

基于不同模型的迁移学习垃圾分类

蒲晓海 周 丽

四川大学锦城学院 计算机与软件学院 四川 成都 611731

【摘要】随着我国垃圾分类的进一步推进,如何正确地做好垃圾的分类是关键所在。在垃圾分类问题上可以利用机器视觉和传感器等技术帮助提高分类的正确率。本文基于三个不同的预训练模型 GoogLeNet、ResNet18 和 MobileNetV2 在迁移学习垃圾分类任务上的结果对比分析,实验结果表明:采用 ResNet18 预训练模型进行迁移学习的垃圾分类任务效果最好。

【关键词】迁移学习;垃圾分类;GoogLeNet;ResNet18;MobileNetV2

1 引言

人工垃圾分类投放是垃圾处理的第一步,但其分类结果往往不够准确。据一项在北京、广州等 17 个城市小区调查发现居民在回答垃圾分类问题上的准确率仅为 27.7%^[1]。垃圾处理厂是处理海量垃圾的最后一步,然而,我国的垃圾处理厂大都采取人工流水线分拣的方法进行垃圾分类,效率较低,分出的可回收垃圾仅占小部分,大部分垃圾最终的处理方式都是填埋,这样就大大减少了资源的利用率。随着深度学习技术在图像识别分类上的发展,通过相机捕获垃圾图片,然后采用卷积神经网络自动提取图像特征并进行分类,从而可以让机器自动进行垃圾分拣,有效地提高垃圾分类的正确率。本文采用不同模型进行迁移学习来解决垃圾分类任务,通过实验对比,分析不同模型进行迁移学习的效率和准确率。

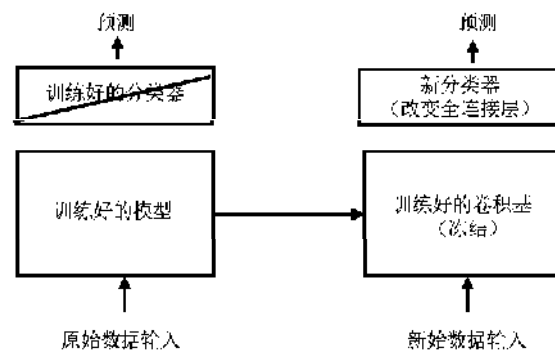
2 迁移学习

迁移学习 (Transfer Learning) 作为机器学习方法的一种,主要思想是将过去已有的知识即源域 (source domain) 迁移到有关联的、要学习的新知识即目标域 (target domain)^[2]。本文基于预训练好的模型

GoogLeNet、ResNet18 和 MobileNetV2 迁移学习到垃圾分类上。在基于预训练模型的迁移学习上有特征提取和微调 2 种方法,本文采用了特征提取方法,如图 1 所示,通过使用预训练好的模型提取特征——这三种模型均通过大型图像数据集 ImageNet 进行了很好的学习训练,得到了图像的大量特征;然后仅训练最后的全连

接层,用较少的迭代次数就可以完成学习,解决垃圾分类的问题。

图 1: 特征提取方法



一般来说,增加网络的深度和宽度可以提高网络的

性能,但这会造成参数量太大、过拟合等问题。为解决

3 GoogLeNet 模型结构

图 2: GoogLeNet 模型具体结构^[3]

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	124M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	164K	129M
inception (4b)		14×14×512	2	160	112	224	24	64	64	451K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	248	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	22	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

此类问题, Google 研究人员提出了 22 层神经网络模型 GoogLeNet^[3], 如图 2 所示, 其在 2014 年的 ImageNet 分类竞赛中获得冠军。该模型最大特点是构建了如图 3 所示的 9 个 Inception 模块, 将不同大小的卷积核组合在一起可以融合不同的特征信息, 在增加网络的深度和宽度的同时可以提高网络的鲁棒性, 此外还通过加入 1*1 的卷积核, 降低参数量, 减少了计算量, 从而提高模型训练速率。基于此模型的迁移学习在苹果缺陷检测^[4]、风电机组检测^[5] 等方面都具有较高的准确率。

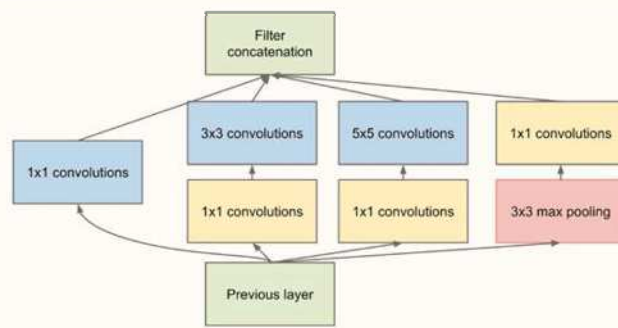


图 3: Inception 模块^[3]

4 ResNet18 模型结构

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

图 4: ResNet 模型具体结构^[6]

ResNet(Residual Network)^[6] 是 2015 年由何恺明等人提出的, 其结构如图 4 所示, 同年该模型也在 ImageNet 分类竞赛中获得冠军。随着神经网络的不断加深其效果反而不好即产生退化问题。该模型通过 Residual 模块如图 5 所示, 除了用到 2 个 3*3 的卷积, 还通过“shortcut connections”将输入 x 送到输出。使得原来的目标值从 $H(x)=x$ 变为 $H(x)=F(x)+x$ 。ResNet 相当于将学习一个完整的输出的目的变为学习目标值 $H(x)$ 和 x 的差值 $F(x) = H(x)-x$ (也叫残差)。所以, 之后的训练目标就是要将残差结果趋于 0。这种结构使得浅层与深层网络产生恒等映射, 从而很好地解决退化问题。因此该模型作者最终通过残差结构堆叠搭建了超 1000 层的网络结构, 错误率最终也比较低。本文所使用的模型为 ResNet18, 18 表示其网络层有 18 层。

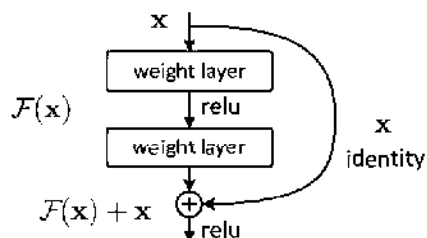


图 5: Residual 模块^[6]

5 MobileNetV2 模型结构

MobileNet^[7] 由 Google 团队在 2017 年提出, 主要用于移动端的轻量级卷积神经网络。其参考了深度分离卷积, 深度分离卷积核心是将传统卷积分解为一个深度卷积和一个逐点卷积 (1*1 卷积)^[8]。MobileNetV2^[9] 在 MobileNet 的基础上进行了改进, 使得模型准确率得到了提高, 模型也变得更小, 其结构如图 6 所示。其借鉴 ResNet 的残差结构提出倒残差结构, 同时将激活函数 ReLU 改为 ReLU6。在加强网络结构性能时也可以减少特征信息的丢失。

图 6: MobileNetV2 结构^[9] (t 表示通道扩张倍数, c 表示输出通道数, n 表示重复次数, s 表示步长)

6 实验设计与结果分析

本实验所使用数据集来源于华为垃圾分类挑战赛数据集。一共有 14802 张图片, 分为 4 大类, 40 小类, 每张图片对应 40 小类中的一种, 图像种类及数量见表 1。将数据集按照 8:2 划分为训练集和测试集 (训练集有 11824 张, 测试集有 2978 张)。为了实验的准确性, 把图片的大小都调整到 224*224, 并使用图像增强技术将这些图片进行缩放、旋转、翻转、改变对比度。

使用 Anaconda3 搭建了开发和运行环境, 硬件设备为 Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz 2.30 GHz、16GB 内存、显卡为 NVIDIA GeForce GTX 1050 Ti, 采用了 64 位 Windows 系统、Python3.6 Pytorch 1.7.1, 使用 Jupyter Notebook 作为开发环境。实验采

用了 GPU 进行加速。

通过冻结 GoogLeNet、ResNet18 和 MobileNetV2 模型除最后一层全连接层之外的所有网络权重将参数固定。然后由于本实验需要将图像分为 40 个类别，而上

述模型基于 ImageNet 的 1000 分类，所以要将模型最后一层“1000 fully connected”全连接层替换为具有随机权重的“40 fully connected”全连接层，然后只训练该层。

表 1: 数据集图像种类与数量

图像类别 1	图像类别 2	图像数量	总计	图像类别 1	图像类别 2	图像数量	总计
其他垃圾	0 一次性快餐盒	242	1652	20 快递纸袋	226	8611	
	1 污损塑料	370		21 插头电线	657		
	2 烟蒂	279		22 旧衣服	375		
	3 牙签	85		23 易拉罐	309		
	4 破碎花盆及碟碗	387		24 枕头	318		
厨余垃圾	5 竹筷	289	3389	25 毛绒玩具	550		
	6 剩饭剩菜	395		26 洗发水瓶	351		
	7 大骨头	362		27 玻璃杯	536		
	8 水果果皮	300		28 皮鞋	302		
	9 水果果肉	309		29 砧板	416		
	10 茶叶渣	387		30 纸板箱	321		
	11 菜叶菜根	736		31 调料瓶	446		
	12 蛋壳	331		32 酒瓶	280		
可回收物	13 鱼骨	409		33 金属食品罐	322		
	14 充电宝	357		34 锅	395		
	15 包	419		35 食用油桶	351		
	16 化妆品瓶	352		36 饮料瓶	265		
	17 塑料玩具	309		37 干电池	322		
	18 塑料碗盆	362		38 软膏	391		1150
	19 塑料衣架	312		39 过期药物	437		

本实验设计了一组对比实验，实验中采用相同的参数：学习率为 0.001，批处理大小为 64，迭代次数为 10。分别对数据集进行训练和测试。分别采用 GoogLeNet、ResNet18、MobileNetV2 预训练模型进行迁移学习训练集和测试集的损失值曲线和准确率曲线如图 7、图 8、图 9 所示（其中蓝色表示训练集，

红色表示测试集）。通过实验结果表 2 可以看到采用 ResNet18 进行迁移模型的准确率较高为 73.50%。3 个实验在平均单次迭代中训练和测试中所需时间分别为 680, 675, 660s。综合比较，ResNet18 模型更适合迁移学习的垃圾分类任务。

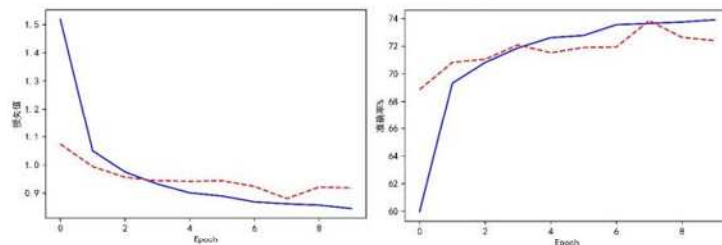


图 7 (左边为损失值曲线，右边为准确率曲线，下同)

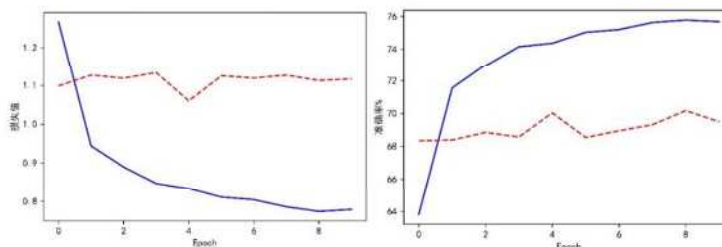


图 8

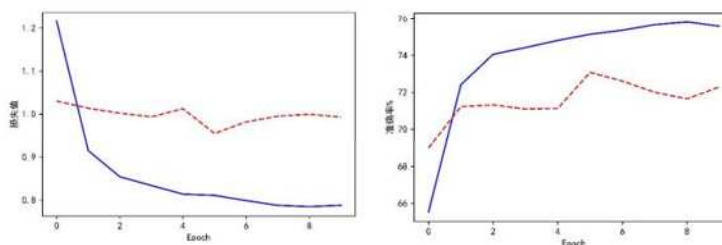


图 9

表 2 实验结果

实验结果	GoogLeNet	ResNet18	MobileNetV2
训练准确率 %	73.91	75.68	75.58
测试准确率 %	72.41	73.50	72.29
训练损失值	0.85	0.78	0.79
测试损失值	0.92	1.12	0.99
平均单次迭代用时/s	680	675	665

总结

本实验通过采用 GoogLeNet、ResNet18、MobileNetV2 预训练模型迁移学习进行 40 类别的垃圾分类。可以看到 ResNet18 实验结果最为理想。因此，基于迁移学习的垃圾分类是可行的，但后续仍需通过增加垃圾种类、改变模型中的参数和采用迁移学习的微调方法进一步提高垃圾分类效率和准确率，以更好贴近生活实际进行垃圾分类。垃圾分拣自动化有利于提高垃圾分类的效率和准确度，通过对可回收垃圾的准确识别，可以大大增加可回收资源化利用率，让生态环境可持续健康的发展。

【参考文献】

[1] 陈耘. 抓垃圾分类应精准施策 [J]. 宁波经济 (财经视点), 2019(08):44-45.

[2] Pan S J, Qiang Y. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359.

[3] Szegedy C, Wei L, Jia Y, et al. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[4] 薛勇, 王立扬, 张瑜, 沈群. 基于 GoogLeNet 深度迁移学习的苹果缺陷检测方法 [J]. 农业机械学报, 2020, 51(07):30-35.

[5] 徐一鸣, 张娟, 刘成成, 顾菊平, 潘高超. 迁移学习模式下基于 GoogLeNet 网络的风电机组视觉检测 [J]. 计算机科学, 2019, 46(05):260-265.

[6] [1] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.

[8] 陈宇超, 卞晓晓. 基于机器视觉与深度学习的医疗垃圾分类系统 [J]. 电脑编程技巧与维护, 2019(05):108-110.

[9] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.