鲁东大学学报(自然科学版)
Journal of Ludong University(Natural Science Edition) 2021,37(2) : 139—145

基于迁移学习的无监督细粒度图像分类模型

韩天宇[1],张利锋[1],王喜亮[2]

(1.鲁东大学 信息与电气工程学院,山东 烟台 264039; 2.上海通用汽车有限公司 南厂冲压车间,山东 烟台 264006)

摘要: 图像分类是计算机视觉领域的一个重要研究分支,普通图像分类关注主类对象差异性的判别,而细粒度图像分类则重点研究主类下不同子类的区分。考虑到较小的子类间差异(只在某个局部上有细微差异) 和由于拍摄角度、背景、姿态等因素导致的较大的子类内差异,细粒度图像识别成为一项颇具挑战的任务。为了缓解细粒度图像数据集难以获取和扩充的问题,本文提出了一种基于迁移学习和孪生网络实现的无监督细粒度图像分类模型。该模型采用两阶段训练策略,第一阶段采用有监督方式预训练基础网络,使模型快速学习到细粒度图像的通用特征; 第二阶段通过搜索引擎扩充数据集,设计孪生网络并采用无监督方式微调模型以实现知识迁移。实验表明,孪生网络模型在 Stanford Car 和 Aircraft 数据集上能够获得较好的性能表现。

关键词: 迁移学习; 孪生网络; 无监督学习; 细粒度图像分类

中图分类号: TP391.413 **文献标志码**: A **文章编号**: 1673-8020(2021) 02-0139-07

在深度学习研究的推动下,计算机视觉领域取得了巨大的进步。得益于 CIFAR[1]、ImageNet[2]、COCO[3] 等通用数据集的支持,普通图像分类已经取得显著进展。但是,许多应用需要视觉模型能够区分主类中不同子类之间的细微差异(例如,对奥迪 A4 和 A6、波音 747 和 777 进行分类) ,上述应用场景的问题属于细粒度分类问题,它重点关注如何区分主类内的不同子类对象。由于细粒度图像具有子类间差异小且子类内差异大的性质,为了获得好的性能表现,细粒度分类模型必须具有区分子类间细微差异的能力。

已有的细粒度分类研究,给出了 3 种研究路径: 1) 利用对象的局部显著特征设计深度网络; 2) 使用特征编码的深度学习方法; 3) 基于迁移学习的模型方法。前两种主要是设计深度网络来捕获细粒度图像的特征表示,但细粒度图像数据集的有限性限制了模型性能的进一步改善。因此,许多研究人员尝试使用迁移学习,针对性地获取细粒度图像精细化特征区域知识,以提高细粒度分类的准确性。

利用局部特征构建深度学习模型往往需要对图像的局部特征边界框进行标注。Zhang 等[4] 提出了一种基于候选部件(如车灯或轮胎) 的 R-CNN 模型,使用候选部件训练网络以提取局部特征,从而缓解了细粒度图像类间变化小的限制。Lin 等[5] 提出了一种 LAC(定位、对齐和分类) 框架来对齐和定位具有区分度的部件,其中使用了 VLF(阈函数) 自适应地折衷分类和对齐的误差,从而更新部件定位坐标。但是,以上利用局部特征的方法通常需要大量的部件标注数据,耗时费力,往往还需要专家知识,很难大规模扩展应用。因此,许多研究[6—7] 针对缺少数据标注的限制,利用注意力机制[8] 来发现图像中部件的位置,从而缓解了标注数据的压力。

使用特征编码的方法设计深度模型,对细粒度图像进行特征升维,在高维特征空间中寻找细粒度图像的可区分特征。双线性模型(Bilinear-CNN) [9] 将两个卷积神经网络(convolutional neural network,CNN) [10] 输出的特征进行外积运算,从而实现对细粒度图像的高维空间编码。双线性模型具有特征的细节捕获能力,显著提高了细粒度图像分类的准确性。但是由于双线性模型容易引发维度爆炸问题,模型维度的数量级通常在数十万到几百万,并不适用于资源受限的设备

收稿日期: 2021-02-09; 修回日期: 2021-03-05

第一作者简介: 韩天宇(1994—) ,男,山东潍坊人,硕士研究生,研究方向为图像处理与模式识别。E-mail: teeyohan@ 163.com

通信作者简介: 张利锋(1977—) ,男,宁夏银川人,讲师,博士,研究方向为分布式计算、计算机视觉。E-mail: lifengzhang@ ldu.edu.cn

中。针对该问题,许多工作[11—12]使用张量绘制[13]来缓解双线性模型中维度爆炸问题。

以上两种方法最主要的挑战在于标注数据集困难与计算资源耗费大。因此,有研究者引入迁移学习,寻求一种不需要大量的数据标注且资源花费小的模型。注意到细粒度数据集与大型主流数据集中有许多重叠的图像,因此,使用大型数据集的预训练模型有可能会减少细粒度图像的需求。Cui 等[14]提出了一种使用推土机距离的算法来估计不同领域间的相似性,并证实了迁移学习可以从使用源域的预训练模型中受益。

针对细粒度图像子类间差异小且子类内差异大的性质以及细粒度数据集难以获取和扩充的问题,本文借鉴迁移学习中领域适应研究的思路,期望设计一种特殊的网络结构,寻求通过扩充数据集来提升模型性能。

1 孪生网络模型

CNN 是一种强大的基于深度学习技术的端到端分类方法[10],它可以直接从图像中学习特征表示,以表示图像的多样性。本研究利用两个 CNN 网络结合迁移学习构造两阶段孪生网络架构,实现无监督领域迁移能力。本文提出的孪生网络模型以及两阶段训练流程如图 1 所示。

第一阶段为训练 Base-ResNet,首先采用 ResNet[15]作为基础骨干网络,选择大型通用数据集 ImageNet 对其进行预训练,这可以使网络学习到一般图像的通用知识,然后修改 ResNet 最后的分类层,使其符合细粒度数据集的分类数,最后选择细粒度数据集对其进行有监督训练; 第二阶段为微调 Grey-ResNet,首先使用搜索引擎对选择的细粒度数据集进行图像检索,将检索到的图像组成细粒度检索数据集,然后将 Base-ResNet 的第一个卷积层进行修改,使其符合灰度图像的通道数,与 Base-ResNet 联合形成孪生网络,最后选择细粒度检索数据集并执行灰度化对其进行无监督微调。

在第一阶段,模型学习到细粒度图像空间的一般性特征; 在第二阶段,模型会更注重细粒度图像的纹理、轮廓等非颜色细节特征。下面将详细介绍基于迁移学习孪生网络的两阶段训练策略。

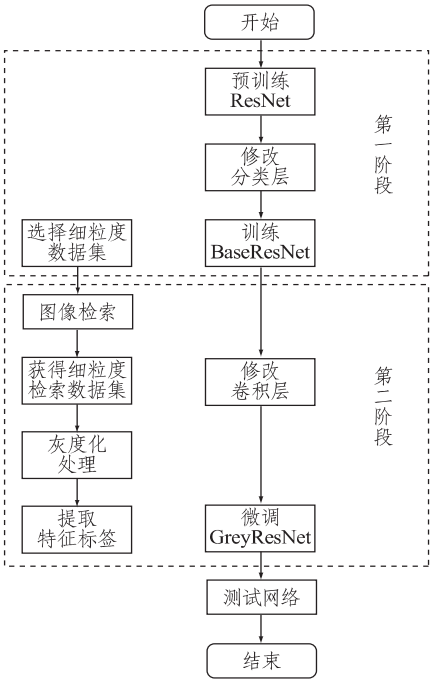

图 1 孪生网络两阶段训练流程

Fig.1 Two-stage training process of twin network

2 两阶段训练策略

2.1 训练 Base-ResNet

为了降低模型对大规模数据集的依赖,应该避免从零开始训练网络,因此,本文使用在 ImageNet 上预训练好的 ResNet-50 作为 Base-ResNet 的基础骨干网络,这样可以从 ImageNet 中学到一般图像的现有特征并对其进行调整,以完成新的细粒度任务。为了匹配细粒度数据集与通用数据集在空间维度上的差异,将 ResNet-50 最后分类层进行修改,作为 Base-ResNet。第一阶段训练过程如图 2 所示。

在细粒度数据集上训练 Base-ResNet 的过程中,将图像和标签按批送入网络,使用交叉熵损失对 Base-ResNet 的所有层进行训练。常规的迁移训练方法可能仅仅对网络的最后一层或者最后数层进行训练,而本文选择对网络中的所有层进行调整,目的是使网络对细粒度数据集更加敏感,能够更准确地捕获细粒度数据集的特征空间编码,有利于第二阶段策略的性能提升。

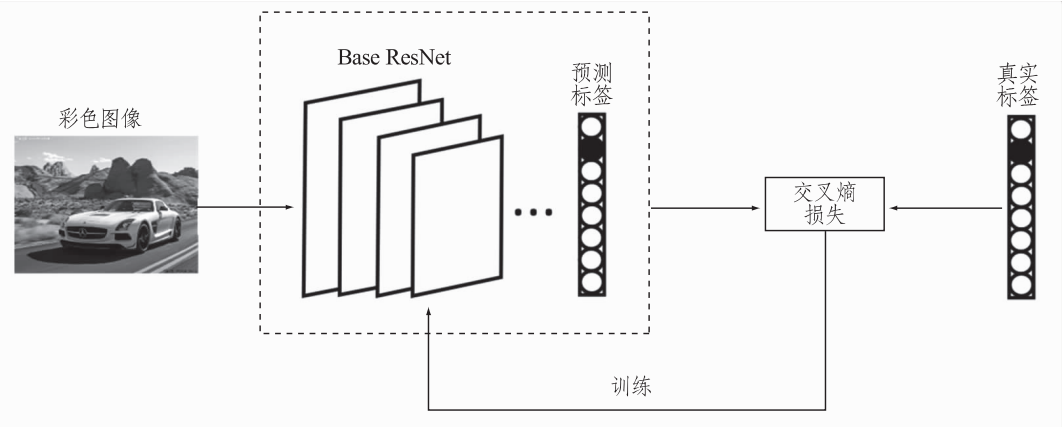

图 2　预训练 Base-ResNet

Fig.2 Pretraining of Base-ResNet

2.2　微调 Grey-ResNet

2.2.1　检索数据集获取

深度网络的学习性能极大程度上依赖于数据集的丰富程度,为了让深度网络“见多识广”,最有效的手段往往是为模型扩充数据集。但细粒度数据集的扩展往往需要专家参与,成本昂贵。因此,本文通过已有数据集的标签使用搜索引擎进行图像检索,并将检索到的图像作为第二阶段微调训练的检索数据集来使用。

本文分别对 Stanford Car[16] 和 Aircraft[17] 数据集进行了图像检索,直接选用 196 类 Car 和 100 类 Aircraft 的标签,针对每类标签返回了 100 张检索结果。结果显示: 本文采用的图像检索效果较优,仅存在少量的重复且大多数图像都与原始标签语义相关。实验中发现,若使用同一检索关键词但增加检索数量,会导致检索结果中的图像出现大量重复,这会严重增加检索数据集的冗余性。重复的图像虽然不会降低孪生网络模型性能,但会增加训练时间。如果想要进一步扩充检索数据集,可以使用标签的同义词进行图像检索。

2.2.2　孪生网络的实现

孪生网络模型的训练以第一阶段训练好的 Base-ResNet 为基础并在检索数据集上做适应性微调训练,检索到的新数据能扩充模型的认知,从而丰富深度网络的表示空间。为了确保模型能重点关注细粒度图像的关键特征而降低颜色导致的偏度,比如,同一型号的汽车可能具有不同的颜色,但颜色并不会为分类提供帮助反而会干扰分类器导致误判。孪生网络对图像进行了灰度化处理,迫使网络更加关注于纹理、轮廓等非颜色细节特征。孪生网络结构如图 3 所示。

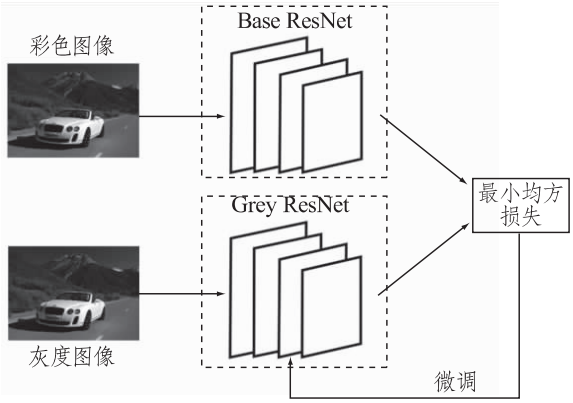

图 3　孪生网络

Fig.3 Twin network

具体实现上,将 Base-ResNet 的第一个卷积层进行修改,使其匹配灰度图像的通道数,修改后的网络称为 Grey-ResNet。接下来将 Base-ResNet 与 Grey-ResNet 进行联合形成孪生网络,将检索数据集中同一图像的彩色和灰度图像送入 Base-ResNet 和 Grey-ResNet 中。由于在第一阶段预训练模型 Base-ResNet 已经对细粒度数据集敏感,且已经能够很好地完成分类任务,因此选择两个网络输出向量的最小均方误差进行无监督微调训练。本文分别尝试了对 Grey-ResNet 的第一个卷积层与全部层进行微调,微调单层的情况下,只有少量的参数进行更新,因此该过程能快速收敛,不会占用太多算力; 而微调全层的情况下,网络会对检索数据集更加敏感。无监督训练带来的好处是不需要对扩充后的检索图像进行数据标注,这为扩充数据集提供了新的通用途径和方法。

3 领域适应分析

迁移学习是机器学习的子研究领域,它专注于复用已有问题的解决模型,并将其利用在其它不同但相关的问题上,领域适应是迁移学习的重要分支。数学上,给定源域 $D_s=\{X_s,f_s(X)\}$ 和源任务 T_s,目标域 $D_t=\{X_t,f_t(X)\}$ 和目标任务 T_t,领域适应旨在目标域 D_t 不同于源域 D_s 但目标任务 T_t 与源任务 T_s 相同的条件下,通过源域 D_s 和源任务 T_s 所获取的知识来帮助模型解决在目标域 D_t 上目标任务 T_t 的预测函数 $f_t(\cdot)$。下面针对本文提出的两阶段策略进行分析。

在第一阶段训练过程中,源域为 $D_s=\{X_s,f_s(X)\}$,其中 X_s 为细粒度图像和标签,$f_s(\cdot)$ 为Base-ResNet,源任务 T_s 为细粒度图像分类。使用彩色图像和标签对 Base-ResNet 进行全网络训练,训练完成后的 Base-ResNet 已经具备初步的解决细粒度分类任务的能力。对于两阶段训练策略来说,该阶段是对目标任务进行预训练,目的是为了使网络学习到细粒度图像的通用特征,更新全网络的参数是为了使网络更全面的捕获数据集的空间特征。

在第二阶段微调过程中,目标域为 $D_t=\{X_t,f_t(X)\}$,其中 X_t 为检索数据集中的彩色图像和灰色图像,$f_t(\cdot)$ 为Base-ResNet和Grey-ResNet联合形成的孪生网络,目标任务 T_t 为更加注重灰度特征的细粒度图像分类。由于无监督带来的好处,不需要图像标注,但是 Grey-ResNet 参数的更新依赖于 Base-ResNet,虽然这在一定程度上增加了模型的复杂度,但是孪生网络的设计会为无监督训练提升网络表达能力提供新的思路。Grey-ResNet 是可拔插的,不侵入已有网络,可以用于其它有监督训练网络的数据扩充和性能提升。同样地,送入第二阶段训练的图像也可以根据数据集的特性进行特殊处理,灰度化只是其中的一种针对性处理方式而不是绝对方式。

从预训练到微调,本文运用迁移学习领域适应的方法来提高细粒度图像识别准确率,迁移了细粒度图像的通用特征,在考虑颜色特征的权衡中突出了纹理、轮廓等细节特征在细粒度图像中的重要性。

4 实验结果分析

4.1 实验配置

本实验涉及的数据集包括: 大型通用数据集 ImageNet 针对分类任务的子集 ISLVRC、细粒度数据集 Stanford Car 和 Aircraft、使用搜索引擎进行图像检索获得的检索数据集 Retrieved Car 与 Retrieved Aircraft,如表1所示。

表1 模型使用的数据集

Tab.1 The datasets used in the model

数据集	类别数	训练样本数	测试样本数	总样本数
ISLVRC	1000	1 281 167	50 000	1 331 167
Stanford Car	196	12 657	3528	16 185
Aircraft	100	7968	2032	10 000
Retrieved Car	196	9911	1960	11 871
Retrieved Aircraft	100	4515	1000	5515

在第一阶段,对细粒度数据集进行分割,得到训练集和测试集。在第二阶段,使用搜索引擎进行图像检索,针对每个细粒度标签返回了100张检索结果,去除明显的噪声图像后,组成检索数据集,同样也对检索数据集进行了分割。

根据两阶段训练策略,分别对 Base-ResNet 和 Grey-ResNet 进行了训练和微调,两个阶段都采用随机裁剪和随机水平翻转进行数据增强,模型参数设置如表2所示。

表2 实验参数设置

Tab.2 Parameter settings

网络	监督类型	准则函数	优化器	批次大小	训练轮数	学习率
Base-ResNet	有监督	交叉熵	SGD,M=0.9	16	200	0.005,G=90%
Grey-ResNet	无监督	最小均方	SGD,M=0.9	8	100	0.005,G=90%

第一阶段准则函数使用交叉熵函数,优化器选择动量为0.9的随机梯度下降(SGD,Momentum=0.9),批次大小为16(Batchsize=16),一共训练200轮(Epochs=200),初始学习率为0.005,每20轮调整为原来的90%(LearningRate=0.005,Gamma=0.9)。

第二阶段准则函数使用最小均方函数,优化器同样选择动量为0.9随机梯度下降,批次大小

为 8,一共训练 100 轮,初始学习率为 0.005,每 10 轮调整为原来的 90%。

4.2　实验结果

本节评估两阶段孪生模型在细粒度数据集上的性能表现。本文主要针对分类准确率、模型大小、测试时间三个方面与主流方法[9—12,18—19]进行对比,验证孪生网络模型的性能。

4.2.1　模型准确率

本文分别在 Stanford Car 和 Aircraft 以及与之对应的检索数据集 Retrieved Car 和 Retrieved Aircraft 数据集上进行了实验,实验结果如表 3~4 所示。

表 3　细粒度数据集上的准确率

Tab.3 Accuracy on fine-grained datasets

模型	发表会议	Stanford Car/%	Aircraft/%
Grey-ResNet(微调单层)	—	88.5	**84.0**
Grey-ResNet(微调全层)	—	**90.6**	83.7
FGR without PA[18]	CVPR 2015	92.6	—
HSnet[19]	CVPR 2017	93.9	—
RA-CNN[6]	CVPR 2017	92.5	88.2
MA-CNN[7]	ICCV 2017	92.8	89.9
PA-CNN[20]	IEEE TIP 2020	93.3	91.0
Bilinear-CNN[9]	ICCV 2015	91.3	84.1
Compact Bilinear-CNN[11]	CVPR 2016	91.2	84.1
Low-rank Bilinear-CNN[12]	CVPR 2017	90.9	87.3
DBTNet[21]	NIPS 2019	94.1	91.2
MOMN[22]	IEEE TIP 2020	92.8	90.4

表 4　检索数据集上的准确率

Tab.4 Accuracy on retrieved datasets

模型	Retrieved Car/%	Retrieved Aircraft/%
Grey-ResNet(微调单层)	83.5	62.5
Grey-ResNet(微调全层)	80.7	68.4

由表 3 可知,与利用特征编码的方法相比,在 Stanford Car 上,Grey-ResNet 的性能基本可以与 Low-rank Bilinear-CNN 保持持平; 在 Aircraft 上,Grey-ResNet 的性能基本与 Bilinear-CNN、Compact Bilinear-CNN 保持持平。与 Bilinear-CNN 相比,同样需要两个卷积网络,Grey-ResNet 将单个网络的输出向量作为特征标签,而后者将两个网络的输出向量进行池化外积,因此 Grey-ResNet 的模型复杂度更小。此外,与利用局部特征的方法相比,在 Stanford Car 上,Grey-ResNet 的性能稍差 1%~2%。但是 FGR without PA 需要部件边界框,HSnet 需要部件锚点,且均采用有监督的训练方式,严格要求图像标签,而采用无监督方式的 Grey-ResNet 不要求图像标签,大大放宽了数据来源的限制。

由表 4 可知,Grey-ResNet 在 Retrieved Car 和 Retrieved Aircraft 两个检索数据集上也能保持较好的性能表现。细粒度数据集中的图像经过领域专家严格筛选,去除了大量的噪声数据,而检索到的图像更能代表实际中的图像,这说明 Grey-ResNet 具有较强的适应性。

进一步分析,Grey-ResNet 的性能可能与下列因素有关: 1) 检索数据集规模限制。在实验中,选取了细粒度数据集每个标签的 100 个检索结果。在检索关键词不变的条件下,若继续增加检索结果则会大大增加图像的重复率,重复的图像虽然不会降低 Grey-ResNet 的性能,但会增加训练时间。若想继续扩充数据集,可以考虑使用同义词进行检索。2) 检索数据集质量限制。在实验中,对检索数据去除了明显的噪声图像(例如,设计图和驾驶舱内的图像) ,但仍然可能存在语义不相关的图像干扰。3) 微调处理方式。汽车或飞机等类型实体可能存在多种颜色版本,但它们却属于同一子类。在这种情况下,颜色特征成为迷惑视觉模型误判分类的重要因素。针对这种观察,孪生网络模型在微调前对图像进行了灰度化,而针对颜色敏感的数据集,则可以对数据集的颜色做扩充,来提高实体的丰富性。4) 特征标签提取。在孪生网络结构中,Base-ResNet 相当于特征标签提取器,虽然已经具备较好的完成细粒度分类任务的能力,但仍然存在误差。由于 Grey-ResNet 是可拔插的,不会侵入原有网络,因此,可以将其它有监督训练网络作为 Base-ResNet 进行

更换,减少标签误差,进一步提升性能。

基于以上分析,本文认为两阶段孪生网络模型在 Stanford Car 和 Aircraft 以及与之对应的 Retrieved Car 和 Retrieved Aircraft 上都可以达到较好的性能表现,同时又可以作为一种通用的数据扩充和性能提升方法提供思路。

4.2.2 模型复杂度

实验基于 Python 语言和 Pytorch 框架实现,硬件环境为 2×8 GB 2666 MHz 内存,Intel(R) Core(TM) i5-8500 @3.00 GHz CPU,NVIDIA GeForce GTX 1060 6 GB GPU,模型复杂度如表 5 所示。

表 5 模型复杂度

Tab.5 Model complexity

网络	显存占用/GB	图像大小	测试时间
Base-ResNet	2.3	224×224	26.1 ms/张
Grey-ResNet(微调单层)	1.4	224×224	33.0 ms/对
Grey-ResNet(微调全层)	1.6	224×224	39.5 ms/对

模型大小上,孪生网络在 GPU 上进行训练时,第一阶段训练 Base-ResNet,显存占用约为 2.3 GB。第二阶段微调 Grey-ResNet 时,微调单层的情况下显存使用约为 1.4 GB,微调全层的情况下显存占用约为 1.6 GB。

测试时间上,第一阶段训练 Base-ResNet 时,输入图像随机裁剪成 224×224 大小,输入单张图像需要消耗时间约为 26.1 ms。第二阶段微调 Grey-ResNet 时,输入图像随机裁剪成 224×224 的彩色和灰度图像,在微调单层的情况下输入一对图像(彩色和灰度图像)需要消耗时间约为 33.0 ms,在微调全层的情况下输入一对图像需要消耗时间约为 39.5 ms。

5 结论

本文提出了一种基于搜索引擎数据扩充的无监督迁移学习模型,用于细粒度图像的分类应用。设计两阶段训练策略,首先基于细粒度数据集有监督训练 Base-ResNet,然后利用迁移学习结合无监督方式对 Grey-ResNet 进行适应性微调。考虑到细粒度图像在纹理、轮廓等非颜色特征与颜色特征之间的关系与权重,在微调训练前对新扩充的图像进行了灰度化处理。经过实验分析发现,孪生网络模型能够提高在细粒度数据集上的准确率,说明其确实能够捕获纹理、轮廓等细节特征,同时又作为一种通用的数据扩充和性能提升方法提供了思路。

参考文献:

[1] KRIZHEVSKY A. Learning multiple layers of features from tiny images [R]. Toronto: University of Toronto, 2009.

[2] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database [C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.

[3] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context [C]// European Conference on Computer Vision, 2014: 740-755.

[4] ZHANG N, DONAHUE J, GIRSHICK R, et al. Part-based R-CNNs for fine-grained category detection [C]// European Conference on Computer Vision, 2014: 834-849.

[5] LIN D, SHEN X Y, LU C W, et al. Deep lac: deep localization, alignment and classification for fine-grained recognition [C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1666-1674.

[6] FU J L, ZHENG H L, MEI T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4438-4446.

[7] ZHENG H L, FU J L, MEI T, et al. Learning multi-attention convolutional neural network for fine-grained image recognition [C]// 2017 IEEE International Conference on Computer Vision, 2017: 5209-5217.

[8] YANG Z, LUO T G, WANG D, et al. Learning to navigate for fine-grained classification [C]// European Conference on Computer Vision (ECCV), 2018: 420-435.

[9] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear cnn models for fine-grained visual recognition [C]// 2015 IEEE International Conference on Computer Vision, 2015: 1449-1457.

[10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[11] GAO Y, BEIJBOM O, ZHANG N, et al. Compact bilinear pooling [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 317-326.

[12] KONG S, FOWLKES C. Low-rank bilinear pooling for fine-grained classification [C] // 2017 IEEE

Conference on Computer Vision and Pattern Recognition,2017: 365-374.
[13] PHAM N, PAGH R. Fast and scalable polynomial kernels via explicit feature maps [C] // 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2013: 239-247.
[14] CUI Y,SONG Y,SUN C,et al.Large scale fine-grained categorization and domain-specific transfer learning [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018: 4109-4118.
[15] HE K M,ZHANG X Y,REN S Q,et al.Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[16] KRAUSE J,STARK M,DENG J,et al.3D object representations for fine-grained categorization [C] // 4th International IEEE Workshop on 3D Representation and Recognition,2013: 554-561.
[17] MAJI S,RAHTU E,KANNALA J,et al. Fine-grained visual classification of aircraft [J]. arXiv, 2013: 1306.5151.
[18] KRAUSE J,JIN H,YANG J,et al.Fine-grained recognition without part annotations [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5546-5555.
[19] LAM M, MAHASSENI B, TODOROVIC S. Fine-grained recognition as hsnet search for informative image parts [C] // the IEEE Conference on Computer Vision and Pattern Recognition,2017: 2520-2529.
[20] ZHENG H L,FU J L,ZHA Z J,et al.Learning rich part hierarchies with progressive attention networks for fine-grained image recognition [J]. IEEE Transactions on Image Processing,2019,29: 476-488.
[21] ZHENG H L,FU J L,ZHA Z J,et al.Learning deep bilinear transformation for fine-grained image representation [J].arXiv,2019: 1911.03621.
[22] MIN S B,YAO H T,XIE H T,et al. Multi-objective matrix normalization for fine-grained visual recognition [J]. IEEE Transactions on Image Processing, 2020, 29: 4996-5009.

Unsupervised Fine-grained Image Classification Model Based on Transfer Learning

HAN Tianyu[1], ZHANG Lifeng[1], WANG Xiliang[2]

(1.School of Information and Electrical Engineering,Ludong University,Yantai 264039,China;
2.Stamping Workshop of South Factory,Shanghai General Motors Co.,Ltd.,Yantai 264006,China)

Abstract: Image classification is an important branch in computer vision applications.The common image classifications focus on discriminating the category of different main class objects, while the fine-grained image classifications concentrating on distinguishing the different sub-categories in a main class.The fine-grained image recognition tasks are quite challenging,considering the small differences between the subcategories(slightly differences of some specific parts of objects) and larger intra-subclass differences caused by shooting angles, background,postures and other factors.To alleviate the difficulty in acquiring and expanding of the fine-grained image datasets,an unsupervised fine-grained image classification model based on transfer learning in the form of twin networks was proposed.The model exploits a two-stage training strategy to train the twin network.The first stage pretrains the basic network by a supervised way so that the model can quickly learn the common features of the fine-grained images,while in the second stage,search engine is executed to expand the datasets and employ the twin network to fine-tunes it in an unsupervised manner in order to achieve the knowledge transferring.The experimental results show that the twin network model can achieve good performance on the Stanford Car and Aircraft datasets compared with the counterparts.

Keywords: transfer learning; twin network; unsupervised learning; fine-grained image classification

(责任编辑　李秀芳)