

# 图识垃圾：基于迁移学习的图像识别实证研究

方纯涵 肖逸东 向悦 王仁武

华东师范大学经济与管理学部，上海 200241

**摘要：**深度学习是目前图像识别的主流方法，迁移学习是机器学习的分支，可有效提高训练效率。为了实现生活垃圾图像分类任务，并验证迁移学习方法的优越性，本文以垃圾材质、外形及综合因素为分类依据，分别提出 A/B/C 三套分类标准并构建含三图集的 GC 垃圾图像数据库，利用深度学习和迁移学习方法设计垃圾图像识别模型。对数据集进行非均衡数据处理与数据增强；设计垃圾分类 CNN 模型，基于三图集得到模型 ModelA/B/C-CNN，识别准确率分别为 74.68%、76.55%、66.83%；使用迁移学习，利用 DenseNet201 为预训练模型，采取底层冻结并训练最后一个卷积块的方式微调，并采取 dropout 处理、L2 正则化参数使用等防止过拟合措施，得到 ModelA/B/C-TL 三个模型，准确率分别为 90.87%、90.09%、88.37%，较 CNN 模型其识别性能有显著提高。并将 ModelC-TL 模型部署于微信小程序，实现“图识垃圾”实际应用。

**关键词：**深度学习；迁移学习；卷积神经网络；垃圾分类；微信小程序

**中图分类号：**TP39

## Image recognition garbage category

### An empirical study on Image recognition based on Transfer Learning

FANG Chunhan XIAO Yidong XIANG Yue WANG Renwu

( Department of Information Management, Faculty of Economics and Management, East China Normal University, Shanghai 200241, China)

**Abstract:** Deep learning is the mainstream method of image recognition. As a branch of machine learning, transfer learning can effectively improve training efficiency. To achieve the task of household garbage image classification, and verify the excellence of the transfer learning method, we based on the waste material, shape and comprehensive factors for classification basis, respectively propose three sets of classification standards A/B/C, and build the GC garbage image database, using the method of depth and migration study design garbage image recognition model. Implement unbalanced data processing and data enhancement, design CNN model of

garbage classification, and the recognition accuracy of ModelA/B/C-CNN based on three atlas is 74.68%, 76.55% and 66.83% respectively. By transfer learning and using DenseNet201 as the pre-training model, this paper fine-tuned the bottom layer by freezing and training the last convolution block, and took dropout processing, L2 regularization parameter use and other measures to prevent overfitting. ModelA/B/C-TL models are obtained with accuracy of 90.87%, 90.09% and 88.37%, which are significantly improved compared with the CNN model in recognition performance. And the ModelC-TL model is deployed in WeChat small program to realize the practical application of "Image recognition garbage category".

**Key words:** deep learning; transfer learning; convolutional neural network; garbage classification; WeChat applet

## 0 引言

近年,随着 5G 时代的到来和移动互联网的再次升级,图像传输和获取的效率显著提高,因此图像逐渐成为信息交流的重要媒介。图像数据具有直观性、高传播效率、高参考价值等特点,便于信息接收者快速理解信息内容。图像的高信息价值使得针对图像处理的技术研究具有重大意义。

图像在大数据分析领域中也有巨大的应用价值。通信技术的快速增长使得互联网图像信息源的质量和数量呈现增长趋势。互联网用户越来越倾向于在淘宝、天猫等电商平台采用图片评论的方式来给商品以反馈,并且由于传送图片的便利性和直观性,用户传输的图像信息量将会不断攀升。用户、企业的行为都给图像处理和分析提供了海量的信息源,使得图像信息的识别和提取成为了大数据领域的核心课题。

从 2013 年开始,图像就是全球用户在社交媒体上分享量最大的内容类型,远远超过文字类型;在搜索引擎领域,图像搜索也已经成为流量仅次于网页搜索的重要搜索项目;在图书情报领域,图像数据也可以作为舆情监控的重要参考依据。随着各领域对大数据信息的依赖,文字分析已经不足以支持数据时代的信息需求,因此数据分析的信息源也逐渐从文本数据转向了图像数据。

为了更高效地处理图像数据，图像识别的技术应运发展，成为了人工智能领域的重点研究技术。图像识别是一种利用计算机对图像进行处理，分析和理解，以识别各种不同模式的目标和对象的技术。已经被广泛应用于各种场景，包括图片搜索、电子商务、人脸识别等。

生活垃圾分类是图像识别的新场景，随着我国越来越多的城市加入到垃圾分类的行列中，垃圾分类俨然成为“新风口”，而其带来的产业“新蓝海”也悄然而至；面对人们“垃圾分类难”的问题，生活垃圾场景的图像识别任务亟待研究与解决。本文将利用图像识别的主流方法深度学习，设计基于 CNN 的垃圾分类模型；并使用迁移学习方法以提高模型的训练效率与识别性能，从而得到高准确率的垃圾图像分类模型；最后将其部署于微信小程序以实现“图识垃圾”应用。

## 1 相关研究

深度学习概念由 Hinton<sup>[1]</sup>等人于 2006 年提出，它是一种含多隐藏层、多感知器的网络结构；可自动学习并提取更高维、更抽象的数据以描述属性类别和特征。随着计算机运算能力和速度的提升，高维矩阵运算得以保障，因而近年来深度学习发展迅速。图像识别领域常用的深度学习模型主要为：深层信念网络（DBN）<sup>[2]</sup>、卷积神经网络（CNN）<sup>[3]</sup>、循环神经网络（RNN）<sup>[4]</sup>、生成式对抗网络（GAN）<sup>[5]</sup>、胶囊网络（CapsNet）<sup>[6]</sup>。相较其他经典网络结构模型，卷积神经网络模型在图像识别领域的表现更为优异，是目前图像识别领域的主流网络结构<sup>[7]</sup>。

### 1.1 卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）是一种包含卷积计算的前馈神经网络。CNN 有许多经典的网络结构：最早的 CNN 模型 LeNet-5 由 LeCun<sup>[3]</sup>等人使用反向传播网络 BP 算法设计，使用 ReLU 为激活函数并用 softmax 回归；AlexNet 由 Krizhevsky<sup>[8]</sup>等改进，其扩大了网络规模，采用

dropout、数据增强等技术减轻过拟合问题；VGG 由 Simonyan<sup>[9]</sup>等改进，使用  $3 \times 3$  小卷积核构建网络并增加网络深度；ResNet 由 Kaiming He<sup>[10]</sup>等提出，引入残差块，将低层特征直接映射至高层网络，改善网络退化问题；DenseNet 由 Huang G<sup>[11]</sup>等提出，将网络中每一层特征图都与其他层建立连接，加强特征传播与特征重用，减少模型参数量，缓解深层网络梯度消失问题。

除了对 CNN 模型结构进行改进，许多学者研究 CNN 网络参数设置优化。Mishkin<sup>[12]</sup>等人根据经典 CNN 结构提出参数选取建议：对于激励函数，可选取有 BN 的 ReLU 非线性函数；对于池化方法，可采用最大池化及平均池化之和；对于学习率，可使用线性学习率衰减方法；对于批量尺寸，建议按 64 的比例选取，例如 128、256 等；对于输入图像大小，实验发现增大图像时模型识别准确率呈线性提升，但同时所需的计算空间呈二次增长。另外，刘万军<sup>[13]</sup>在研究不同池化模型的 CNN 学习性能时实验发现，最大池化法模型的准确率与收敛速度均优于其他单一池化方法；叶林飞<sup>[14]</sup>在研究出血图像检测时对比不同卷积核大小的 CNN 识别性能，实验表明相比大卷积核，全部使用  $3 \times 3$  小卷积核时模型准确率最高。

部分学者研究 CNN 识别性能与数据集的关系，并提出数据集优化方法。对于数据集规模与质量，Mishkin<sup>[12]</sup>等指出当前深度学习性能较大依赖于数据集的大小，训练集规模需达到下限以保证模型性能，且指出数据集质量也将大幅影响识别性能；邢世宏<sup>[15]</sup>等根据 VC 维<sup>[16]</sup>理论估算二分类 $d_{vc}$ 的宽松数据集容量（上限）为  $10000d_{vc}$ ，实验表明实际数据集容量下限约为  $10d_{vc}$ （ $d_{vc}$  为模型参数量）。对于非均衡数据处理<sup>[17]</sup>，Buda<sup>[18]</sup>等表示类间不均衡对分类性能存在不利影响，而过采样方法可有效消除不均衡现象且不会导致过拟合；另有学者<sup>[19]</sup>采用加大低频标签权重值的方法应对。数据增强技术也可用以实现数据集的类间均衡，并有效改善过拟合；褚晶辉<sup>[20]</sup>在研究乳腺肿瘤识别时采用过采样与数据增强处理以均衡数据集；高友文<sup>[21]</sup>对苹果表面病变数据集进行数据增强后，基于 AlexNet 模型的测试准确率由 78% 上升至 83%。

## 1.2 迁移学习

基于深度学习的迁移学习<sup>[22]</sup>是指“运用已存有的知识对不同但相关领域问

题进行求解的一种机器学习方法”，Pan<sup>[23]</sup>等将迁移学习研究内容总结为归纳式迁移学习、直推式迁移学习及无监督式迁移学习三种范型，以及基于实例、特征、参数、关系进行迁移的四种方法。基于实例的迁移学习<sup>[24,25]</sup>指有选择性地从源域迁移部分样本至目标域；基于特征的迁移学习<sup>[24,26]</sup>指学习特定的特征变换，并利用它使源域实例有效映射至目标域；基于参数的迁移学习<sup>[24,27]</sup>指将源域模型参数作为初始值训练目标域模型；基于关系的迁移学习<sup>[24]</sup>指将源域内数据间的关系型知识迁移至目标域。

对于深度学习中的过拟合现象，除了常用的数据增强技术和正则化方法（ $L_2$ 正则化<sup>[28]</sup>、Dropout<sup>[29]</sup>、DropConnect<sup>[30]</sup>及 Maxout<sup>[31]</sup>），基于参数的迁移学习也可有效防止过拟合。

基于参数的迁移学习即基于模型迁移学习，早期研究重心为支持向量机<sup>[28]</sup>；近几年转向深度卷积神经网络，其主要方法为单源有监督迁移学习的 CNN Fine-tuning<sup>[32]</sup>，即利用预训练的源域 CNN 模型进行微调以实现目标域 CNN 训练。WenXuan Liao<sup>[33]</sup>等使用 VGG19 网络训练数据集，并基于该模型参数实现迁移学习，实验表明模型识别准确率从 80.08%提升至 91.41%；褚晶辉<sup>[20]</sup>等基于预训练模型 VGG16 的参数对自定义的 CNN 进行迁移学习及二次迁移，得到准确率为 97.95%的优异性能乳腺肿瘤诊断模型。

另有许多学者<sup>[34,35]</sup>对合适的微调方式展开研究，结果表明对于大规模且与源域样本相似的数据集或小规模不相似数据集，预训练模型的底层卷积可实现良好的基础特征提取，宜采用冻结底层、微调高层的方式进行迁移学习；对于大规模不相似数据集建议全部微调；对于小规模相似数据集可冻结卷积基。

### 1.3 垃圾分类任务研究

目前国内外利用深度学习解决生活垃圾图像分类问题的研究较少。国内学者吕程熙<sup>[36]</sup>分别基于 DNN 和 CNN 训练多分类的垃圾识别模型，实验表明 CNN 模型表现更好，其准确率为 63%；秦斌斌<sup>[37]</sup>基于 Xception 模型并加入迁移学习和部分微调方式对含 5 类材质的垃圾数据集进行训练，测试得到 91%平均精度。

实践部署方面，国内部分专利阐述了使用特征值比对<sup>[38]</sup>、PB 模型<sup>[39]</sup>、卷

积神经网络<sup>[40]</sup>等方法进行垃圾分类并部署为智能垃圾桶的思路；吕文杰<sup>[41]</sup>基于 Inception-v3 模型开发垃圾分类应用程序；王爽<sup>[42]</sup>提出垃圾分类微信小程序的设计框架。

综上，针对生活垃圾图像分类任务，尚缺乏逻辑性标签分类以及高覆盖率、高准确率、理论与实践结合的生活垃圾分类模型；因此本文将基于深度卷积神经网络与迁移学习训练生活垃圾图像分类模型，并使其与微信小程序相结合，实现“图识垃圾”的实际应用，在理论与实践上拓展图像识别领域的研究。

## 2 实验原理

### 2.1 非均衡数据处理与数据增强

数据不平衡现象普遍存在于自然图像分类问题中。对于生活垃圾图像分类问题，主要表现为类间不平衡，即由各类垃圾的产生比例不同、图像获取难度不同而使样本图集中部分品类垃圾偏少。非均衡数据会导致训练时算法和网络更多地提取大样本量类别的图片特征信息，而忽略少样本量类别，最终难以达到普遍理想的分类准确率。针对非均衡数据处理<sup>[43]</sup>通常有优化样本数据集、采样重建样本集、集成学习等。本文采用过采样与数据增强方法多步构建均衡数据集。过采样方法的优势在于结果较为稳定，且在保留数据集全部原有信息的基础上，重复少数类别的样本；过采样也可通过数据增强代替实现类间均衡的目的。数据增强方法可有效扩大数据集规模，降低 CNN 过拟合可能性，提高模型泛化能力。数据增强可通过对原始数据集样本进行旋转、翻转等几何变换，获得图像副本以扩充数据集；该方法增加冗余信息，以提高网络泛化能力；并使神经网络减少对低相关特征的关注，增强对高相关性、高一致性特征的关注与学习，从根本上提升整体性能。

### 2.2 卷积神经网络

卷积神经网络是深度学习的代表算法之一，随着计算机算力的提高与大数

据时代的到来,深度卷积神经网络在计算机视觉任务上表现出优越性能。

CNN 基本结构<sup>[44]</sup>由输入层、卷积层、激活层、池化层、全连接层和输出层构成。卷积神经网络层级之间通过不同的神经节点建立连接,逐层传递输入信息,经历多次卷积和池化过程的特征提取、特征映射后,由全连接层综合各种局部特征,最后进行分类输出。

其核心是卷积计算层。卷积层有多个特征图,每个特征图有多个神经元,输出特征图的每个神经元通过一组权值与上一层特征图进行局部连接<sup>[45]</sup>,并将加权求和得到的局部加权值加上偏置值,传递给一个非线性函数,如 ReLU 函数;得到各个神经元的输出值;然后下采样得到各个特征映射层,且同一个映射面上的神经元共享权值<sup>[46]</sup>。CNN 的图像处理中,图像数据以二维矩阵形式输入神经网络,记原始图像为  $A$ 。用  $X^l$  表示卷积神经网络第  $l$  层的特征图,  $X^0 = A$ 。假设  $X^l$  是卷积层,卷积过程可描述为:

$$X^l = f(X^{l-1} \otimes W^l + b^l) \quad (1)$$

式中:  $W^l$  表示第  $l$  层卷积核的权值向量,  $b^l$  为第  $l$  层的偏置向量,运算符  $\otimes$  表示卷积核与上一层特征图进行卷积操作,  $f(\cdot)$  表示非线性的激励函数。

卷积层得到特征映射后,通常由紧连着的下采样层对特征图进行下采样<sup>[47]</sup>。下采样时,卷积层存在  $N$  个输出特征图,假设  $X_j^l$  是一个第  $l$  层的特征图,一个子采样过程可描述为:

$$X_j^l = f(w_j^l \text{down}(X_j^{l-1}) + b_j^l) \quad (2)$$

式中:  $\text{down}(\cdot)$  表示下采样函数,  $w_j^l$  和  $b_j^l$  分别为相应特征图可训练的权重和偏置向量。

假设  $X^l$  是下采样层,则整体下采样过程可表示为:

$$X^l = \text{subsampling}(X^{l-1}) \quad (3)$$

下采样可实现特征降维,减少学习参数量,防止过拟合;并一定程度保持特征不变性。

## 2.3 迁移学习

迁移学习旨在将已有知识迁移至新环境中以解决该目标域的问题。在深度

学习领域存在许多可提取准确、细微特征的神经网络模型，我们可以利用这些成熟的模型作为预训练模型进行迁移学习。

迁移学习的关键概念为域和任务。域  $D$  分为源域  $D_s$  和目标域  $D_t$ ，一个域分别由一个特征空间  $X$  和特征空间上的边际概率  $P(X)$  组成，其中  $X = x_1, x_2, \dots, x_n \in X$ 。一个域的任务分别表示为  $T_t = \{Y_t, f_t(\cdot)\}$  和  $T_s = \{Y_s, f_s(\cdot)\}$ ，其中  $D_s \neq T_s$  或  $D_t \neq T_t$ 。

迁移学习的目的是完成知识在相关领域的迁移。迁移学习可利用  $D_s$  和  $T_s$  的已有知识，获得目标函数，甚至提高目标函数  $f(\cdot)$  在目标域  $D_t$  的分类预测性能。

### 3 实验模型设计

#### 3.1 CNN 模型构建

本节研究基于卷积神经网络的生活垃圾图像分类算法。本文设计的 CNN 模型由三个卷积块、三个池化层、四个 dropout 层、一个全连接层及一个 softmax 回归层组成，使用 Adam 优化器优化神经网络。模型架构如下图 1 所示：

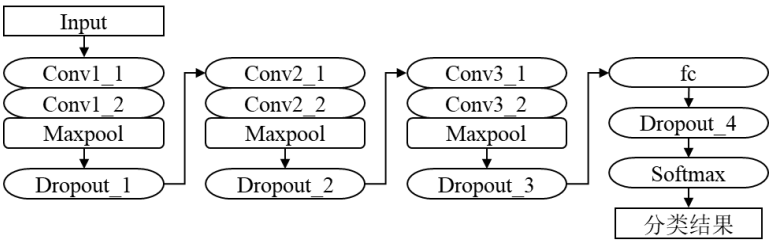


图 1 CNN 模型架构图

Figure 1 CNN model architecture diagram

##### 3.1.1 卷积与池化

卷积层可从多角度逐层提取图像数据特征图的局部特征，并将特征映射划分至若干区域进行池化操作，池化的本质即采样。由图可知，该卷积结构包含 3 个卷积块，每个卷积块均含 2 个卷积层，共 6 个卷积层；每个卷积块后含对应的 ReLU 激活函数<sup>[48]</sup>（未展示）；且每个卷积块后连接一个池化层，采用最大值池化法（Maxpooling）。

所有卷积核大小为  $3 \times 3$ ，步长为 1；池化层的窗口大小也为  $3 \times 3$ ，步长同



为 3；因此经一次池化操作后，输出特征维度降至输入特征维度的三分之一。卷积层的卷积核个数设置随卷积块依此增多，分别为 32、64、128 个；随着卷积核个数增多，各卷积块提取特征的数量增加，使学习特征逐渐从简单的图像物体边缘信息向图像高层抽象特征过度。

### 3.1.2 dropout 处理

使用 dropout 方法可在训练样本较少时有效减轻过拟合现象，甚至得到更强健的模型。Dropout 过程为在深度学习网络训练时按一定概率  $p$  随机地将部分神经元暂时从网络中丢弃，暂时忽略其权重。对比原神经网络公式，使用 dropout 处理的公式为：

$$r_j^l \sim \text{Bernouli}(p) \quad (4)$$

$$\tilde{X}^l = r^l * X^l \quad (5)$$

$$X^l = f(\tilde{X}^{l-1} \otimes W^l + b^l) \quad (6)$$

模型共四次使用 dropout 处理，三次分别于三个卷积块后，其丢弃比例均为 0.25，最后一次为全连接层与输出层之间，其概率为 0.5；其中第一个丢弃层 dropout\_1 的输出张量继续输入后几个丢弃层中。

### 3.1.3 全连接层与输出层

全连接层接收第三个卷积块经 dropout 处理后的张量数据。设置全连接层的神经元数为 4608，处理输入的张量，使其转换维度至一维，运算并经历激活函数后，再次采用 dropout 处理。

输出层采用 softmax 分类器，根据不同训练样本类别数确定神经节点个数。

Softmax 计算公式如下：

$$\sigma_i(X) = \frac{\exp(w_i^T \cdot x_i)}{\sum_{j=1}^m \exp(w_j^T \cdot x_j)} \quad \begin{matrix} i = 1, 2, \dots, m \\ x = [x_1, x_2, \dots, x_n]^T \in R^n \end{matrix} \quad (7)$$

式中： $\sigma_i(X)$  可视为输入数据  $x$  属于类别  $i$  的概率， $w$  为模型参数。

## 3.2 使用迁移学习的模型构建

本节研究利用迁移学习，将一种已在大规模带注释数据集 ImageNet 上训练

的模型迁移到目标主题“生活垃圾图像分类”任务上。迁移学习的大致流程<sup>[49]</sup>为：首先利用相关领域的大规模数据集对特定网络模型的随机初始化参数进行训练；并利用该预训练模型对目标领域的数据进行特征提取；最后利用提取的特征与目标域数据完成特定分类任务的 CNN 训练。

本文选取合适的深层神经网络模型 DenseNet 系列模型作为预训练模型，利用模型在 ImageNet 自然图像数据集上预先训练好的权重进行训练；为模型设计全局平均池化（global average pooling）操作与 dropout 处理；增设两个全连接层与批量标准化处理层（batch normalization），其中全连接层神经元个数分别为 2048 和 1024，并结合使用 L2 参数正则项；最终输出层 softmax 基于学习特征进行分类预测。为研究适合 GC 图集的高效迁移学习训练方式，分别采取两种形式训练：方法 1 为冻结 DenseNet 模型卷积基的冻结方式；方法 2 为冻结底层并训练最后一个卷积块的微调方式。迁移学习模型结构如图 2 所示：

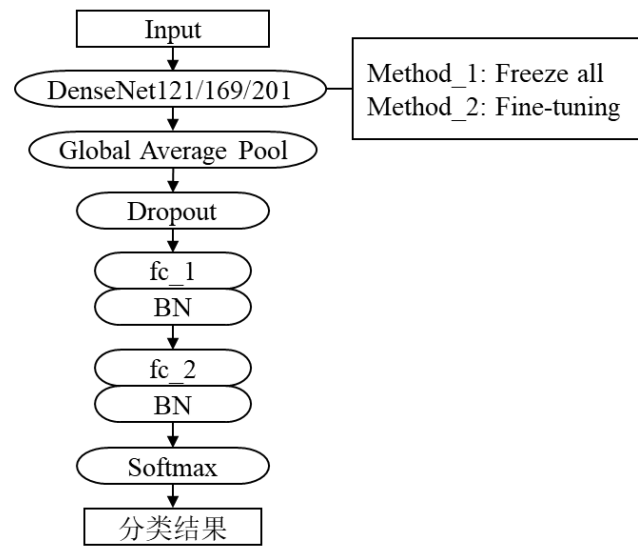


图 2 使用迁移学习的模型架构

Figure 2 Model architecture of transfer learning

4 实验分析

4.1 构建生活垃圾分类数据库 GC

目前尚未存在专用于垃圾图像分类的数据库，为验证本文算法的有效性及时解决生活垃圾图像分类问题，将构建一个常见生活垃圾图像数据库 GC。由于生活垃圾分类主题本身具有专业性与知识储备需求，因此数据库分类标签的确定至关重要，其标签分类的逻辑合理性将最终决定垃圾图像分类任务的识别准确率与实用性。

本文依照上海市垃圾分类规定，以垃圾材质及外形作为参考因素，总结：常见生活垃圾大类包括纸类、塑料类、金属类、玻璃类、砖瓦陶瓷、织物、木材、电池、尘土、残渣类、餐饮垃圾、外卖垃圾等；垃圾分类标准为可回收垃圾、干垃圾、湿垃圾、有毒有害垃圾。为探寻生活垃圾分类场景的适宜标签分类标准，分别构建三种标签分类标准：标准 A 为垃圾材质分类，共 8 类；标准 B 为垃圾外形分类，共 6 类；标准 C 为垃圾综合分类，汇总共 16 类。生活垃圾分类应用时，可结合标准 A 与标准 B，或单凭标准 C 判断垃圾类别。三种标签分类标准如下表所示：

表 1 生活垃圾图像分类标准

Table 1 Standards for classification of domestic garbage images

| 标准 A |                   |           | 标准 B |                   |          |
|------|-------------------|-----------|------|-------------------|----------|
| 编号   | 标签                | 备注        | 编号   | 标签                | 备注       |
| A0   | 0paper            | 纸         | B0   | 0texture          | 材质       |
| A1   | 1plastic          | 塑料        | B1   | 1texture-fragment | 材质碎片     |
| A2   | 2metal            | 金属        | B2   | 2product          | 制品+容器    |
| A3   | 3glass            | 玻璃        | B3   | 3packing          | 包装       |
| A4   | 4tile             | 砖瓦陶瓷      | B4   | 4food-packing     | 食品包装     |
| A5   | 5fabric           | 织物        | B5   | 5tableware        | 一次性餐具餐盒  |
| A6   | 6wood             | 木         |      |                   |          |
| A7   | 7food             | 湿垃圾       |      |                   |          |
| 标准 C |                   |           |      |                   |          |
| 编号   | 标签                | 备注        | 编号   | 标签                | 备注       |
| C0   | 0paper_recyclable | 纸制垃圾-可回收  | C8   | 8glass            | 玻璃-可回收   |
| C1   | 1paper_residual   | 纸制垃圾-干垃圾  | C9   | 9tile             | 砖瓦陶瓷-干垃圾 |
| C2   | 2paperpacking     | 纸包装-可回收   | C10  | Afabric           | 织物-可回收   |
| C3   | 3plastic          | 塑料材料-可回收  | C11  | Bwood             | 木制品-可回收  |
| C4   | 4plastic_bottle   | 塑料瓶-可回收/害 | C12  | Cdust             | 尘土-干垃圾   |

|    |                    |          |     |               |          |
|----|--------------------|----------|-----|---------------|----------|
| C5 | 5plastic_container | 其他塑料容器   | C13 | Dfood         | 湿垃圾      |
| C6 | 6plastic_packing   | 塑料包装-干垃圾 | C14 | Etakeaway_box | 餐饮外带-干垃圾 |
| C7 | 7metal             | 金属-可回收   | C15 | Ffood_packing | 食品包装-干垃圾 |

结合以上标准与各垃圾分类属性构建细分项，如纸类可细分为可回收纸类、不可回收纸类（纸巾、票据等干垃圾）、一次性纸餐具（干垃圾）、纸类容器（可回收）、食品纸包装（干垃圾）等。考虑 GC 数据库构建的完备性，本文共提出垃圾细类 39 项，预计覆盖 90%左右的生活垃圾品类。

生活垃圾图像样本依据细分项进行收集，主要数据来源为谷歌图库、百度图库、沙沙野等图像资源平台及自行拍摄采集的垃圾图像。样本数据质量良好，图像分辨率均在  $300 \times 300$  以上，且均为日常场景下拍摄的自然图片；其数据含大量特征与噪声，可保证训练时数据良好的冗余度。各细类收集 300–2000 张不等图像样本，共得 21667 张样本。与 CNN 与迁移学习模型的参数量比较可知，该数据集规模远超下限、满足大规模条件。将细类图像样本分别依据三种标签分类标准归档存储，以此构建 GC-A，GC-B，GC-C 三个 GC 数据库子图集。

分别对三个 GC 图集进行相同的数据预处理。考虑数据库样本平均大小、训练效果及计算机性能，将所有图像统一缩至  $192 \times 192$  像素；且对本图集数据整体做标准化处理（Normalization）；并按 7: 3 的比例将数据集划分为训练集和测试集。另外，由于 GC 图集均存在类间数据不平衡现象，因此在整体标准化前，各图集视情况进行非均衡数据处理及数据增强（水平翻转及旋转 20 度）。

## 4.2 训练模型

实验在配备 AMD Ryzen 5 处理器，内存为 20G 的 Windows 10 操作系统下进行，CNN 训练利用深度学习框架 TensorFlow<sup>[50]</sup>及深度学习库 Keras<sup>[51]</sup>。训练时采用 Adam 为优化算法，学习率默认为 0.001；CNN 训练时迭代次数（epochs）设为 80 至 120 次，迁移学习时则设为 10 至 20 次。

首先，利用 GC 数据库原始图集对 CNN 模型进行训练并测试，以 GC-A 图集为例，数据批量（batch\_size）大小设置为 64；训练及验证得到损失函数变化如下图 1 所示：

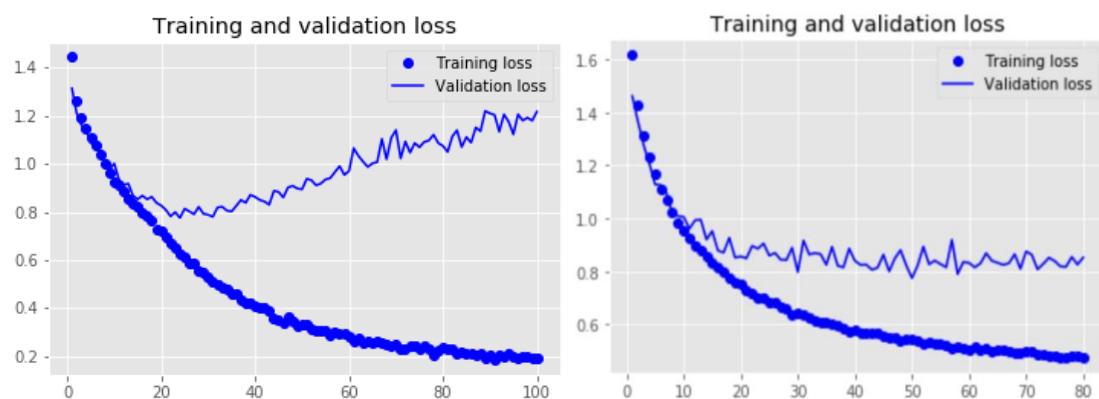


图 1 GC-A 原始图集 (a) 与数据增强 GC-A 图集 (b) 的损失函数变化曲线

Figure 1 GC-A loss function curve of the original dataset (a) and the data augmented dataset (b)

由图可知，在原始图集样本数量较少且神经网络模型较复杂的情况下，训练出现过拟合现象，随着训练迭代次数增加，模型在训练集上的分类错误仍然逐渐减小，但在测试集上的测试分类错误却呈增大趋势。另使用数据增强后的图集 GC-A 进行实验，并适当增大 batch 值为 100，训练得损失函数变化如图 2 所示，模型训练有效、拟合状况良好。

## 4.3 实验结果

### 4.3.1 非均衡数据处理与数据增强的有效性验证

为验证训练集样本数量及分布对图像分类 CNN 性能的影响，以 GC-A 图集为例，分别使用原始图集与数据非均衡处理并数据增强后的图集训练深度卷积神经网络。

非均衡数据处理使各类别样本量不同程度地增减，使其类间分布趋于均衡；大多数类别的识别性能均存在不等幅提升，模型宏平均 Macro Average 明显提升，模型整体识别性能良好。表 2 记录了 CNN 对 GC-A 图集的实验结果。由表可知，与原始图集训练相比，用非均衡数据处理后的数据集训练 CNN，其平均分类正确率提高了 0.74%；说明非均衡数据处理及数据增强能改善分类效果。

表 2 非均衡数据处理与数据增强前后的 GC-A 图集实验结果对比

Table 2 Comparison of GC-A dataset experimental results before and after unbalanced data processing

|         | 原始图集   | 非均衡数据处理后 |
|---------|--------|----------|
| 样本数量/个  | 21667  | 28736    |
| 平均分类正确率 | 0.7394 | 0.7468   |

已知其他条件相同时，训练集样本数量扩增、样本类间分布趋于均衡，利于 CNN 模型分类性能提升。本实验采取同样的非均衡数据处理与数据增强方法对 GC 数据库其余数据集进行 CNN 训练，结果如下表 3 所示。可知多分类模型准确率一定程度与其类别数相关，类别数少的分类任务易获得更高识别准确率，如基于 CNN 网络的 GC-B 图集实验准确率远高于 GC-C 图集。

表 3 GC 数据库 A/B/C 图集 CNN 实验结果

Table 3 CNN experimental results of GC database A / B / C dataset

| 模型名称     | ModelA-CNN | ModelB-CNN | ModelC-CNN |
|----------|------------|------------|------------|
| 数据集      | GC-A       | GC-B       | GC-C       |
| 类别数      | 8          | 6          | 16         |
| 样本数量     | 28736      | 27322      | 33076      |
| val_acc  | 0.7468     | 0.7655     | 0.6683     |
| val_loss | 0.8827     | 0.8382     | 1.1912     |

#### 4.3.2 迁移学习优异性验证

为验证使用迁移学习方法的优异性，以 GC-C 图集为例，采用多种方法进行实验，得模型性能对比如下表 4。结果显示，以 DenseNet121 为预训练模型使用迁移学习方法的网络性能相较 CNN 大幅提升，其平均分类正确率提高了 13.8%；其中，使用微调迁移学习方式(Method\_2)的网络性能远优于冻结方式(Method\_1)；且采取防止过拟合措施优化后，模型平均分类正确率高达 86.08%。表明，对一定规模数据集进行多分类实验时，可用较深层网络，并使用迁移学习方法以提升模型分类性能；对于与原始图集较相似的大数据集，如本文的 GC 数据库各图集，宜采用部分冻结，最后一个卷积块参与训练的微调

方式进行迁移学习。且增加 dropout 层、BatchNormalization 层与使用正则化参数可有效防止过拟合与提升网络性能。

表 4 GC-C 图集迁移学习的不同方法性能比较

| Table 4 Performance comparison of GC-C dataset by different transfer learning methods |         |          |              |  |
|---|---------|----------|--------------|--|
| Transfer Learning Method  | val_acc | val_loss | Train_time   |  |
| method_1 based on DenseNet121   | 0.5814  | 1.7746   | 6788s/epoch  |  |
| method_2 based on DenseNet121   | 0.8063  | 0.7122   | 7310s/epoch  |  |
| method_2 based on DenseNet121 (Prevent Overfitting)                                   | 0.8608  | 0.6895   | 7662s/epoch  |  |
| method_2 based on DenseNet169 (Prevent Overfitting)                                   | 0.8700  | 0.6545   | 9870s/epoch  |  |
| method_2 based on DenseNet201 (Prevent Overfitting)                                   | 0.8802  | 0.6372   | 12878s/epoch |  |

如表 4，对比 DenseNet 三个预训练模型实验结果，模型分类正确率随预训练模型深度增大而提升，loss 值下降，但相应训练时间成本递增；可知同算法的 DenseNet 预训练模型结构越深，其迁移学习得到的模型分类效果越好，其中 DenseNet201 取得最高测试正确率 88.02%。测试该模型各类别图像的分类效果，其最佳分类与最差分类展示如下图 2：



图 2 GC-C 图集以 DenseNet201 为预训练模型的迁移学习实验结果

Figure 2 GC-C database experimental results of transfer learning with DenseNet201

GC-C 图集中分类效果最好的五类为玻璃类、尘土类、纸类（不可回收）、木质类、塑料瓶类垃圾，均达 92% 以上的分类准确率。分类效果最差的类别是

“餐饮外带垃圾”，其准确率为 69%，易被错分为“其他塑料容器垃圾”。部分类别被误分类的主要原因是 GC-C 图集中另存在与这些类别垃圾高相似性的垃圾图像，如其他塑料容器类别中部分图像与餐饮外带垃圾中的塑料餐盒相似，塑料袋垃圾与部分食品包装垃圾相似，可回收纸板与塑料板材相似，这些相似度较高的垃圾在垃圾的材质与外形上具有重合，致使垃圾图像具有相似的色彩、形状、纹理等特征，对最后的分类结果产生干扰。针对此误分类问题，对数据集质量优化处理，进一步筛查归类细分标签，如严格剔除 C6plastic\_container 中的一次性塑料容器，并在 C15takeaway\_box 图像中选择沾有油污或携带残渣的外卖餐盒，而非干净的餐盒图像。优化数据集质量后重复实验，可得平均分类正确率提升了 0.35%；其中分类较佳的类别识别性能变化不大，C15takeaway\_box 类准确率提升至 74%，其余较差类别准确率均有不同程度的提升。表明，增强标签细项的分类合理性与优化数据集质量有利于进一步提升模型识别性能。

本文利用 CNN 与迁移学习方法所得 GC-C 数据库最佳实验结果与前人研究结果比较如下：

表 5 GC 数据库 A/B/C 三图集实验结果与前人研究成果比较

Table 5 Comparison of the experimental results of the three datasets of the GC database A / B / C with previous research results

| Model Name | Classification task / Data set                        | Algorithm                     | Accuracy |
|------------|---|-------------------------------|----------|
| -          | Multi-class-10 items <sup>[36]</sup>                  | CNN                           | 63%      |
| -          | Two-class-recyclable / non-recyclable <sup>[36]</sup> | CNN                           | 80%      |
| Xception   | Multi-class-5 <sup>[37]</sup>                         | CNN                           | 91%      |
| ModelA-CNN | Multi-class-8-GC-A                                    | CNN                           | 74.68%   |
| ModelB-CNN | Multi-class-6-GC-B                                    | CNN                           | 76.55%   |
| ModelC-CNN | Multi-class-16-GC-C                                   | CNN                           | 66.83%   |
| ModelA-TL  | Multi-class-8-GC-A                                    | Transfer Learning-DenseNet201 | 90.87%   |
| ModelB-TL  | Multi-class-6-GC-B                                    | Transfer Learning-DenseNet201 | 90.09%   |
| ModelC-TL  | Multi-class-16-GC-C                                   | Transfer Learning-DenseNet201 | 88.37%   |



对比前人算法中基于深度神经网络的垃圾图像分类任务，本实验的多分类模型垃圾种类覆盖面更全、识别性能更优、综合效用更佳。对比 CNN 模型，使用迁移学习之后的模型利用预训练的权重，使训练效率提高；且识别准确率均大幅提升，尤其是 GC-C 的准确率提升了 21.54%。实验说明，使用迁移学习的深度卷积神经网络算法识别性能优异，能够有效实现生活垃圾图像分类。

## 5 模型应用

本文基于迁移学习的各垃圾图像识别模型性能良好，为进一步探究其应用价值，故选择将 ModelC-TL 模型部署应用于微信小程序端以实现“图识垃圾”应用；该生活垃圾综合分类模型可区分 16 种垃圾类群，具备 88.37% 的识别准确率。本章利用小程序云开发环境及 TensorFlow.js 插件实现该模型在小程序端的实际应用。

### 5.1 小程序云开发

小程序云开发是腾讯方提供的原生 serverless 云服务。借助云平台，小程序开发者无需经历搭建或购置服务器、注册及备案域名等传统 web 应用开发工作，可直接将代码运行所需的服务器部署与运营交由云端平台托管；由此降低开发门槛，并使开发者更专注于业务逻辑部署，提高开发效率。

小程序云开发提供数据库、存储器和云函数三大基础能力支持。云数据库是文档型数据库，可包含多个集合，每个集合为一个 JSON 数组，数据库 API 具有数据增加、删除、修改、查询等基本操作能力；云存储是云端提供的免费存储空间，开发者可带权限上传、下载等文件管理操作；云函数是管理端操作接口，实质是一段云端可运行的代码，可上传部署自行编写的后端代码或调用第三方 API 进行操作。开发者只需在小程序端调用原生接口，即可以实现云数据库、云存储、云函数的操作。

### 5.2 小程序架构和流程

“图识垃圾”小程序借助微信开发者工具开发实现。小程序搭建使用的语

言有 WXML、WXSS 以及 JavaScript 语言。应用开发体现了三层架构设计思想<sup>[52]</sup>：视图层，主要为小程序界面呈现，由 WXML 和 WXSS 定义组件及其样式而实现；逻辑层，为功能实现的核心，利用小程序原生接口实现，由 JavaScript 编写；云端后台，包含数据、模型、文件等资源的存储，通过原生接口、云函数及云端 SDK 等实现，JavaScript 编写。

该小程序垃圾识别流程图如下图 3 所示

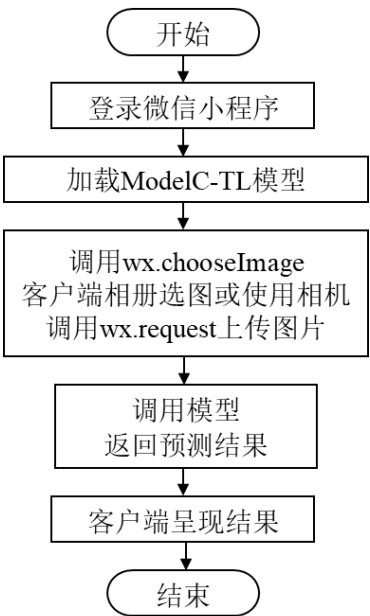


图 3 微信小程序“图识垃圾”流程图

Figure 3 Flowchart of WeChat Mini Program Garbage Image Recognition

### 5.3 模型部署

“图识垃圾”应用基于 ModelC-TL 模型完成识别，小程序利用 TensorFlow.js 插件完成模型的部署，其关键操作有：转换模型、加载模型和预测。

原 ModelC-TL 模型基于 keras 框架完成训练，以 HDF5 文件形式保存其结构及权重。现需使用 tensorflowjs 中的 tfjs-converter 进行模型转换，生成小程序中可用的 web 格式模型，即 json 拓扑文件和 bin 二进制权重文件。配置云开发环境，将转换后的 ModelC-TL 模型文件全部上传至云存储的同一文件夹中待用。加载运行所需的第三方库，如 tfjs-core 和 tfjs-layers 等；并通过读取云端模型的 json 文件存储路径实现 ModelC-TL 模型加载。将客户端上传的图片作为输

入数据，进行数据预处理，将图像裁剪、缩放至 192\*192 像素大小，此过程可使用 tidy 函数以节约内存、提高执行效率；再利用 ModelC-TL 模型进行预测；最终输出结果并与客户端交互。

### 5.4 运行效果

“图识垃圾”小程序运行效果如下，图 4(a)为主界面；点击“拍照识别”按钮上传待识别的生活垃圾自然图像，如图(b)所示，以使用过的纸巾（干垃圾）为例；经部署的 ModelC-TL 模型识别后预测该图像属于类别 1paper\_residual，故输出结果“纸质垃圾-干垃圾”以及相应图标，完成界面交互，如图(c)所示。

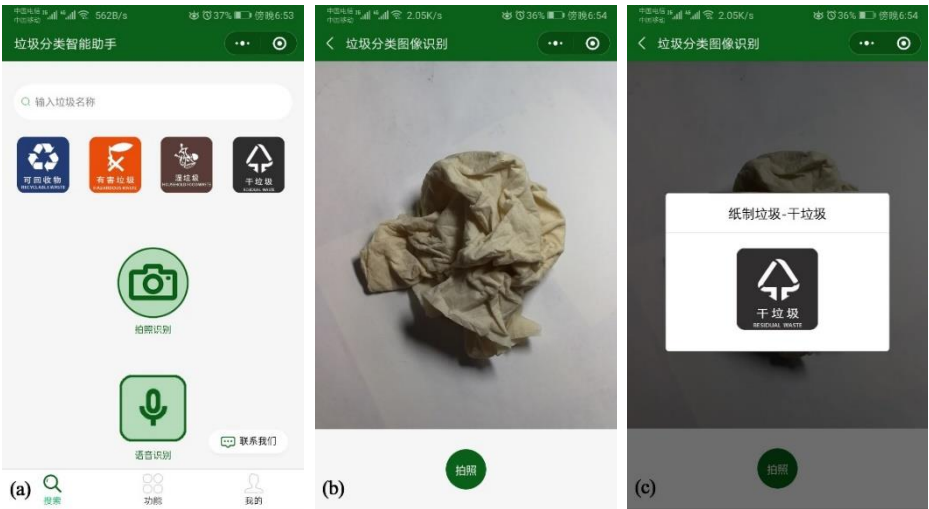


图 4 图识垃圾微信小程序运行效果

Figure 4 Image of WeChat Mini Program *Garbage Image Recognition*

本章实现了 ModelC-TL 模型在微信小程序端的部署，通过“图识垃圾”应用进一步证明了该模型的可用性及其优异的识别性能，体现了生活垃圾分类模型的应用价值。

## 6 结语

本文构建了高覆盖率的常见生活垃圾图像数据库 GC，并创建三种图集标签分类标准。研究了基于 CNN 的垃圾图像分类模型，并在此基础上使用迁移学习，提升模型的识别性能及训练效率。并探究了数据集与迁移学习方法对模型性能的影响，结果表明非均衡数据处理与数据增强等数据集质量优化对模型性能有提升作用；加入合适的迁移学习方式可以大幅提升模型分类性能，对于 GC 数据库各图集，其规模较大且与预训练模型的原始图集相似，宜采用微调方式进行训练。此外，实现了 ModelC-TL 模型在微信小程序端的部署，证明了模型的可用性与识别性能。

GC 数据库的标签分类标准、深度卷积网络模型、使用迁移学习的方法及模型在小程序端的应用均可为生活垃圾图像分类任务提供参考依据，后续研究可进一步优化模型算法，寻求识别性能优异的更轻量级模型；同时增加“电子设备”等标签，扩充各标签样本量，实现加大数据库覆盖率、优化 GC 数据集质量。

## 参考文献:

- [1] Hinton G E, Salakhutdinov R R.Reducing the dimensionality of data with neural networks[J].Science, 2006, 313 (5786) : 504.
- [2] Hinton G E, Osindero S, Teh Y W.A fast learning algorithm for deep belief nets[J].Neural Computation, 2006, 18 (18) : 1527-1554.
- [3] Lecun Y, Bottou L, Bengio Y, et al.Gradient-based learning applied to document recognition[J].Proceedings of IEEE, 1998, 86 (11) : 2278-2324.
- [4] 杨丽,吴雨茜,王俊丽,刘义理.循环神经网络研究综述[J].计算机应用,2018,38(S2):1-6+26.
- [5] Goodfellow I J, Pouget-Abadie J, Mirza M, et al.Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems.Montreal, Canada: MIT Press, 2014: 2672-2680.
- [6] Sabour S, Frosst N E, Hinton G.Dynamic routing between capsules[EB/OL]. (2017) [2019].<https://arxiv.org/abs/1710.09829>.
- [7] 郑远攀, 李广阳. 深度学习在图像识别中的应用研究综述[J] 计算机工程与应用, 2019,55(12):20-36.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1106 — 1114.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2015-11-04]. <http://www.robots.ox.ac.uk:5000/~vgg/publications/2015/Simonyan15/simonyan15.Pdf>.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[EB/OL]. [2016-01-04]. [https://www.researchgate.net/publication/286512696\\_Deep\\_Residual\\_Learning\\_for\\_Image\\_Recognition](https://www.researchgate.net/publication/286512696_Deep_Residual_Learning_for_Image_Recognition).
- [11] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [12] Mishkin D, Sergievskiy N, Matas J. Systematic evaluation of convolution neural network advances on the Imagenet. arXiv:1606.02228v2, 2016
- [13] 刘万军,梁雪剑,曲海成.不同池化模型的卷积神经网络学习性能研究[J].中国图象图形学

报,2016,21(09):1178-1190.

- [14] 叶林飞,徐建省,孔德昱,赵璇,张龙,王星远.基于卷积神经网络的出血图像检测[J].软件,2019,40(02):78-82.
- [15] 邢世宏,施闻明,任荟洁.不同数据集容量下的卷积神经网络图像识别性能[J].舰船科学技术,2019,41(21):188-193.
- [16] VAPNIK V N, CHERVONENKIS A Y. On the uniform convergence of relative frequencies of events to their probabilities[J]. Theory of Probability & Its Applications, 1971, 17(2): 264–280.
- [17] He H B, Garcia E A. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263–1284
- [18] Mateusz Buda,Atsuto Maki,Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks[J]. Elsevier Ltd,2018,106.
- [19] 陈磊,李菲菲,陈虬.基于迁移学习与权重支持向量机的图像多标签标注算法[J].电子科技,2020,33(03):12-16.
- [20] 褚晶辉, 吴泽葵等. 基于迁移学习和深度卷积神经网络的乳腺肿瘤诊断系统[J] 激光与光电子学进展, 2017(08):1-7.
- [21] 高友文,周本君,胡晓飞.基于数据增强的卷积神经网络图像识别研究[J].计算机技术与发展,2018,28(08):62-65.
- [22] Long M S, Wang J M, Cao Y, Sun J G, Yu P S. Deep learning of transferable representation for scalable domain adaptation. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(8): 2027–2040
- [23] Pan S J, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345–1359
- [24] 张雪松,庄严,闫飞,王伟.基于迁移学习的类别级物体识别与检测研究与进展[J].自动化学报,2019,45(07):1224-1243.
- [25] Dai W Y, Yang Q, Xue G R, Yu Y. Boosting for transfer learning. In: Proceedings of the 24th International Conference on Machine Learning. Corvalis, USA: ACM, 2007. 193–200
- [26] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Sydney, Australia: ACM, 2006. 120–128

- [27] Li X. Regularized Adaptation: Theory, Algorithms and Applications [Ph.D. thesis], University of Washington, USA, 2007
- [28] 吕国豪,罗四维,黄雅平,蒋欣兰.基于卷积神经网络的正则化方法[J].计算机研究与发展,2014,51(09):1891-1900.
- [29] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958
- [30] Wan L, Zeiler M, Zhang S, et al. Regularization of neural networks using Dropconnect//Proceedings of the International Conference on Machine Learning. Atlanta, USA, 2013: 058-1066
- [31] Goodfellow I J, Warde-farley D. Mirza M, et al. Maxout networks//Proceedings of the International Conference on Machine Learning. Atlanta, USA, 2013: 1319-1327
- [32] Tajbakhsh N, Shin J Y, Gurudu S R, Hurst R T, Kendall C B, Gotway M B, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Transactions on Medical Imaging, 2016, 35(5): 1299-1312
- [33] Wen-Xuan Liao, Ping He, et. Automatic Identification of Breast Ultrasound Image Based on Supervised Block-Based Region Segmentation Algorithm and Features Combination Migration Deep Learning Model[J] IEEE Journal of Biomedical and Health Informatics, 2019.
- [34] 方向. 预训练模型迁移学习[EB/OL] (2018-11-02).  
<https://yq.aliyun.com/articles/663228?spm=a2c4e.11153940.0.0.1fd873ab8EdXEc&type=2>
- [35] 贾宇霞,樊帅昌,易晓梅.基于显著性增强和迁移学习的鱼类识别研究[J].渔业现代化,2020,47(01):38-46.
- [36] 吕程熙. 基于深度学习实现自动垃圾分类[J] 软件开发, 2019(24):36-38.
- [37] 秦斌斌,何级.基于卷积神经网络的垃圾分类研究[J].无线通信技术,2019,28(03):51-56.
- [38] 西安航空学院. 一种基于图像处理的垃圾分类处理系统: 中国, 110683240:A[P]. 2020.01.14
- [39] 石家庄邮电职业技术学院. 一种基于视觉识别的自动分类垃圾桶及分类方法: 中国, 11189565.8[P]. 2020.03.06
- [40] 山东科技大学. 垃圾智能分类与管理系统及垃圾分类识别: 中国, 110893956:A[P].

2020.03.20

- [41] 吕文杰,魏孝虎,陈众孚,童灏,马燕.基于卷积神经网络的生活垃圾自动分类软件的实现[J].电脑知识与技术,2020,16(05):203-204.
- [42] 王爽. 微信小程序在垃圾分类中的应用研究[J] 信息与电脑, 2019(22):66-68.
- [43] Japkowicz N. Learning from imbalanced data sets: a comparison of various strategies [C] AAAI Workshop on Learning from Imbalanced Data Sets, 2000, 68:10-15.
- [44] Andrej Karpathy. Cs231n: Convolutional neural networks for visual recognition[EB/OL] 2016. <http://cs231n.github.io/>.
- [45] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015,512(7553):436-444
- [46] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998,86(11): 2278-2324
- [47] ZEILER M D, FERGUS R. Stochastic pooling for regularization of deep convolutional neural networks[EB/OL] [2016-01-11] <http://www.matthewzeiler.com/pubs/iclr2013/iclr2013.pdf>.
- [48] Hahnloser R H, Sarpeshkar R, Mahowald M A, et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit [J] Nature, 2000, 405 ( 6789 ) : 947-951.
- [49] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J] 计算机应用, 2016[09]:2508-2515.
- [50] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large scale machine learning on heterogeneous systems[EB/OL] (2015-12-12). <http://download.tensorflow.org/paper/whitepaper2015.pdf>.
- [51] François Chollet ,et al. Keras: The Python Deep Learning library [EB/OL] (2016-09-18). <https://keras.io/>
- [52] 董傲通,文俊浩.基于小程序·云开发的实验室设备管理系统的设计与实现[J]实验技术与管理,2019,36(10):282-284+288.