能相对下降。但系统全时段最低准确率仍不低于 98.7%,最高单张视频帧总耗时不超过 64 ms,处理帧率不低于 15 帧/s,总体性能优良满足地铁线路正常运营的需求。

3 结语

由于现有地铁站台门与列车门间空隙异物检测方法存在检测精度差、误报率高,有一定的安全隐患,影响了列车的安全运营。本文通过模拟地铁正常运营中上、下客方式建立图像数据集,搭建并训练基于深度残差神经网络 ResNet50 模型的自动

异物检测系统,利用真实车站发车指示器视频帧数据对自动异物检测系统进行测试。测试结果表明,利用该方法实现对地铁站台门与列车门间空隙的异物自动检测是可行的,且具有较高的检测精度与实时能力。

参考文献

[ 1 ] 施伟,郑兆鹏,班琪. 基于改进 Hough 算法的地铁屏蔽门异物检测技术[J]. 城市轨道交通研究,2016(11): 81.  
SHI Wei, ZHENG Zhaopeng, BAN Qi. Detection technique of metro shielding door foreign body based on the improved Hough algorithm[J]. Urban Mass Transit,2016( 11 ): 81.  
[ 2 ] 卢昌仪. 防止地铁屏蔽门与列车间隙夹人的方案[J]. 都市轨道交通,2008(5): 82.  
LU Changyi. Prevention scheme of metro screen door and train gap clamp[J]. Urban Rapid Rail Transit,2008( 5 ): 82.  
[ 3 ] 魏巍. 基于立体视觉的异物侵入检测方法研究[D]. 北京:北京交通大学,2014.  
WEI Wei. Research on method of intrusion detection based on stereo vision system [ D ]. Beijing: Beijing Jiaotong University,2014.  
[ 4 ] 王飞,张莹,邵豪,等. 多尺度残差网络模型的研究及其应用[J]. 电子测量与仪器学报,2019(4): 19.  
WANG Fei, ZHANG Ying, SHAO Hao. et al. Research and application of the multi-scale residual networks model[J]. Journal of Electronic Measurement and Instrumentation, 2019 ( 4 ): 19.  
[ 5 ] 郭继昌,吴洁,郭春乐,等. 基于残差连接卷积神经网络的图像超分辨率重构[J]. 吉林大学学报(工学版),2019(5): 1726.  
GUO Jichang, WU Jie, GUO Chunle, et al. Image super-resolution reconstruction based on residual connection convolutional neural network[J]. Journal of Jilin University (Engineering and Technology Edition),2019( 5 ): 1726.  
[ 6 ] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [ J ]. Journal of Machine Learning Research, 2014 ( 1 ): 1929.  
[ 7 ] 林琦,陈婧,曾焕强,等. 基于多尺度特征残差学习卷积神经网络的视频超分辨率方法[J]. 信号处理,2019(1): 50.  
LIN Qi, CHEN Jing, ZENG Huanqiang, et al. Video super-resolution method based on multi-scale characteristics residual learning convolutional neural network[J]. Journal of Signal Processing,2020( 1 ): 50.  
[ 8 ] 罗会兰,童康. 时空压缩激励残差乘法网络的视频动作识别[J]. 通信学报,2019(10): 189.  
LUO Huilan, TONG Kang. Spatiotemporal squeeze-and-excitation residual multiplier network for video action recognition[J]. Journal on Communications,2019( 10 ): 189.

( 收稿日期:2020-04-30 )