

引用格式:汤健,田昊,夏恒,等. 基于多元特征异构集成深度学习的图像识别模型及其应用[J]. 北京工业大学学报, 2024, 50(1): 27-37.

TANG J, TIAN H, XIA H, et al. Image recognition model based on multivariate feature heterogeneous ensemble deep learning with its application[J]. Journal of Beijing University of Technology, 2024, 50(1): 27-37. (in Chinese)

# 基于多元特征异构集成深度学习的图像识别模型及其应用

汤健<sup>1,2</sup>, 田昊<sup>1,2</sup>, 夏恒<sup>1,2</sup>, 王子轩<sup>1,2</sup>, 徐喆<sup>1</sup>, 韩红桂<sup>1,2</sup>

(1. 北京工业大学信息学部, 北京 100124; 2. 智慧环保北京实验室, 北京 100124)

**摘要:** 随着城市矿产资源循环利用技术的不断发展,废旧手机回收已成为当前研究热点。受限于计算资源和数据资源的相对缺乏,目前基于线下智能回收装备的废旧手机识别精度难以达到实际应用。针对上述问题,提出一种基于多元特征异构集成深度学习的图像识别模型。首先,利用字符级文本检测算法(character region awareness for text detection, CRAFT)提取手机背部字符区域,再利用 ImageNet 预训练的 VGG19 模型作为图像特征嵌入模型,利用迁移学习理念提取待回收手机的局部字符特征和全局图像特征;然后,利用局部特征构建神经网络模式光学字符识别(optical character recognition, OCR)模型,利用全局和局部特征构建非神经网络模式深度森林分类(deep forest classification, DFC)模型;最后,将异构 OCR 和 DFC 识别模型输出的结果与向量组合后输入 Softmax 进行集成,基于权重向量得分最大准则获取最终识别结果。基于废旧手机回收装备的真实图像验证了所提方法的有效性。

**关键词:** 废旧手机; 图像识别; 迁移学习; 多元特征; OCR; 深度森林; 异构集成

中图分类号: U 461; TP 308

文献标志码: A

文章编号: 0254-0037(2024)01-0027-11

doi: 10.11936/bjtxb2022030011

## Image Recognition Model Based on Multivariate Feature Heterogeneous Ensemble Deep Learning With Its Application

TANG Jian<sup>1,2</sup>, TIAN Hao<sup>1,2</sup>, XIA Heng<sup>1,2</sup>, WANG Zixuan<sup>1,2</sup>, XU Zhe<sup>1</sup>, HAN Honggui<sup>1,2</sup>

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. Beijing Laboratory of Smart Environmental Protection, Beijing 100124, China)

**Abstract:** With the continuous development of recycling technology for urban mineral resources, recycling of used mobile phones has become a hotspot for current research. Restricted by the relative lack of computing resources and data resources, the accuracy of used mobile phone recognition based on current off-line intelligent recycling equipment is difficult to meet the practical application. Therefore, an image identification method based on multivariate feature heterogeneous ensemble deep learning method was proposed. First, the character region on the back of the mobile phone was extracted by using character region awareness for text detection (CRAFT) algorithm, the VGG19 model pre-trained by ImageNet was used as the image feature embedding model, and the local character feature and global

收稿日期: 2022-03-21; 修回日期: 2022-07-06

基金项目: 国家重点研发计划资助项目(2018YFC1900800-5)

作者简介: 汤健(1974—), 男, 教授, 博士生导师, 主要从事小样本数据建模、固废处理智能控制方面的研究, E-mail: freeflytang@bjut.edu.cn

image feature were extracted by using the transfer learning mechanism. Then, the optical character recognition (OCR) character recognition model based on neural network (NN) mode was constructed by using the local feature, and the improved deep forest classification (DFC) model based on non-NN model was constructed by using the global and local features. Finally, the outputs of heterogeneous OCR and the DFC model were integrated and fed into the Softmax to ensemble, and the final recognition result was obtained based on the criterion of maximum category weight vector. The effectiveness of the proposed method was verified based on real images of used mobile phone from recycling equipment.

**Key words:** used mobile phones; image recognition; transfer learning; multivariate feature; optical character recognition (OCR); deep forest; heterogeneous ensemble

随着科技的发展和5G的迅速普及,智能手机的更迭速度不断加快<sup>[1]</sup>。据Strategy Analytics预测,2021年全球智能手机的出货量将同比反弹6.5%,总量达13.8亿部<sup>[2]</sup>。更换速度的加快是导致手机出货量不断提高的主要原因,这也导致个人闲置手机的累计量逐年增多。废旧手机作为一种典型的城市矿产资源<sup>[3]</sup>,利用无人化、智能化的回收装备对其进行回收,能够节省大量人工成本和提升居民交投意愿<sup>[4]</sup>。废旧手机的外观大多具有唯一性,根据图像对其品牌等特征进行识别的方法具有成本低、实用性强等特点<sup>[1]</sup>。如何采用相关数据集构建分类模型对废旧手机进行智能化识别已成为当下提高智能化回收装备回收效率的研究重点<sup>[5]</sup>。

传统图像识别方法模型较为简单,无须基于大规模数据训练。文献[6]改进小波阈值去噪方法及Otsu阈值设置方法,提高传统Canny算子的边缘检测性能,实现了生产线手机膜的缺陷检测。文献[7]提出了基于形态学的红外图像边缘检测算法,改进了仅使用单一方向和固定方向权重的边缘检测算法,采用多方向结构元素提取图像边缘,通过自适应加权得到最终结果。上述传统图像算法虽然广泛应用于工业目标检测,但该类算法精度较低,难以应对回收装备所拍摄废旧手机图像训练样本稀缺场景。

目前,以卷积神经网络为基础的图像识别技术已在目标检测、人脸识别等领域广泛应用<sup>[8]</sup>。文献[9]将卷积注意力机制引入残差网络提出一种基于注意力特征融合的无人机多目标跟踪算法。文献[10]将方向梯度直方图(histogram of oriented gradient, HOG)特征和注意力模型作为孪生结构构建目标跟踪模型。文献[11]将自监督预训练和监督微调用于医疗图像分类,在图像标记稀缺的情况下提高了模型的分类精度。通常,以卷积神经网络

和大型自监督模型构建的图像分类模型通常需要海量数据集进行预训练。但是,本文所述废旧手机识别问题的数据集仅来源于回收装备样机实拍图片,废旧手机存在不同程度的磨损并且图片样本的数据量较少,难以构建行之有效的神经网络分类模型;并且,因考虑设备制造成本等因素导致拍摄图像清晰度较低,以及用户交投时的不规范操作等因素导致废旧手机图像存在完整度较差、部分区域具有镜像等问题。如何在样本量少且样本质量不高的前提下完成对手机品牌的分类已成为当前待解决的主要问题。

深度森林分类(deep forest classification, DFC)模型已被证实在小样本分类问题中具有良好表现<sup>[12-13]</sup>,如:文献[14]将DFC模型应用于卫星姿态控制系统的执行机构与传感器故障诊断;文献[15]基于GrabCut预处理算法和DFC模型构建了衣物识别模型;文献[16]提出基于DFC的多级特征融合的合成孔径雷达(synthetic aperture radar, SAR)目标识别,表明该方法相比传统模型和深度神经网络模型具有更高的识别率。

基于上述研究现状,文献[4]提出“基于并行差分进化-梯度特征深度森林的废旧手机识别模型”,利用废旧手机背部图像构建手机品牌识别模型,准确率达到80.12%;但该识别方法仅考虑手机全局纹理特征,并未对字符等重要局部特征进行融合和识别处理,其精度仍有待进一步提升。显然,目前废旧手机识别存在的难点包括:如何借助其他领域知识解决训练样本稀少、如何同时融合局部字符特征和全局图像特征、如何融入专家知识以克服废旧手机磨损和交投不规范等因素造成的图片清晰度与完整度差、如何有效提取深度特征等问题。

综上,面向废旧手机识别,本文提出一种基于多元特征异构集成深度学习的图像识别模型。首先,利用字符级文本检测算法(character region

awareness for text detection, CRAFT) 提取手机背部字符区域;接着,使用 ImageNet 预训练的 VGG19 模型<sup>[17]</sup>对手机背部图像和其字符特征区域进行特征提取,使用高维卷积特征代替单一维度特征;基于局部字符特征构建光学字符识别 (optical character recognition, OCR) 模型,基于全局图像特征和局部字符特征构建改进 DFC 模型;最后,将不同模型分类结果线性拼接,通过 Softmax 函数得出最终分类结

果。基于废旧手机回收装备实拍的图像数据集,验证了该算法在废旧手机识别中的有效性。

### 1 算法模型

为了保证回收装备在训练样本少且图像精度不高的前提下完成废旧手机品牌识别,提出了一种基于多元特征异构集成深度学习的图像识别方法,其结构如图 1 所示。

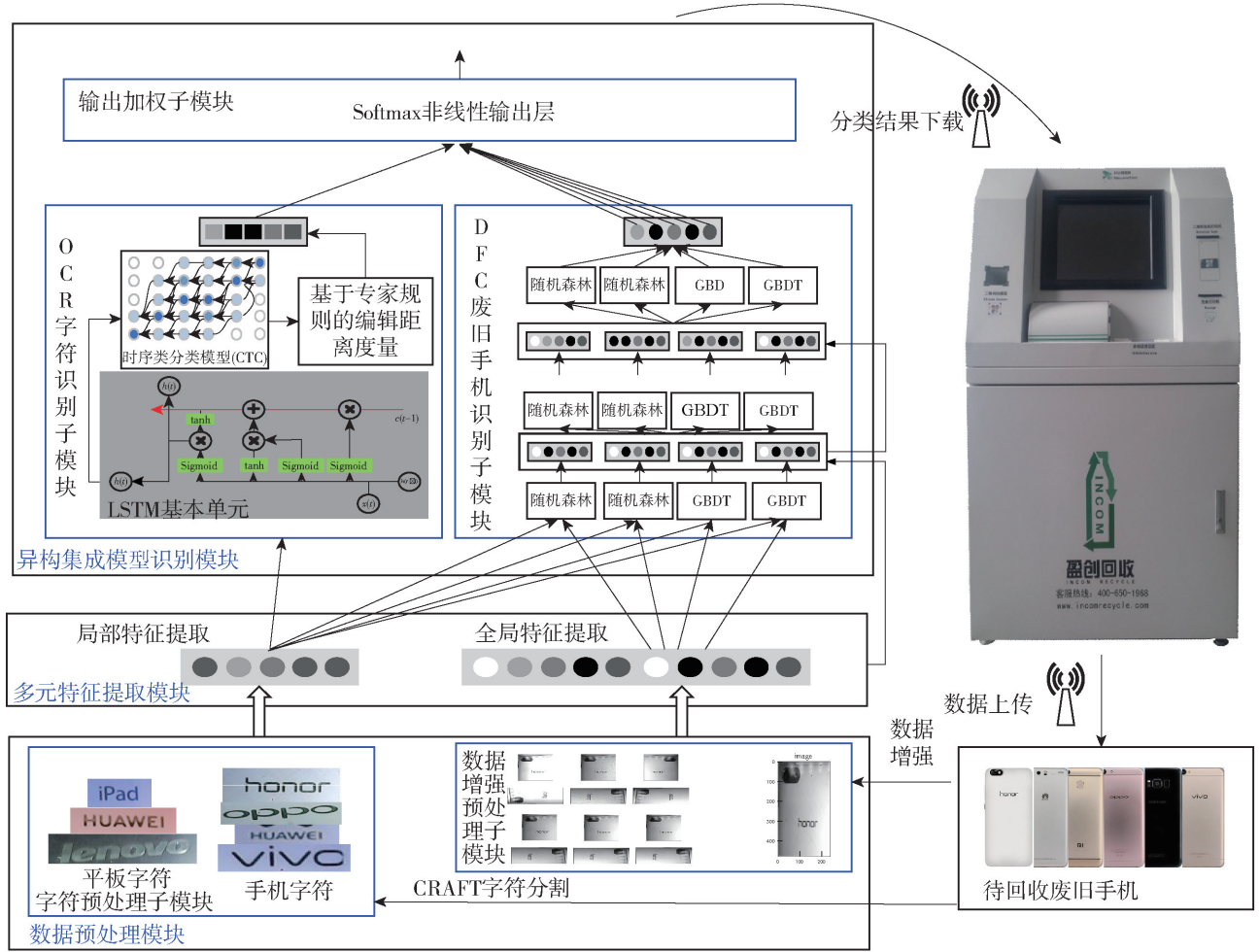


图 1 基于多元特征异构集成深度学习的图像识别模型结构

Fig. 1 Structure of multivariate feature heterogeneous ensemble deep learning method for image recognition

不同模型的功能如下。

1) 图像预处理模块:输入为  $X_{img}^{original}$ , 输出为数据增强预处理结果  $X_{img}$  和 CRAFT 字符级目标检测算法的字符预处理结果  $X_{digit}$ 。

2) 多元特征提取模块:使用基于 ImageNet 预训练的 VGG19 网络获取局部字符特征与全局图像特征在高维空间的表示,其输入为  $X_{img}$  和  $X_{digit}$ , 输出为  $X_{img}^{1,024}$  和  $[X_{digit}^{1,512}, X_{digit}^{2,512}, \dots, X_{digit}^{m,512}]$ 。

3) 异构集成模型识别模块包括 OCR 字符识别、DFC 废旧手机识别和输出加权子模型共三部

分,其中,OCR 子模块输入为  $[X_{digit}^{1,512}, X_{digit}^{2,512}, \dots, X_{digit}^{m,512}]$ , 输出为  $[\hat{y}_1]_{OCR}$ ; DFC 子模块输入为  $[X_{digit}^{1,512}, X_{digit}^{2,512}, \dots, X_{digit}^{m,512}, X_{img}^{1,024}]$ , 输出为  $[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]_{DF}$ ; 输出加权子模块将分类子模块输出结果进行加权, 获得  $[\hat{y}_1^{OCR}, \hat{y}_1^{DF}, \hat{y}_2^{DF}, \dots, \hat{y}_n^{DF}]$ , 得分最高的标签即为最终输出  $\hat{y}_{predict}$ 。

#### 1.1 图像预处理模块

在废旧手机回收流程中,用户的不规范操作会导致采集图像不完整、手机表面反光、手机图像镜像等问题,这使得直接使用回收装备获取的图片构建的识



别模型效果不佳。数据增强通过随机改变训练样本以降低识别模型对某些属性的依赖,进而提高模型的泛化能力,方法包括几何变换、颜色空间变换、内核过滤器、混合图像、随机擦除、基于生成对抗网络的增强和神经风格转移等<sup>[18]</sup>。由于几何变换可解决训练样本的位置偏差问题,而本文所述废旧手机在回收流程中会因用户摆放位置不同而使手机图像产生位置偏差。因此,本文所用数据增强以几何变换为主,包括旋转、翻折、镜像、平移和加入高斯噪声等。

此外,手机背部字符是识别其品牌的重要依据,但用户使用过程中会造成字符的磨损、遮挡等问题,仅将其背部字符作为分类依据会导致识别模型具有很大的局限性。因此笔者提取手机背部图像中的字符特征作为分类依据之一。采用 CRAFT 字符级图像定位算法确定手机字符位置并分割,将字符特征

与整体图片线性拼接,作为后续模型的输入,进而解决难以依据手机图像构建分类模型、单字符特征构建模型的局限性。

研究表明,YOLO3<sup>[19-20]</sup>、Fast-RCNN<sup>[21]</sup>等目标检测算法在人脸检测、车牌检测等领域获得广泛应用,但该类算法检测的目标长宽比较为固定,且大多不会出现形变、磨损等问题。在手机回收问题中,摆放位置的差异会导致字符发生旋转和形变,同时废旧手机的部分字符存在磨损现象,因此直接标定字符所在位置进行训练的效果较差。此外,对手机图像数据集中单个字符进行人工标注的工作量难以估计,且已知目标检测数据集鲜有字符级别的标注。CRAFT 算法本质是通过字符之间紧密程度预测文本所在区域。本文所用 CRAFT 模型的训练过程如图 2 所示。

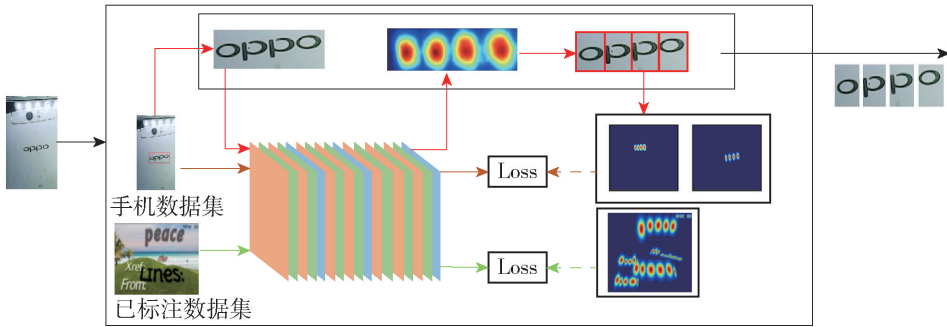


图2 CRAFT 定位裁剪模块结构

Fig. 2 Structure of CRAFT positioning and cropping module

已标注数据集包含单个字符的高斯热度图,CRAFT 算法对其进行有监督训练。对于手机背部图像数据集,首先,标记手机图像中的文本框区域,将其经过透视变换拉伸到比较正的文本框;接着,利用分水岭算法得到单个字符的位置框,生成对应的高斯热度图;最后,变换后重新贴回原图对应的标签图的相应位置。

分水岭算法的分割结果评分公式为

$$S_{\text{conf}}(w) = \frac{l(w) - \min(l(w), |l(w) - l^c(w)|)}{l(w)} \quad (1)$$

式中: $l(w)$ 表示手机图像文本框长度; $l^c(w)$ 为分水岭算法分割字符串长度结果。

进一步,计算分水岭算法的置信度

$$S_c(p) = \begin{cases} S_{\text{conf}}(w), & p \in R(w) \\ 1, & \text{其他} \end{cases} \quad (2)$$

因此,分水岭算法分割得到字符串长度后,根据式(1)得到算法评价值,若该值与真实字符长度一

致,则置信度  $S_c(p) = 1$ 。显然,评分越低说明分割结果可信度越差。

## 1.2 多元特征提取模块

本文将回收装备采集的手机图像表示为  $X_{\text{img}}$ ,其尺寸为  $400 \times 300$ 。相应地,预处理后的字符图像表示为  $X_{\text{digit}} = [X_{\text{digit}}^1, X_{\text{digit}}^2, \dots, X_{\text{digit}}^m]$ ,其中  $X_{\text{digit}}^i$ 表示字符图像中第  $i$  个字符;因单个字符尺寸为  $50 \times 50$ ,以五字符图像为例, $X_{\text{digit}}$  尺寸为  $5 \times 2\,500$ 。此处,将均源于手机背部图像的全局图像特征和局部字符特征称为多元特征。显然,上述 2 种特征因维度不同而无法直接线性组合。此处,采用基于 ImageNet 预训练的 VGG19 网络获取局部字符特征与全局图像特征在高维空间的表示,将降维后的向量特征进行线性组合,其结构如图 3 所示。

## 1.3 异构集成模型识别模块

仅由纹理特征、字符特征等单一特征构建手机识别模型,其精度难以达到实际应用需求。因此,本文采取集成学习理念<sup>[22]</sup>,将不同角度的特征进行线

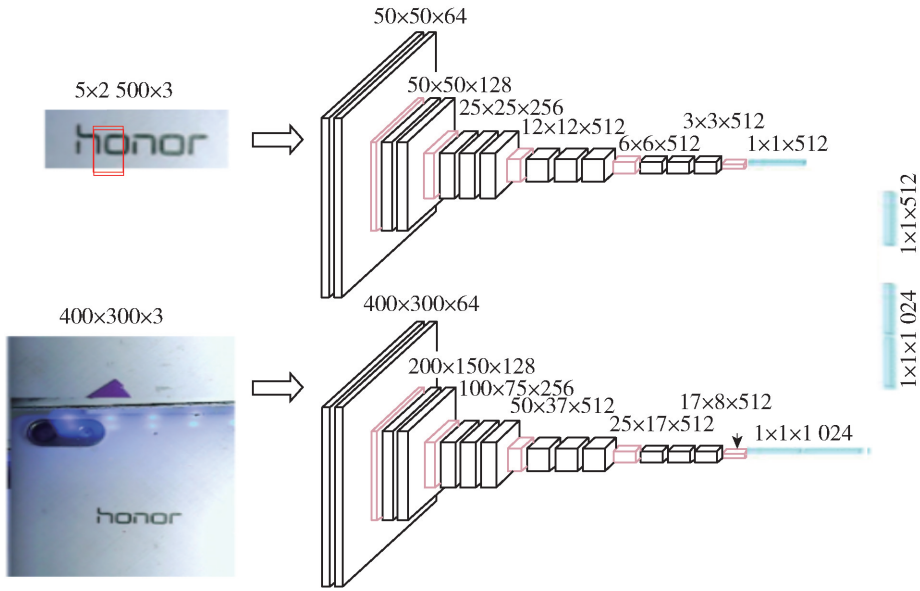


图 3 多元特征提取模块结构

Fig. 3 Structure of multivariate feature extraction module

性组合, 构建异构集成多模型, 以期通过具有多个不同机理的子模型的集成提高模型识别精度。针对前文所述废旧手机品牌分类问题, 构建 OCR 子模型、DFC 废旧手机识别子模型和输出加权子模型, 具体如下所示。

1) OCR 模型子模块

在 OCR<sup>[23]</sup> 手机背部字符识别流程中, 仅使用字符特征  $[X_{digit}^{1,512}, X_{digit}^{2,512}, \dots, X_{digit}^{m,512}]$  作为输入。首先, 通过双向 LSTM 提取含有完整上下文信息的字符序列特征; 接着, 通过 CTC 网络解决输入特征与输出序列无法对应的问题; 最后, 通过莱文斯坦距离确定 OCR 输出字符串与已知标签之间的距离, 获得手机品牌分类结果  $[\hat{y}_1]_{OCR}$ 。该子模型结构如图 4 所示。

该模块采用基于 ImageNet 预训练的 VGG19 模型作为基模型。首先, 将 VGG19 模型中的卷积层和池化层参数固化, 以充分利用迁移学习机制解决废旧手机训练样本缺失的问题; 接着, 针对不同图像特征构建不同大小的全连接层; 最后, 将不同图像的输出进行线性组合, 作为后续分类模型的输入特征。多元特征提取后的特征维度由全连接层的维度决定。针对不同大小的手机图像  $X_{img}$  和字符图像  $X_{digit}$ , 其特征提取过程为

$$\begin{cases} X_{img} \xrightarrow{f_{VGG}(\cdot)} X_{img}^{1024} \\ [X_{digit}^1, X_{digit}^2, \dots, X_{digit}^m] \xrightarrow{f_{VGG}(\cdot)} \\ [X_{digit}^{1,512}, X_{digit}^{2,512}, \dots, X_{digit}^{m,512}] \end{cases} \quad (3)$$

式中  $f_{VGG}(\cdot)$  表示 VGG19 模型输出过程。

首先, OCR 模型根据图像预处理部分所得字符特征  $[X_{digit}^{1,512}, X_{digit}^{2,512}, \dots, X_{digit}^{m,512}]$  构建  $k(k > m)$  个 LSTM 基本单元。双向 LSTM 网络包含 2 个分别表示前项和后项传递的子网络结构:

$$\vec{h}_i = (x_i, \vec{h}_{i-1}) \quad (4)$$

$$\overleftarrow{h}_i = (x_i, \overleftarrow{h}_{i+1}) \quad (5)$$

式中:  $k$  为 LSTM 基本单元超参数;  $\vec{h}_i$  表示前向 LSTM 在时刻  $i$  的输出结果;  $\overleftarrow{h}_i$  表示后向 LSTM 在时刻  $i$  的输出结果;  $x_i$  表示第  $i$  个输入。

第  $i$  时刻的双向 LSTM 输出为

$$h_i = \vec{h}_i + \overleftarrow{h}_i \quad (6)$$

接着, 采用 CTC 网络对双向 LSTM 网络输出  $[h_1, h_2, \dots, h_x]$  中的重复识别字符进行去重, 使其变为  $[y_1, y_2, \dots, y_n]$ 。由于双向 LSTM 的基本单元数多于手机字符数量  $n$ , 导致字符重复划分, 如“honor”会被分割为“hoonorr”。因此, 需将“hoonorr”等多个子串映射为正确结果“honor”, 表示为

$$\begin{cases} B(\pi_1) = B(\text{hoonorr}) \longrightarrow \text{honor} \\ B(\pi_2) = B(\text{honor}) \longrightarrow \text{honor} \\ B(\pi_n) = B(\text{hoonorr}) \longrightarrow \text{honor} \end{cases} \quad (7)$$

CTC 网络通过给定输入  $X$  的最大化后验概率  $P\langle Y|X \rangle$  获得最终结果  $Y$ , 表示为

$$P\langle Y|X \rangle = \sum_{\pi \in B^{-1}(Y)} P\langle \pi|X \rangle \quad (8)$$

式中  $\pi \in B^{-1}(Y)$  表示所有可整合为  $Y$  的子串集合。最后, 加入基于专家经验的分类结果映射模块,

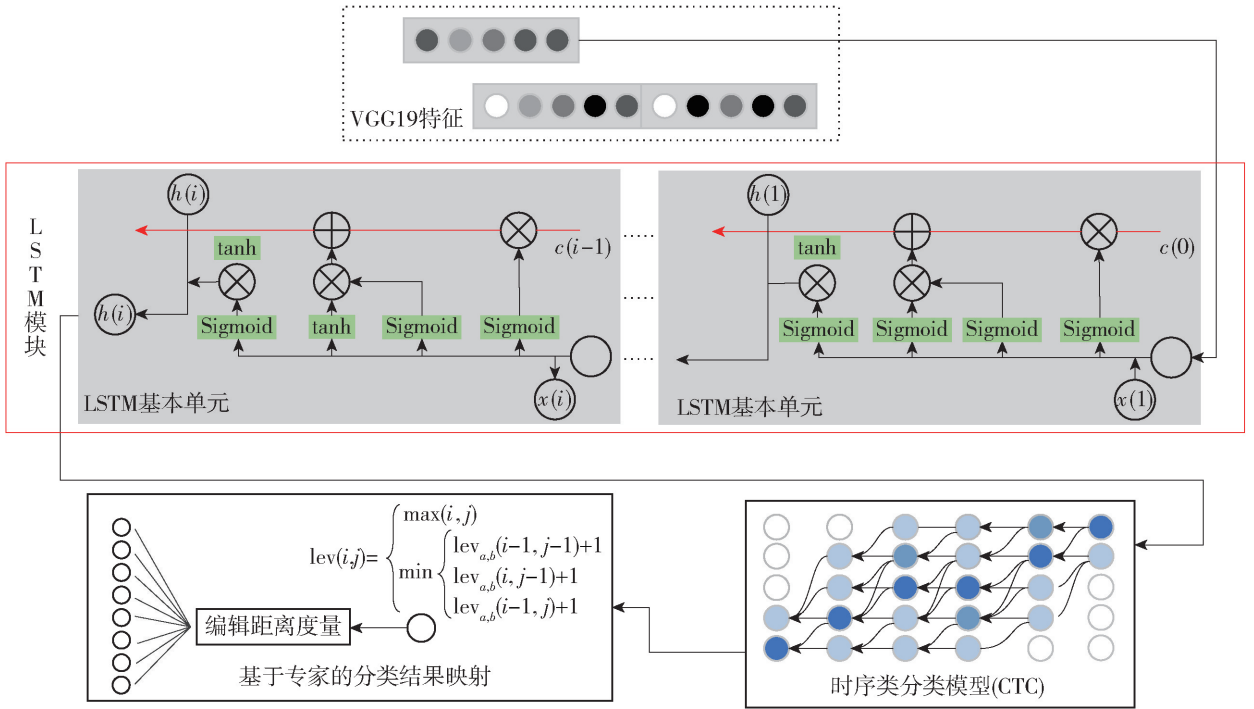


图4 OCR模型结构

Fig.4 Structure of OCR model

主要是通过莱文斯坦距离构建分类结果距离度量准则,表示为

$$\text{lev}_{\text{pre, true}}(i, j) = \begin{cases} \max(i, j) \\ \min \begin{cases} \text{lev}_{a, b}(i-1, j-1) + 1_{(\text{pre}_i \neq \text{true}_j)} \\ \text{lev}_{a, b}(i, j-1) + 1 \\ \text{lev}_{a, b}(i-1, j) + 1 \end{cases} \end{cases} \quad (9)$$

式中:pre 表示电子产品背部字符识别模块结果;true 表示标签集中标签; $i$  和  $j$  分别表示 pre 第  $i$  个字符和 true 第  $j$  个字符。

通过上述准则表征预测识别结果与真值标签的相似度,根据相似度对识别结果进行映射,解决由于单一字符的错误识别导致整体识别结果与真实标签的差异问题,进而提高模型识别准确率。

## 2) DFC 废旧手机识别模型子模块

在基于 DFC 的废旧手机图像识别流程中,将上文所述局部字符特征  $[X_{\text{digit}}^{1,512}, X_{\text{digit}}^{2,512}, \dots, X_{\text{digit}}^{m,512}]$  和全局图像特征  $X_{\text{img}}^{1,024}$  进行线性组合,作为改进 DFC 的输入特征  $X_{\text{DF}}$ ,表示为

$$X_{\text{DF}} = [X_{\text{digit}}^{1,512}, X_{\text{digit}}^{2,512}, \dots, X_{\text{digit}}^{m,512}, X_{\text{img}}^{1,024}] \quad (10)$$

首先,采用  $X_{\text{DF}}$  构建初层随机森林 (random forest, RF) 和梯度提升决策树 (gradient boosting decision tree, GBDT) 模型,获得不同输出  $\hat{y}_i^{\text{RF}}$ ;接着,

将不同模型的输出  $\hat{y}_i^{\text{RF}}$  与  $X_{\text{DF}}$  进行线性组合,作为输入传入下一层模型以构建新的 RF 和 GBDT 模型,通过当前模型的分类精度决定是否继续构建下一层网络模型;最后,当模型精度不再提高时结束模型增长,将末层 RF 和 GBDT 的分类结果加权后获得最终识别结果  $[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]_{\text{DF}}$ 。

所提改进 DFC 模型的结构如图 5 所示。

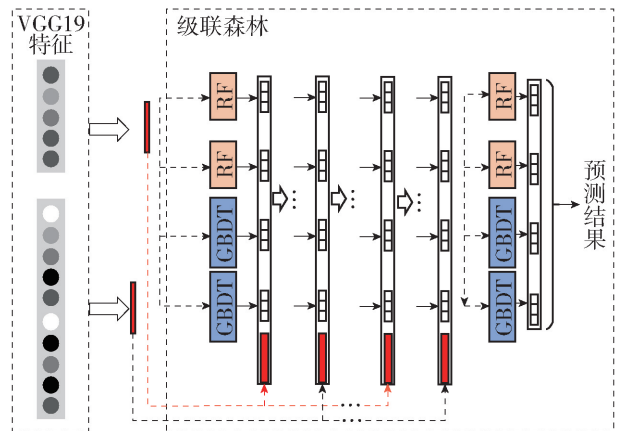


图5 改进 DFC 模型结构

Fig.5 Structure of improved DFC recognition model

## 3) 输出加权子模块

上文所述异构模型识别子模块中,OCR 模型输出的  $[\hat{y}_1]_{\text{OCR}}$  为连续字符串,经过基于距离度量的映射后输出为某一待识别手机品牌;而改进 DFC 模型

输出的  $[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]_{DF}$  为所有待识别手机品牌的概率。为解决异构子模型输出形式不同或输出结果不一致的情况, 本文提出对 2 个分类子模型进行输出加权的策略, 如图 6 所示。

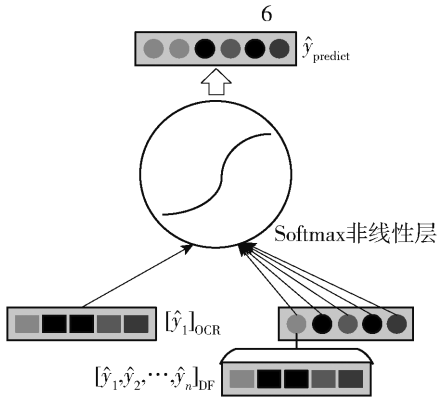


图 6 输出加权子模块结构

Fig. 6 Structure of output weighting sub-module

在本文所述识别模型中, 假定废旧手机标签为  $n$  个, 则 OCR 结果为  $[\hat{y}_1]_{OCR}$ , DFC 识别结果为  $[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]_{DF}$ , 线性拼接后可获得  $n + 1$  维结果向量  $[\hat{y}_1^{OCR}, \hat{y}_1^{DF}, \hat{y}_2^{DF}, \dots, \hat{y}_n^{DF}]$ , 将其输入目前广泛用于深度学习监督学习分类器的归一化指数函数 Softmax, 后者表示为

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{j=1}^K e^{z_j}}, j \in [1, K] \quad (11)$$

式中:  $K$  表示原始向量维度;  $e$  表示自然对数。该函数将任意实数的  $K$  维向量  $\mathbf{z}$  压缩为向量  $\sigma(\mathbf{z})$ , 并使得每个元素  $\sigma(\mathbf{z})_j$  范围都在  $(0, 1)$  内。

进而, 最终获得  $[\hat{y}_1^{OCR}, \hat{y}_1^{DF}, \hat{y}_2^{DF}, \dots, \hat{y}_n^{DF}]$  所对应的权重。此处, 将权重最高的标签作为最终分类结果  $\hat{y}_{predict}$ 。

## 2 仿真实验

### 2.1 实验数据描述

为验证本文所述算法的有效性, 选用废旧手机回收装备实拍图作为数据集进行仿真测试。废旧手机回收装备的应用场景如图 7 所示。数据集共 123 张图像, 包含 8 个类别的废旧手机品牌, 分别是华为手机 (HUAWEI)、荣耀 (HONOR)、小米 (MI)、中兴 (ZTE)、OPPO、VIVO、苹果 (iPhone), 其他品牌 (Others)。

实验阶段将样本按 8:2 的比例随机划分训练集和测试集, 在训练过程中通过十折交叉验证的方法进一步划分为训练子集和验证子集。

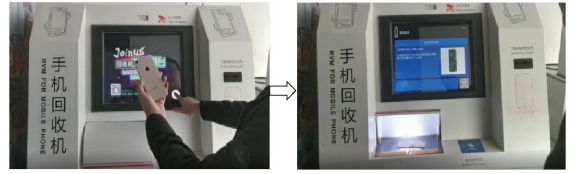


图 7 废旧手机回收装备应用场景

Fig. 7 Application scenario of waste mobile phone recycling equipment

本文所述识别模型基于 python3.7 及 TensorFlow-GPU-2.3.0 框架构建, 其训练和测试硬件环境如下: 系统为 Ubuntu20.04 LTS, 内存为 64 GB, GPU 为 GeForce RTX20280, CPU 为 Intel Xeon CPU-1650。

### 2.2 实验结果及分析

首先, 构建分类模型前采用数据增强手段对训练集和测试集样本进行扩充。以 HONOR 手机背部图像为例, 将手机背部图像进行旋转、翻折、加噪声等操作, 将 1 张手机背部图像样本扩充至 12 张。样本扩充示意图如图 8 所示。

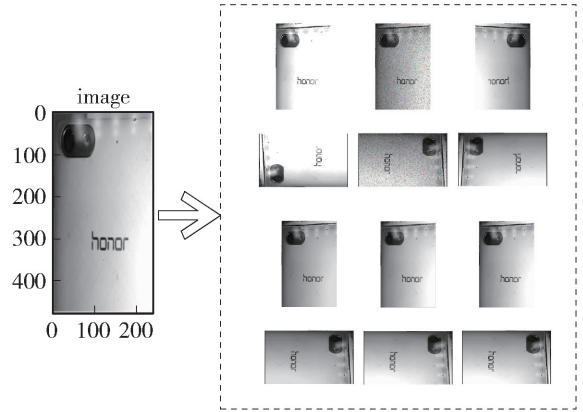


图 8 数据增强效果

Fig. 8 Data enhancement renderings

接着, 采用 CRAFT 字符分割算法对待回收废旧手机图像进行分割, 获得对应手机字符数据集, 图像预处理结果如图 9 所示。



图 9 图像预处理结果

Fig. 9 Image preprocessing results

### 2.3 多元特征提取结果

多元特征提取采用基于 1 400 万张图片 2 万类



ImageNet 数据集的 VGG19 预训练模型  $f_{VGG}(\cdot)$ 。根据输入图像的不同,在 VGG 模型中加入不同大小的全连接层,其中针对  $400 \times 300$  的全局手机图像加入 1 024 维全连接层,针对  $50 \times 50$  的局部字符图像加入 512 维全连接层。

为说明字符特征在废旧手机分类问题中的重要性,抽取 VGG19 模型中的某层输出进行可视化。全局手机图像和局部字符图像的 VGG19 特征提取图如图 10 和 11 所示。如图所示,进行 VGG19 特征提取后,字符纹理信息能够有效保留。

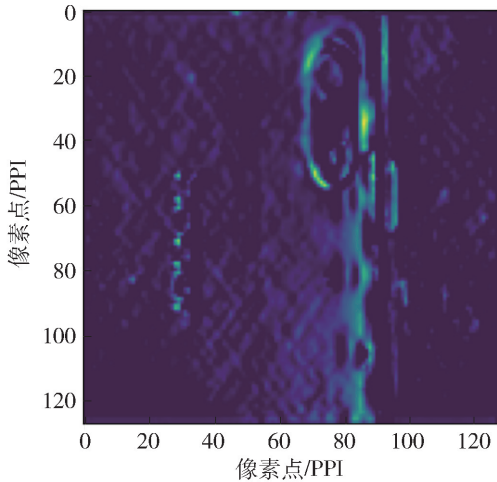


图 10 VGG19 提取的全局图像特征

Fig. 10 Global feature map of VGG19 extraction

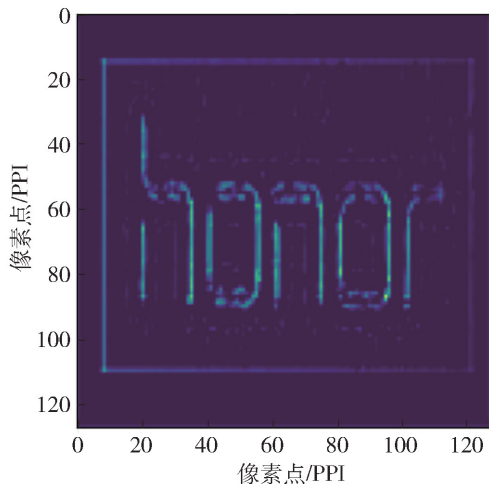


图 11 VGG19 提取的局部特征

Fig. 11 Local feature map of VGG19 extraction

## 2.4 异构集成模型识别结果

本文的 OCR 模块利用 EasyOCR 中英文字符预训练模型<sup>[23]</sup>,包括 128 个 LSTM 基本单元,即  $k = 128$ 。DFC 废旧手机识别模型使用 RF 和 GBDT 作为每层模型的基分类器,均采用 250 棵决策树,后者

选用 L1 + L2 正则化优化损失函数。

使用图像预处理后的字符图片构建的 OCR 模型分类混淆矩阵如图 12 所示,使用废旧手机图像和字符图片构建的 DFC 废旧手机识别模型分类混淆矩阵如图 13 所示。

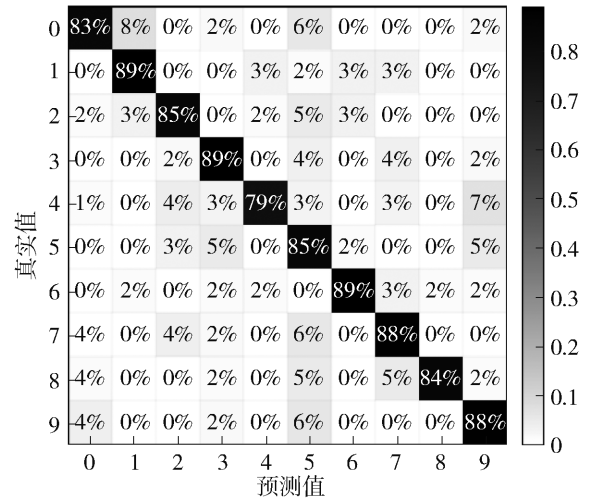


图 12 OCR 模型混淆矩阵

Fig. 12 Confusion matrix of OCR model

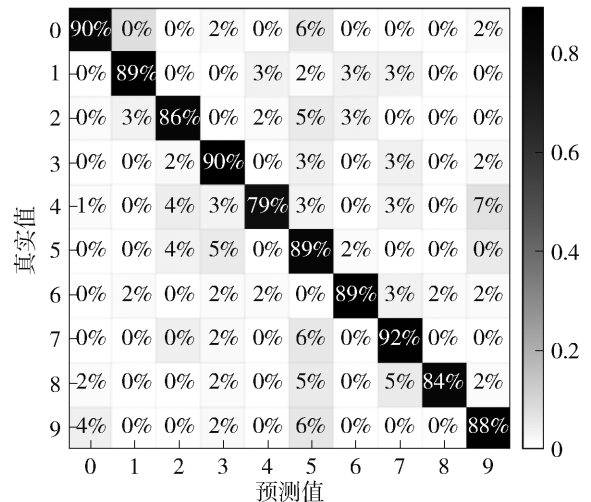


图 13 DFC 废旧手机识别模型混淆矩阵

Fig. 13 Confusion matrix of DFC model

通过输出加权子模块对上述 2 个子模型的结果进行集成,进而得到多元特征异构集成深度学习的识别模型混淆矩阵,其分类精度可达 90.17%,如图 14 所示。

## 2.5 对比实验

为验证本文所述方法有效性,将所提方法与当前主流小样本分类模型进行比较,其中:微调(fine-tuning)法用 ImageNet 作为原始数据集训练 VGG19 和 ResNet50<sup>[24]</sup>,将卷积层权重参数固定,采用本文



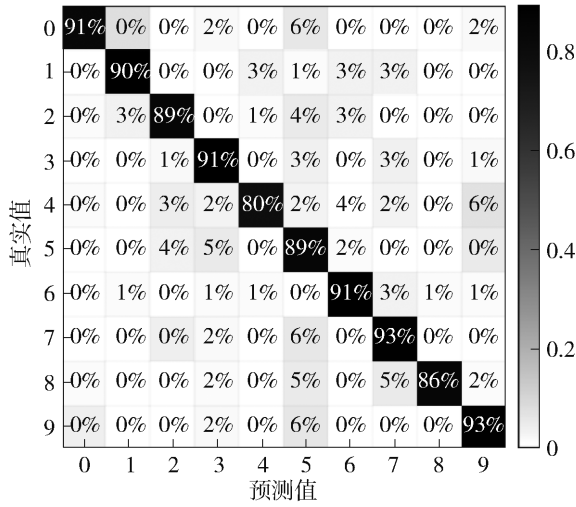


图 14 多元特征异构集成深度学习模型混淆矩阵

Fig. 14 Confusion matrix of multivariate feature heterogeneous ensemble deep learning model

所述废旧手机图像训练最后的全连接层参数;元学习中采用基于度量准则的 Siamese network<sup>[25]</sup>,采用网络爬虫获取手机背部图像训练一个 VGG19 网络,再通过废旧手机图像训练另一个 VGG19 网络,利用欧氏距离计算相同类别高清样本与实拍样本在对应模型中的高维特征相似度,具体结果如表 1 所示。此外,本文分别构建了单一特征 + 深度森林、VGG 特征 + 深度森林、VGG 特征 + OCR 模型的分类模型,精度如表 2 所示。由结果可知,本文所提方法具有最佳性能。

表 1 不同方法的对比结果

Table 1 Accuracy comparison of different methods

方法	具体实现	准确率/%
微调法	VGG19 <sup>[17]</sup>	75.83
	ResNet50 <sup>[24]</sup>	81.93
元学习法	Siamese network <sup>[25]</sup>	78.38
本文方法	VGG19 特征 + 异构集成	90.17

表 2 废旧手机识别模型精度对比结果

Table 2 Accuracy comparison results of used mobile phone recognition models

方法	精度
HOG 特征 + DFC 模型	80.12
VGG19 特征 + DFC 模型	87.93
VGG19 特征 + OCR 模型	86.37
VGG19 特征 + 异构集成模型	90.17

### 3 参数分析

本文所述异构集成深度学习识别方法中的 2 个子模型的超参数决定了模型精度。对 OCR 模型中 LSTM 单元数进行超参数实验,结果如图 15 所示。

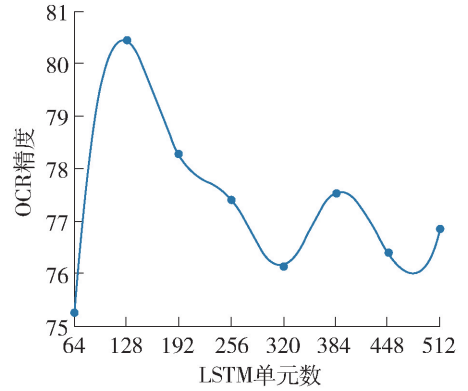


图 15 LSTM 单元数与 OCR 精度关系

Fig. 15 Relationship between the number of LSTM units and the accuracy of OCR accuracy

如图 15 所示,OCR 模型精度受 LSTM 单元数影响:LSTM 单元数增多后,模型精度在训练集表现为持续提高,但在测试集表现为略有降低。综合考虑后,将 LSTM 单元数设为 128 较为合理。

在 DFC 模型中,对森林算法种类及其含有的决策树数量 2 个超参数进行实验分析。其中,森林算法类别在完全随机森林(CRF)、普通随机森林(RF)和 GBDT 这 3 类中任选 2 类。以 VGG19 输出为 DFC 输入进行的分组实验结果如表 3 所示。

表 3 不同森林算法的 DFC 模型精度对比

Table 3 Accuracy comparison of DFC model based on different forest algorithms

模型	分类精度
RF + CRF	80.69
RF + GBDT	87.93
CRF + GBDT	84.17

为简化模型需调整的参数量,本文在不同种类的森林算法中采用相同数量的决策树。针对本文所采用的 RF + GBDT 模型,决策树数量对识别精度的影响见图 16。如图所示,综合模型精度和复杂度考虑,决策树数量设为 250 较为合理。

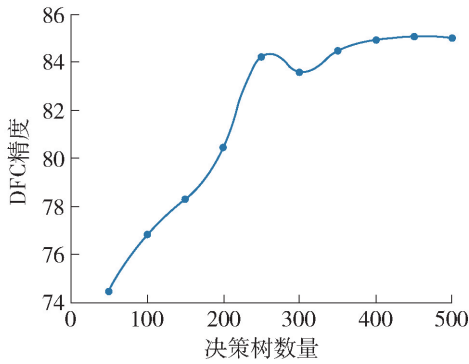


图 16 决策树数量与 DFC 精度关系

Fig. 16 Relationship between the number of decision trees and DFC accuracy

## 4 结论

本文所提基于多元特征异构集成深度学习的图像识别方法能够很好地改善废旧手机回收装备拍摄手机图片样本少、图像质量不高导致的识别精度低的问题,创新性体现在:

1) 使用 ImageNet 预训练的 VGG19 网络对废旧手机图像进行全局和局部特征提取。随着卷积层的深入,单个特征的感受野不断增大,其表征能力不断增强,更优于单一视角的特征提取方法。相比仅采用对纹理敏感的 HOG 特征所构建的 DFC 模型,使用 VGG19 构建的模型精度明显提高。

2) 采用异构集成深度学习方法构建废旧手机分类模型,基于同一数据集对不同任务构建分类模型,最后通过非线性函数对多个模型输出进行加权得到最终分类结果,所提异构集成模型相比单 OCR 和单 DFC 识别模型的精度明显提高。

仿真实验结果表明,在样本数量相同的前提下,所提方法具有较高的识别率,但特征提取模块的计算代价较高,下一步研究将通过继续优化网络结构以提高特征提取模型与所识别对象的关联度,以期进一步提高精度。

### 参考文献:

- [1] 韩红桂,甄琪,任柯燕,等. 基于孪生卷积神经网络的手机型号识别方法[J]. 北京工业大学学报, 2021, 47(2): 112-119.  
HAN H G, ZHEN Q, REN K Y, et al. Mobile phone model recognition method based on Siamese convolutional neural network [J]. Journal of Beijing University of Technology, 2021, 47(2): 112-119. (in Chinese)
- [2] LINDA S. Global handset/smartphone/feature phone sales

forecast for countries: 2007 to 2026[EB/OL]. [2021-03-01]. <https://www.strategyanalytics.com/access-services/devices/mobile-phones/smartphone/smartphones/market-data/report-detail/global-handset-smartphone-feature-phone-sales-forecast-for-88-countries-2007-to-2026?> Related.

- [3] 郭学益,张婧熙,黄国勇,等. 典型电子废弃物中金属资源开采潜力分析[J]. 中国有色金属学报, 2018, 28(2): 365-376  
GUO X Y, ZHANG J X, HUANG G Y, et al. Exploitation potentiality analysis of metal resources in typical electronic waste[J]. Chinese Journal of Nonferrous Metals, 2018, 28(2): 365-376. (in Chinese)
- [4] 王子轩,汤健,夏恒,等. 基于并行差分进化-梯度特征深度森林的废旧手机识别方法[J]. 控制理论与应用, 2022, 39(11): 2137-2148.  
WANG Z X, TANG J, XIA H, et al. Used mobile phone recognition method based on parallel differential evolution and gradient feature deep forest [J]. Control Theory & Applications, 2022, 39(11): 2137-2148. (in Chinese)
- [5] WANG Z, TANG J, CUI C, et al. Review of used mobile phone recognition method for recycling equipment [C]// 2020 39th Chinese Control Conference (CCC). Piscataway: IEEE, 2020: 1105-1110.
- [6] 林琳,吕彦诚,郭昊,等. 基于边缘检测的生产线手机膜缺陷识别方法[J]. 控制与决策, 2021, 36(4): 1017-1024.  
LIN L, LÜ Y C, GUO H, et al. Mobile phone protective film defect recognition method based on edge detection [J]. Control and Decision, 2021, 36(4): 1017-1024. (in Chinese)
- [7] 王睿男,武穆清,陈铁英,等. 基于形态学的红外图像边缘检测[J]. 北京邮电大学学报, 2021, 44(1): 66-71.  
WANG R N, WU M Q, CHEN T Y, et al. Edge detection of infrared image based on morphology [J]. Journal of Beijing University of Posts and Telecommunications, 2021, 44(1): 66-71. (in Chinese)
- [8] 刘晓军,类成玉,张晓玲,等. 基于卷积神经网络的人脸检测综述[J]. 兵工自动化, 2021, 40(6): 58-63.  
LIU X J, LEI C Y, ZHANG X L, et al. Face detection survey based on convolutional neural network [J]. Ordnance Automation, 2021, 40(6): 58-63. (in Chinese)
- [9] 刘芳,浦昭辉,张帅超. 基于注意力特征融合的无人机多目标跟踪算法[J]. 控制与决策, 2023, 38(2): 345-353.  
LIU F, PU Z H, ZHANG S C. UAV multi-target tracking algorithm based on attention feature fusion [J]. Control

- and Decision, 2023, 38(2): 345-353. (in Chinese)
- [10] 宋建辉, 孙晓南, 刘晓阳, 等. 融合 HOG 特征和注意力模型的孪生目标跟踪算法[J]. 控制与决策, 2023, 38(2): 327-334.  
SONG J H, SUN X N, LIU X Y, et al. Twin target tracking network combining HOG features and attention mode[J]. Control and Decision, 2023, 38(2): 327-334. (in Chinese)
- [11] AZIZI S, IMANI F, GHAVIDEL S, et al. Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study[J]. International Journal of Computer Assisted Radiology and Surgery, 2016, 11(6): 947-956.
- [12] ZHOU Z H, FENG J. Deep forest [J]. National Science Review, 2019, 6(1): 74-86.
- [13] 夏恒, 汤健, 乔俊飞. 深度森林研究综述[J]. 北京工业大学学报, 2022, 48(2): 182-196.  
XIA H, TANG J, QIAO J F. Review of deep forest[J]. Journal of Beijing University of Technology, 2022, 48(2): 182-196. (in Chinese)
- [14] 程月华, 江文建, 杨浩, 等. 基于深度森林的卫星 ACS 执行机构与传感器故障识别[J]. 航空学报, 2020, 41(增刊1): 195-205.  
CHENG Y H, JIANG W J, YANG H, et al. Fault identification of actuators and sensors of satellite attitude control systems based on deep forest algorithm [J]. Journal of Aeronautics and Astronautics, 2020, 41(Suppl 1): 195-205. (in Chinese)
- [15] LI H, ZHAO H, YU E, et al. A clothes classification method based on the gcForest [C] // 2018 3rd IEEE International Conference on Image, Vision and Computing (ICIVC). Piscataway: IEEE, 2018: 429-432.
- [16] 李璐, 杜兰, 何浩男, 等. 基于深度森林的多级特征融合 SAR 目标识别[J]. 电子与信息学报, 2021, 43(3): 606-614.  
LI L, DU L, HE H N, et al. Multi-level feature fusion SAR automatic target recognition based on deep forest [J]. Journal of Electronics and Information, 2021, 43(3): 606-614. (in Chinese)
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//2015 The Eleventh International Conference on Learning Representations (ICLR). Ithaca: Cornell University Library, arXiv. org. 2015: 1-14.
- [18] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.  
ZHAO K L, JIN X L, WANG Y Z. Survey on few-shot learning[J]. Journal of Software, 2021, 32(2): 349-369. (in Chinese)
- [19] LI T, MA Y, ENDOH T. A systematic study of tiny YOLO<sub>3</sub> inference: toward compact brainware processor with less memory and logic gate [J]. IEEE Access, 2020, 8: 142931-142955.
- [20] CHOWDARY V S, TEJA G P S, MOUNESH D, et al. Sign board recognition based on convolutional neural network using Yolo-3 [J]. Journal of Computational and Theoretical Nanoscience, 2020, 17(8): 3478-3483.
- [21] SUN X, WU P, HOI S. Face detection using deep learning: an improved faster RCNN approach [J]. Neurocomputing, 2018, 299: 42-50.
- [22] 汤健, 乔俊飞, 郭子豪. 基于潜在特征选择性集成建模的二噁英排放浓度软测量[J]. 自动化学报, 2022, 48(1): 223-238.  
TANG J, QIAO J F, GUO Z H. Dioxin emission concentration soft measurement based on multi-source latent feature selective ensemble modeling for municipal solid waste incineration process [J]. Acta Automatica Sinica, 2022, 48(1): 223-238. (in Chinese)
- [23] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298-2304.
- [24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [25] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Piscataway: IEEE, 2005: 539-546.

(责任编辑 张蕾)