

文章编号: 2095-2163(2021)01-0152-05

中图分类号: TP391

文献标志码: A

基于 GINet 的垃圾分类检测网络

吕泽正, 祁翔

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘要:为了更好地服务于城市垃圾分类,提高垃圾分类前端收集的工作效率,本文提出了一种基于 GINet 的智能分类垃圾网络。首先在 Kaggle 数据集和华为垃圾分类公开数据集的基础上进行了人工标注,并建立了垃圾分类的训练数据集。其次,为了提高模型的泛化能力,扩充训练样本,设计了一种具有针对性的多背景图像增强方法。最后,为了解决垃圾分类数据集中某些同类垃圾之间的尺寸、颜色差异巨大,以 VGG-16 为主干特征提取网络,构建了一个融合多特征提取与注意力机制的垃圾识别网络 (Garbage Identification Net, GINet)。仿真实验表明,该算法在复杂环境下拥有良好的鲁棒性和稳定性,检测准确率可达到 94.5%,很好地满足了工业场景下垃圾检测的准确性。

关键词: 垃圾分类; 深度学习; GINet

A garbage classification and detection method based on GINet

LV Zezheng, QI Xiang

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] In order to better serve the classification of urban garbage and improve the efficiency of front-end collection of garbage classification, this paper proposes an intelligent garbage classification network based on GINet. Firstly, manual annotation is carried out on the basis of Kaggle and Huawei's public garbage classification data set, and a training data set for garbage classification is established. Second, in order to improve the generalization ability of the model and expand the training samples, a targeted multi-background image enhancement method is designed. Finally, in order to solve the huge difference in size and color between some similar garbage in the garbage classification data set, VGG-16 is used as the main feature extraction network, and a garbage identification network that combines multi-feature extraction and attention mechanisms is constructed named Garbage Identification Net (GINet). The simulation experiment shows that the algorithm has good robustness and stability in complex environments, the detection time is only 20 ms, and the accuracy can reach 94.5%, which satisfies the high efficiency detection of garbage targets in industrial scenarios.

[Key words] garbage classification; deep learning; GINet

0 引言

近年来,中国国内城市垃圾产生量每年都在不断增长,因此,垃圾的科学分类与处理就显得尤为重要。一般情况下,垃圾处理主要是通过焚烧与填埋方式,但是这种做法却会导致土壤、空气和水资源受到严重污染,存在很大的弊端。据统计可知,城市生活垃圾主要分为干垃圾、湿垃圾、可回收垃圾和有毒有害垃圾^[1]。就目前来说,国内关于垃圾回收大多采用了人工流水线处理,但在处理过程中则暴露出效率低、成本高且对工人身体健康影响较大的问题与不足。随着工业机器人技术的快速发展,机器人代替人工作业已在多个领域开始落地实施,因此基于机器人的垃圾分类回收作业已经成为该行业的未来发展趋势。

当前,垃圾分类的检测研究相对较少,许多高效

实用的检测方法也仍处于研发阶段,并且普遍存在着检测速度慢、精准率低、适应性差等诸多问题。Diya 等人^[2]研发了一种垃圾分类系统。系统中包含了感应功能、电压调控、红外传感功能,与此同时却又发现该系统的垃圾检测量较小且抗干扰性极差。吴健等人^[3]提出一种基于颜色分割的垃圾识别系统。但是,由于垃圾的背景、尺寸、质量皆不相同,依赖人工提取相应数据不同的特征的传统算法,鲁棒性较差,处理方式十分复杂,所需时间很长,因而无法满足实时性。

近年来,随着深度学习的飞速发展,基于卷积神经网络的垃圾分类方法显示出了鲜明的优越性^[4-6]。其中,Abeywickrama 等人^[4]将垃圾分类视为图像分类,通过支持向量机和卷积神经网络对 6 类垃圾进行了识别分类,并取得了识别准确率为 83% 的识别结果。Mittal 等人^[5]制作了由 2 561 张

作者简介: 吕泽正(1997-),男,硕士研究生,主要研究方向:深度学习。

通讯作者: 吕泽正 Email:2398248206@qq.com

收稿日期: 2020-09-22

哈尔滨工业大学主办 ◆ 专题设计与应用

垃圾图片构成的数据集,并设计了准确率高达 87.69% 的 GarbNet 模型。Seredkin 等人^[6]通过融合 Faster-RCNN^[7]网络进行垃圾分类,其准确度较高,较为有效地实现了垃圾的识别,但由于某些同类垃圾之间的尺寸、颜色差异巨大,导致垃圾识别的效果却仍未臻至理想。

综上所述,本文设计研发了一种基于多尺度特征提取与注意力机制融合的垃圾识别算法(Garbage Identification Net,GINet)。在网络构架中加入了多背景图像增强方法,丰富了垃圾识别种类,增强了模型泛化能力;为了提高模型在垃圾分类任务上的准确性,算法引入了注意力机制与多特征融合模块。

1 算法设计

1.1 多背景图像增强

图像增强技术是提高模型泛化能力的重要途径,主要可分为空域法和频域法。其中,空域法是直接进行图像处理;频域法是在变换域内修正图像系数,得到增强的结果。合理运用图像增强技术,既可以提升视觉效果,增强清晰度,也可以针对给定图像的技术需求,在突出感兴趣特征的同时,抑制不感兴趣特征。

本次研究提出了一种有针对性的多背景图像增强方法。该方法主要是在训练的各个周期过程中,每 4 张图片就保存一次最大激活特征图,再上采样至原图大小,继而进行选择性的裁剪,并加以随机组合,将组合后的图像另行输入至 CNN 中重新进行训练。研究得到的算法伪代码见如下。

for 每次迭代 do

 输入,目标值 = *get_minibatch*(数据集)

 if 模式 == 训练 then

 新输入值,新目标值 = *shuffle_miniback*(输入值,目标值)

$\lambda = \text{Unif}(0,1)$

$r_x = \text{Unif}(0,W)$

$r_y = \text{Unif}(0,H)$

$r_w = \text{Sqrt}(i - \lambda)$

$r_h = \text{Sqrt}(i - \lambda)$

$x1 = \text{Round}(\text{Clip}(r_x - r_w/2, \min=0))$

$x2 = \text{Round}(\text{Clip}(r_x + r_w/2, \max=W))$

$y1 = \text{Round}(\text{Clip}(r_y - r_h/2, \min=0))$

$y2 = \text{Round}(\text{Clip}(r_y + r_h/2, \max=H))$

 输入值[:, :, x1 : x2, y1 : y2] = 新输入值[:, :, x1 : x2, y1 : y2]

$\lambda = 1 - (x2 - x1) * (y2 - y1) / (W * H)$

 新目标值 = $\lambda * \text{目标值} + (1 - \lambda) * \text{新目标值}$

 end if

 输出值 = *model_forward*(输入值)

 loss 值 = *compute_loss*(输出值,目标值)

end for

算法通过将随机的 4 张样本图片按比例混合,产生更加丰富的背景,从而提高模型局部视图识别能力。与此同时,在裁剪区域中添加其他样本的信息,能够进一步增强模型定位能力,使其准确性和鲁棒性都变得更好。多背景图像的增强效果如图 1 所示。

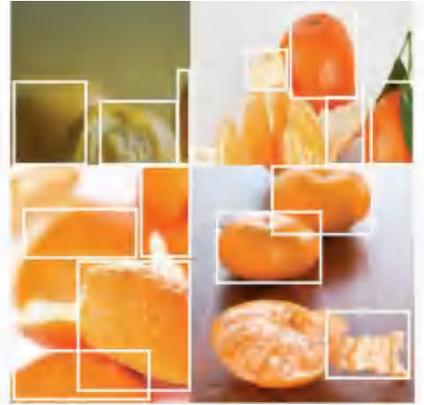


图 1 多背景图像增强效果图

Fig. 1 Schematic illustration of multi-background enhancement

1.2 模型结构

当前有着不错性能表现的目标检测框架都广泛使用了特征金字塔,从而缓解对象实例由于不同比例大小变化而带来的差异性问题。因此,本项目中提出了多层次特征金字塔网络模块来构建更有效的特征金字塔,用于检测不同尺度的对象。首先,将原始图像输入 VGG-16,使用前 13 层卷积提取基本特征,然后,将由浅到深的基本特征分别送入 3 组交替连接的 TUM 模块进行特征融合,并利用每个 TUM 模块的解码器层作为检测对象的特征。随后使用 SFAM 模块将 TUM 模块提取的有效特征聚合起来。最后,根据学习的特征生成密集的边界框和类别分数,并运用非极大值抑制算法(NMS)生成最终预测结果。整体框架如图 2 所示。

1.3 多特征提取 TUM 模块

垃圾分类数据集中某些同类垃圾之间的尺寸、颜色差异巨大,这会导致模型的识别性能变差。针对这一问题,常用的方法是通过多个池化操作,扩大特征提取层的感受野,提高识别性能,但这也导致模型的训练难以收敛。

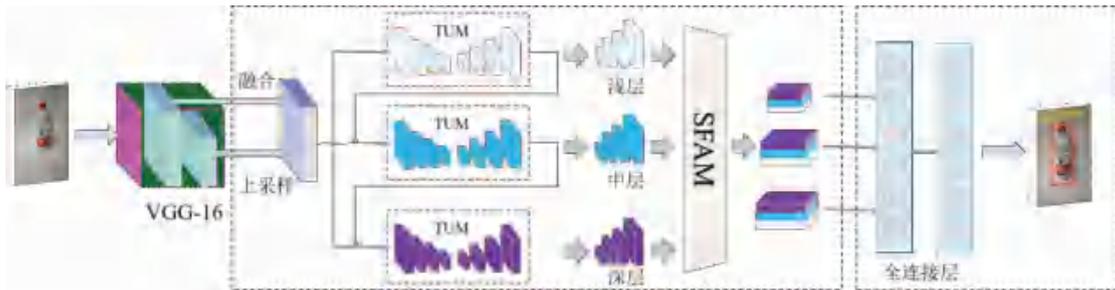


图 2 整体网络结构

Fig. 2 Overall network structure

因此,本文构建了一个多尺度的特征融合模块 (TUM 模块)。模块中采用了简化的 U 形结构,如图 3 所示。其中,编码器是一系列 3×3 ,步长为 2 的卷积层,并且解码器将这些层的输出作为其参考特征,解码器分支的上采样层后添加 1×1 卷积层和按元素求和的操作,以增强学习能力并保持特征的平滑性。每个 TUM 的解码器中的输出都会产生当前级别的多尺度特征。TUM 的输出分为 3 层,具体就是:前 TUM 主要提供浅层特征,中间 TUM 提供中等特征,后 TUM 提供深层特征。

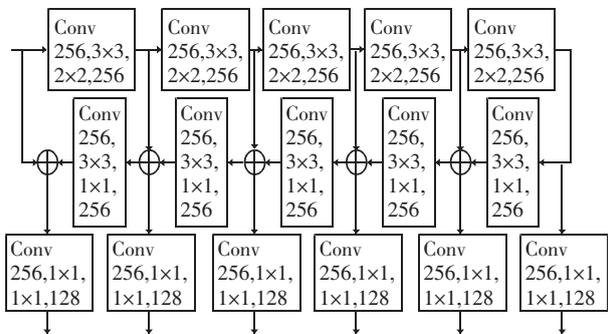


图 3 TUM 模块

Fig. 3 TUM module

1.4 注意力机制 SFAM 模块

由于相同类别垃圾的特征表征差异性可能较大,不利于其正确识别,这就要求准确地关注图像中的显著区域。由此,本文构建了注意力机制模块 (SFAM),使网络模型重点关注有利于分类的特征区域,以实现更好的特征提取功能。

SFAM 旨在将由 TUM 生成的多级多尺度特征聚合成多级特征金字塔,如图 4 所示。SFAM 的第一阶段是沿着信道维度将等效尺度的特征连接在一起。聚合金字塔中的每个比例都包含来自多级深度的特征。同时,引入通道注意力机制促使特征集中在最有益的通道,如图 5 所示。研究中,将不同比例的特征进行上采样,然后聚合在一起。在 SE 模块^[8]之后,则使用全局平均池化压缩通道来生成统计值 $z \in R^C$ 。为了完全获取通道的依赖性,算法通过 2 个全连接层学习注意力机制,其数学原理公式可写为:

$$s = F_{ca}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (1)$$

其中, σ 表示 ReLU 激活函数; δ 表示 sigmoid 函数; $W_1 \in R^{C \times C}$, $W_2 \in R^{C \times C}$, 而 C 表示压缩后的通道数。



图 4 SFAM 模块

Fig. 4 SFAM module

最终,模型通过 VGG-16 融合 TUM 特征提取模块和 SFAM 聚合模块,得出了丰富的特征信息,然后根据学习到的特征生成密集边界框和类别分数,最后对边界框采用非极大值抑制算法 (NMS) 产生最终结果。

2 实验结果与分析

2.1 实验平台

本实验的操作系统为 Ubuntu18.04,主要是利用 Python 语言和 Tensorflow 深度学习框架。本次实验中的硬件环境为 Intel (R) Core (TM) i-9750H

CPU@2.60 GHz 2.59 GHz, 内存大小为 32 GB, 显卡为 NVIDIA GeForce RTX 2080Ti, 显存为 8 G。

2.2 数据采集

数据集是对象描述的集合, 能够精确地反映出对象的特征性质与信息表现能力。本文中的垃圾分类数据主要来源于 Kaggle 和 华为垃圾分类公开数据集, 数据集中共包含 20 种物体、共 91 024 张图片, 结合垃圾分类的要求, 将所有图片分为 4 大类, 即干垃圾、湿垃圾、可回收垃圾和有毒有害垃圾。具体分类及每种垃圾数量见表 1。其中, 训练集样本占 75%, 有 68 268 张; 验证集样本占 15%, 有 13 653 张; 测试集样本占 10%, 有 9 103 张。

表 1 垃圾分类数据集

Tab. 1 Garbage classification data set

干垃圾		湿垃圾		可回收垃圾		有毒有害垃圾	
类别	数量	类别	数量	类别	数量	类别	数量
厕纸	5 000	菜叶	5 080	易拉罐	5 283	废药品	3 162
垃圾袋	4 928	果皮	4 090	书包	4 087	温度计	4 262
竹制品	5 180	蛋壳	4 485	玻璃瓶	5 080	血压计	2 062
陶瓷	4 702	杂草	6 150	镜子	4 790	废胶片	3 362
纺织品	4 809	树叶	5 560	硬塑料	5 986	化妆品	2 962

此外, 为了减少数据集中数据噪声与异常值带来的影响, 系统对样本图片进行了数据清洗。处理流程如图 5 所示。由图 5 可知, 处理中的 2 种主要情况分别是数据噪声清洗与数据异常值清洗。对此可展开剖析分述如下。

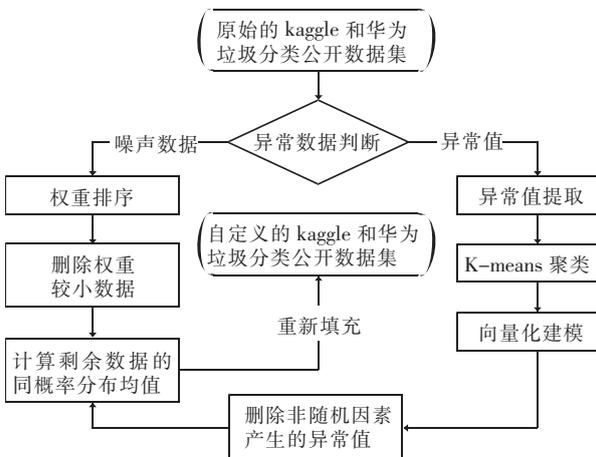


图 5 数据清洗方案

Fig. 5 Data cleaning solution

(1) 数据噪声清洗: 采用噪声删除原则, 对噪声数据进行权值排序, 删除其中权重较小的数据, 并采用同概率分布的均值数据进行自动填充。

(2) 数据异常值清洗: 提取所有异常值, 对其进行 K-means 聚类, 此后进行向量化建模, 去除非随

机因素产生的异常值, 并重新对数据集填充同概率分布的均值数据。

2.3 实验设置

本实验的训练设置主要为: 选用 SGD 优化算法训练模型, 其动量系数设置为 0.98, 一共设置 100 个迭代周期, 批次 *batch size* 设置为 16, 最大训练次数为 20^5 , 初始学习率 0.001, 在第 600 次和第 900 次时降低学习率, 同时在训练过程中保存效果最好的 *loss* 值和最后一次的模型参数。

2.4 结果分析

衡量模型训练效果通常采用 2 个指标, 即: 精确度 (*accuracy*) 和 损失值 (*loss*)。其中, 精确度是算法做出正确分类的样本数占样本总数的比例, 损失值是实际预测输出与样本真实输出间的差异。

图 6 给出了本文研发算法实际的 GINet 和当前优秀的几种分类检测网络在 Kaggle 数据集与 华为垃圾分类公开数据集上进行 20^5 次迭代训练后的精确度 (*accuracy*) 结果对比曲线, 包括 YOLOv3、RetinaNet、M2-det 等。图 7 给出了本文算法实际的 GINet 和当前优秀的几种分类检测网络在 Kaggle 数据集和 华为垃圾分类公开数据集上进行 20^5 次迭代训练后的损失值 (*loss*) 结果对比曲线。

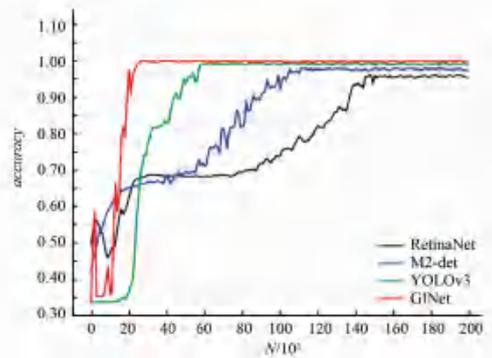


图 6 模型的精确度对比图

Fig. 6 Model accuracy comparison

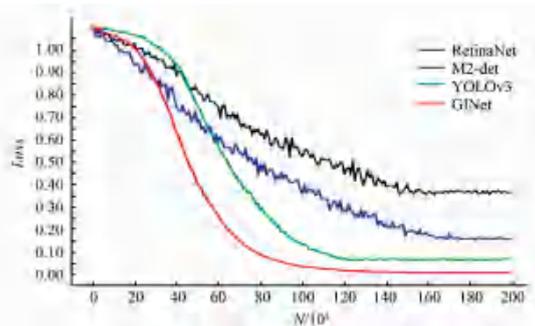


图 7 模型损失值对比图

Fig. 7 Model loss comparison