DOI:10.13718/j.cnki.xsxb.2020.01.009

基于注意机制和循环卷积神经网络的细粒度图像分类算法①

王　伟[1]，　吴　芳[2]

1. 郑州工程技术学院 信息工程学院，郑州 450044；
2. 河南财政金融学院 物理与电子工程学院，郑州 450046

摘要：细粒度图像分类是计算机视觉中非常热的研究方向. 由于同一个大物种的子类别之间具有相似的外观，相似的颜色，所以差别非常细微. 因此，细粒度图像分类非常具有挑战性. 为了解决这个挑战，该文提出一种基于注意机制的循环卷积神经网络用于细粒度图像分类. 首先，根据注意机制循环提取一幅图像中的显著性物体区域；然后，对原始图像和每次提取的显著性区域分别进行分类；最后，融合分类层得分，进行最终分类. 在非常有挑战性的公共数据集 CUB-200-2011，Stanford Dogs 和 Stanford Cars 上进行实验，与比较先进的实验方法进行比较，实验结果表明该文提出的方法非常有效.

关　键　词：细粒度图像分类；显著性检测；注意机制；卷积神经网络

中图分类号：TP393　　**文献标志码**：A　　**文章编号**：1000－5471(2020)01－0048－09

随着社会发展，图像分类越来越重要，是计算机视觉领域中比较热门的研究方向. 通用的图像分类的目的主要是区分出不同的物种，比如区分汽车和鸟. 随着深度学习以及计算机视觉的飞速发展，通用图像分类的准确率越来越高. 然而，随着社会发展以及人类的需求多样性，通用图像分类已经满足不了人们的需求. 例如，当人们在天空中看到一只鸟，却分不清具体是什么鸟. 当人们看到飞机，却不知道是什么类型的飞机. 细粒度图像分类是在区分出基础类别的基础上进一步对子类别进行分类. 由于子类别之间往往都非常相似，一般只能通过细微的局部差异对不同的子类别进行区分. 随着社会不断发展，细粒度分类的需求越来越多，如对飞机和汽车[1]分类可以帮助非专业人士进行准确判断，食物分类、菜品分类和服饰分类可以在吃饭、买菜和购物时给顾客带来帮助. 动物子类别分类如鸟[2]、狗也具有广泛的应用前景，如对不同子类别的昆虫进行分类，可以帮助农民快速识别出害虫的种类，进而进行防治工作；对不同子类别的动物进行分类，可以帮助专家更好地区分和保护稀有物种.

研究者们为了解决这一问题，提出了细粒度图像分类方法. 细粒度分类主要是区分大类别中的子类别，比如区分不同类型的鸟. 解决细粒度图像分类问题最简单的方法就是直接使用一般图像分类模型进行训练，但是这样做会导致分类性能低下，无法应用于实际情况. 导致这种现象的主要原因是细粒度图像分类和一般的图像分类存在明显差异. 图 1 是通用图像分类和细粒度图像分类的图像样本. 其中，图 1(a)是通用图像分类的图像样本，主要区分大的物体类别；图 1(b)是细粒度图像分类的图像样本，主要区分大的

① 收稿日期：2019－04－15
基金项目：河南省重点科技攻关项目(182102210594)；河南省高等学校重点科研项目(18A140013).
作者简介：王　伟(1980－)，男，硕士，副教授，主要从事信息处理及电子通信技术研究.

物体类别中的子类别.

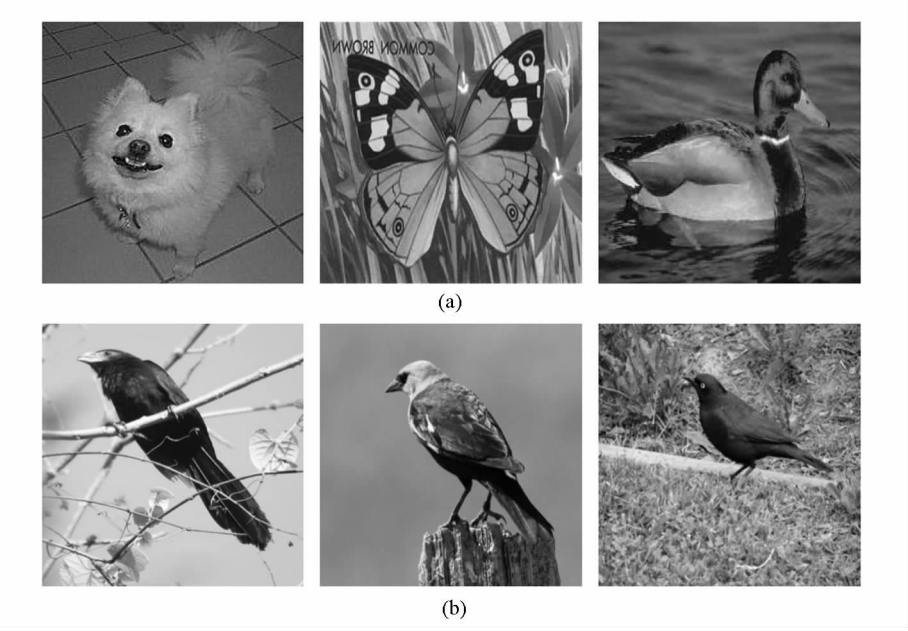

图 1　通用图像分类和细粒度图像分类的图像样本

发现同一个物种之间差别非常细微，如图 1 中第二行的鸟，第一只和第二只，不仅形状很相似，而且颜色都是黑色的，只有细微的差别. 所以，细粒度图像分类是一项极具挑战性的问题. 为此，本文提出了一种细粒度图像分类方法，该方法基于注意机制提取出图像中的显著物体区域，并对提取出的显著区域进行分类，最后通过融合整幅图像和显著物体区域的分类结果，获取最终的图像分类准确率. 为了评估所提出方法的有效性，在公共数据集上做了大量实验. 实验结果表明，本文提出的方法优于其他存在的方法，证明了该方法的有效性.

1　细粒度图像分类、注意机制和卷积神经网络

1.1　细粒度图像分类

图像分类是计算机视觉领域的研究热点之一. 为了解决不同数据集中共同类别图像特征学习能力比较弱的问题，文献[2]提出了一种多重卷积神经网络的跨数据集图像分类方法. 为了解决由于遮挡、光照以及图像姿态变化对图像分类性能的影响，文献[3]提出了一种图像集原型和投影学习算法.

越来越多的学者开始研究极具挑战性的细粒度图像分类. 文献[4]对基于卷积神经网络的细粒度图像分类进行了比较全面的描述，首先介绍了细粒度图像分类的现状，然后分析了强监督和弱监督细粒度图像分类的差异，最后对各种算法进行了总结. 先前的研究者，都是有监督细粒度图像分类[5-6]，根据标注信息提取出显著的物体，然后进行分类. 文献[5]提出一种对深度网络中 filter 进行挑选的方法，基于挑选的 filter 结果构建复杂特征表达. 首先，利用深度 filter 的选择性来挖掘对于某些模式敏感的 filter(比如鸟的喙与腿，如图 1 所示)，从而得到一个 weak 的 Part Detector，进而通过该 Weak Detector 作为初始值来训练一个 Discriminative Part Detector. 文献[6]提出深度模型迁移(DMT)分类方法，该方法可以解决细粒度图像分类中模型复杂度高、很难利用较深的模型等问题.

目前，学者们开始研究弱监督的图像分类[7]. 文献[7]提出一种分层的图像分类方法，该方法联合物体级别和部件级别的特征. 该模型不需要数据集提供的标注信息，而是依赖于自身的算法来获得物体和局部区域. 文献[8]提出一种多任务的域适应方法用于细粒度图像分类，该文章研究了细粒度域适应问题，克服了真实数据难以获得注释这一难题. 文献[9]提出一种低秩的双线性池化方法用于细粒度图像分类，该方

法采用一种深度感知门控模块，该模块根据对象尺度(与深度成反比)自适应地选择卷积网络结构中的池域大小，从而保留图像的细节信息，可以更好地进行分类. 为了利用类间的细微差异，文献[10]提出了基于RPN(Region Proposal Network)与B-CNN(Bilinear CNN)的细粒度分类方法. 为了防止过拟合，首先利用OHEM(Online Hard Example Mine)筛选出对识别结果影响大的图像，然后将筛选之后剩余的图像输入到RPN网络中，得到了对象级别的标注图像，同时将带有对象级别标注信息的图像输入到改进后的B-CNN中，进而进行细粒度图像分类.

1.2 注意机制

当人们在看一样东西的时候，所关注的肯定是当前正在看的这个东西的某一个地方，即当人们的目光转移到别的地方时，注意力会随着目光移动而转移，这就意味着当注意到某个场景或者某个物体时，该场景内以及该目标内每一个位置上的注意力分布不同. 其实，人们在观察图像时，并不是一次性就能把整幅图像每一个位置的像素都看一遍，大多数都是根据需求把注意力集中在图像的特定位置. 人们会根据之前所观察的图像来学习，并且得到未来所要观察图像的注意力应该集中的位置.

注意机制[8,11-12]运用在各行各业，例如图像分类、目标检测、目标跟踪以及姿态估计等等. 注意机制符合人类的视觉机理，首先大致一瞥，第一眼看到感兴趣的区域；然后对感兴趣的区域进行分类、检测、定位等等. 文献[8]提出一种自顶向下的注意机制，利用带有反馈的卷积网络. 文献[12]提出一种细粒度图像检索方法，该方法首先基于显著性注意机制提取出有意义的目标区域，然后提取这些区域中的特征进行图像检索. 实验结果表明这些特征非常具有判别力.

在以上细粒度图像分类方法中，要么需要标注注释信息，而这些注释信息需要人力进行标注，耗费人力财力，大大增加了工作量；要么没有运用注意机制，不能更好地对物体占整幅图比例比较小的图像进行分类. 为了解决以上这些问题，本文提出一种无监督基于注意机制的循环深度卷积神经网络用于细粒度图像分类. 本文提出的方法不仅可以不需要标注信息，节省了大量的人力财力，而且运用了基于注意机制的循环卷积神经网络，可以循环地捕捉微小的细节信息，进而提高细粒度图像分类的性能.

1.3 卷积神经网络

卷积神经网络(Convolutional Neural Networks，CNN)是深度学习(deep learning)的主要算法之一，可以直接将图像作为输入，并且自动地提取特征，还可以对图像进行变形(如比例缩放、平移、倾斜)操作. 卷积神经网络主要包括数据输入层、卷积计算层、ReLU激励层、池化层和全连接层. 数据输入层主要是对原始数据进行预处理，包括去均值、归一化和白化处理. 卷积计算层是卷积神经网络中最重要的层，包括局部关联和窗口滑动2个关键操作. 激励层的作用是对卷积层输出的结果做非线性映射，一般用ReLU做激励层. 池化层在2个卷积层中间，可以减少过拟合. 全连接层是指两层之间的神经元进行两两连接. 本文提出一种基于注意机制的循环卷积神经网络结构，用于细粒图像分类. 实验结果表明，该方法对细粒度图像分类非常有用.

2 基于注意机制的循环卷积神经网络结构

首先，根据预训练模型，生成包含显著物体区域的热图，进而得到显著区域的掩码图；然后，根据掩码图得到显著物体；最后，进行细粒度图像分类.

2.1 显著区域生成

对于给出的尺寸为 $H\times W$ 的图像，卷积层热图是一个包含 $h\times w\times d$ 元素的三维张量，也包含一系列二维的特征图 $S=\{S_n\}(n=1,\cdots d)$. S_n 是尺寸为 $h\times w$ 的第 n 个特征图，对应于第 n 个通道. 深度描述子可以表示为

$$X=\{x_{i,j}\} \tag{1}$$

其中，(i,j) 是一个元组，$i\in(1,\cdots h)$，$j\in\{1,\cdots w\}$. 利用VGGNet-16模型来提取深度描述子，在 $pool_s$ 层

能得到一个 $7\times7\times512$ 的张量. 另一方面，有 512 个尺寸为 7×7 的特征图.

对于一幅图像，有大量无用的部分，需要选取有用的区域. 仅仅用预训练模型，来选取有用信息. 本文方法可以定位出有用物体，忽略噪声，提出一个简单有效的方法来得到有用的特征. 集成特征 A 表示为

$$A=\sum_{n=1}^{d}s_n$$

计算 A 的平均值 a，如果 $A_{i,j}>a$，则掩码图 $M=1$；否则，掩码图 $M=0$. 求出 M 中的连通组件，由于有些图像中有好几个连通组件，选取最大的连通组件. 根据掩码图来得到有用的特征. 具体算法流程如算法 1 所示.

算法 1：计算最大连通组件的算法流程.

算法 1：在图像中找连通的组件

输入：图像

1：用训练好的模型，生成热图；
2：选择热图中的一个像素点 p 作为开始点；
3：**while** True **do**
4：　　用泛洪填充算法去标记包含像素 p 的连通组件中的所有像素；
5：　　**if** 所有的像素都被标记了 **Then**
6：　　Break；
7：　　**end if**
8：　　搜索另一个没有标记的像素作为像素 p；
9：**end while**
10：**return** 返回连通的组件.

图 2 为求最大连通组件的流程图. 图 2(a)是输入的原始的图像；图 2(b)是(a)对应的热图，并用本文算法求出的最大组件. 图 2(c)是利用掩码图求出的一幅图像上的有用信息，即显著区域.

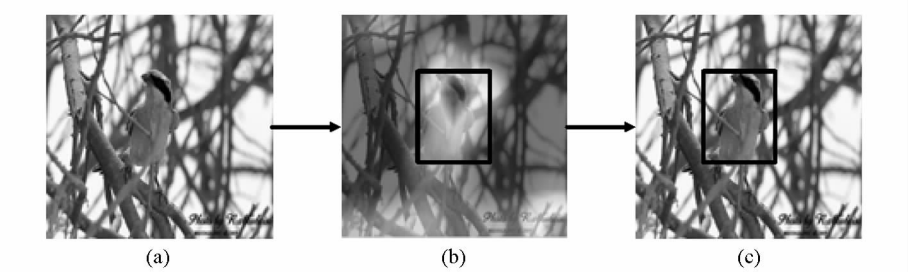

图 2　计算显著区域的流程图

2.2　特征融合

求出一幅图像上有用的信息后进行图像分类. 首先，对于原始的图像图 3(a)，根据图像所对应的热图图 3(c)，得到掩码图 M. 根据章节“2.1 显著区域生成”，可以得到掩码图 M 等于 1 的位置是显著区域；掩码图 M 等于 0 的位置是非显著区域. 人们裁剪矩形区域，使得所有显著区域中所有像素都包含在这个矩形框内，图 3(b)即是对原始图像进行裁剪后的图像. 具体算法流程如算法 2 所示. 然后，可以把物体图 3(b)作为原始图像，根据物体图 3(b)所对应的热图，得到掩码图 M’. 再者，利用掩码图 M’，对物体图 3(b)进行裁剪，得到有用的更小物体部分，人们可以循环地求出更有用、更小的物体区域，这就是本文提出的基于注意机制的循环卷积神经网络结构. 最后，联合原始图像和求得的有用物体部分进行分类，其中图 3(d)分支输入为原始图像，图 3(e)分支输入为裁剪后的图像. 图 3(d)和图 3(e)为卷积神经网络结构，例如 AlexNet 和 VGGNet. 图 3(f)为细粒度图像分类结果.

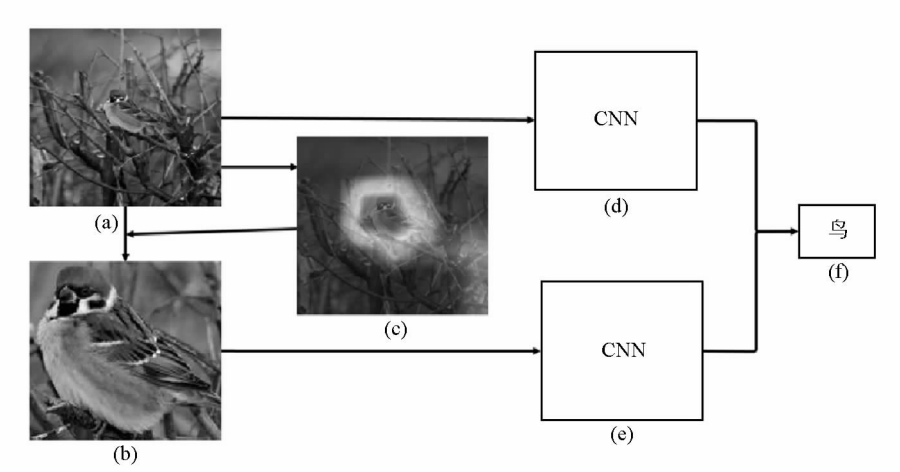

图3 基于注意机制的循环卷积神经网络结构的细粒度图像分类算法流程图

算法2：裁剪过程

输入：图像

1：用训练好的模型，生成热图；

2：根据热图计算出掩码图M；

3：显著图＝原始图像I点成掩码图M；

4：用矩形框框住显著区域；

5：进行裁剪.

输出：裁剪后的图像.

3 实 验

本文为了验证所提出的方法做了大量的实验.

3.1 数据集

为了验证本文提出的细粒度图像分类方法的有效性，在经典的、有挑战性的公共数据集CUB-200-2011数据集[2]、Stanford Dogs数据集[13]和Stanford Cars数据集[1]上进行细粒度分类实验. 另外，为了验证本文提出的定位显著性物体区域的有效性，在经典的、有挑战性的公共数据集PASCAL VOC 2012和MS COCO上进行定位实验.

CUB-200-2011数据集在细粒度图像分类任务中是使用最广泛的一个数据库，它包含200种不同种类，一共有11 788幅鸟类图像数据. 其中，5 994幅图片用于训练，5 794幅图片用于测试. 每张图片都有详细的人工标注，包括1个子类别标签，1个图片主体标注框，15个局部区域位置以及312个二值属性. 所有的属性都与特定部分的颜色、图案或者形状有关. 本文仅仅使用子类别标签.

Stanford Dogs数据集包含120种不同种类，一共有20 580幅鸟类图像数据. 其中，12 000幅图片用于训练，8 580幅图片用于测试.

Stanford Cars数据集包含196种不同种类，一共有16 185幅鸟类图像数据. 其中，8 144幅图片用于训练，8 041幅图片用于测试.

PASCAL VOC 2012数据集一共有11 530张图像，每幅图像都带有标注，标注物体包括人、交通工具(如船、飞机等)、动物(如狗、猫等)、家具(如沙发、桌子等)在内的20个物体类别.

MS COCO数据集由微软构建，包含检测、分割和定位等任务. 与PASCAL VOC数据集相比，COCO数据集中的图像包含了生活中常见图像以及自然图像，图像目标数量较多，背景较复杂，目标物体尺寸较小，因此COCO数据集上的定位等任务更难.

3.2　本文所用的主干网络结构

本文用 VGGNet 网络结构在 Caffe 平台进行实验. VGGNet 每层的参数个数如表 1 所示. VGGNet 是由牛津大学计算机视觉组和 DeepMind 公司一起研发的一种卷积神经网络，并于 2014 年在 ILSVRC 竞赛中获得了图像分类项目第二名和图像定位项目第一名. VGGNet 共有 6 种不同类型的网络结构，每种网络结构都有 5 组卷积层，每组卷积层都用的卷积核，并且每组卷积层后都进行了一个最大池化，接着是 3 个全连接层. 在训练较高级别网络的时候，可以先训练较低级别的网络，然后用前者所得到的权重来初始化高级别的网络结构，这样可以加快网络结构的收敛速度. VGGNet 中比较出名的是 VGGNet-16 和 VGG-Net-19，最常用的是 VGGNet-16. VGGNNet-16 共 16 层，包括 13 个卷积层和 3 个全连接层.

表 1　VGGNet 网络结构的参数

类型	滤波尺寸	滤波数量	步幅	输出尺寸
Input	—	—	—	224×224
Conv	3×3	64	1	224×224
Conv	3×3	64	1	224×224
Maxpool	2×2	128	2	112×112
Conv	3×3	128	1	112×112
Conv	3×3	128	1	112×112
Maxpool	2×2	256	2	56×56
Conv	3×3	256	1	56×56
Conv	3×3	256	1	56×56
Conv	3×3	256	1	56×56
Maxpool	2×2	512	2	28×28
Conv	3×3	512	1	28×28
Conv	3×3	512	1	28×28
Conv	3×3	512	1	28×28
Maxpool	2×2	512	2	14×14
Conv	3×3	512	1	14×14
Conv	3×3	512	1	14×14
Conv	3×3	512	1	14×14
Maxpool	2×2	512	2	7×7
FC	1×1	4096	—	1×1
FC	1×1	4096	—	1×1
FC	1×1	ClassNum	—	1×1

3.3　实验结果

对本文提出的细粒度图像分类方法在 3 个经典的 CUB-200-2011 数据集、Stanford Dogs 数据集和 Stanford Cars 数据集上进行大量的验证. 根据算法 1，求出的掩码图和有用物体区域如图 4 所示. 其中，第一行表示原始图像，第二行表示掩码图，第三行方框中是求出的有用物体区域. 通过实验结果可以看出，本文方法很精确地定位出了有用物体区域.

为了验证本文提出的定位方法的有效性，将本文提出的物体区域定位方法与其他先进的定位方法 OS-Boxes[14] 和 WILDCAT[15] 进行了对比. 在数据集 PASCAL VOC2012 和数据集 MSCOCO 上的实验结果分别如表 2 和表 3 所示. 实验结果表明，本文提出的定位显著物体区域方法优于其他先进的图像定位方法，进一步说明了本文所提方法的有效性.

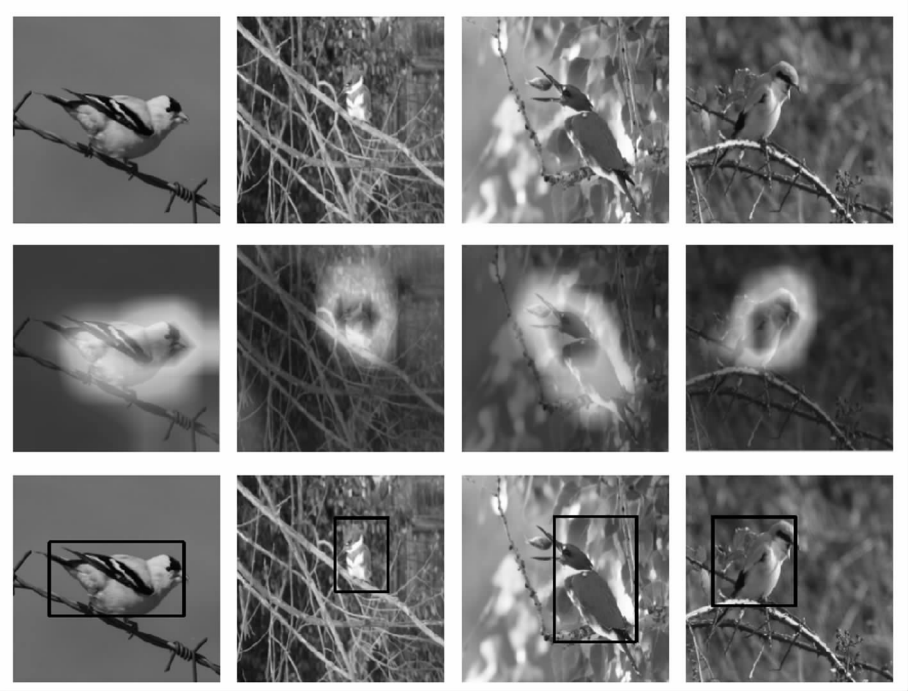

图4 原图、掩码图和有用物体区域

表2 本文物体定位方法和其他定位方法在数据集 PASCAL VOC 2012 上的定位结果

方 法	定位性能/%
OS-Boxes[14]	77.7
WILDCAT[15]	82.9
Ours	84.7

表3 本文物体定位方法和其他定位方法在数据集 MS COCO 上的定位结果

方 法	定位性能/%
OS-Boxes[14]	46.4
WILDCAT[15]	53.4
Ours	57.3

将本文提出基于注意机制的循环卷积神经网络细粒度图像分类方法与其他先进细粒度图像分类方法PDFR[5]、模型迁移[6]、Two-level[7]、Multi-task[8]、Low-rank[9]，Look and Think Twice[11]进行对比. 在实验中，Full 表示只用原始图像进行分类，Object 表示只用本文得到的有用物体进行分类，Full+Object 表示联合原始图像和得到的物体区域进行分类. 在数据集 CUB-200-2011，Stanford Dogs 和 Stanford Cars 上的实验结果分别如表4，表5和表6所示，实验结果表明，Full+Object 比 Full 和 Object 的分类准确率高，证明了本文方法的有效性. 此外，与其他比较好的方法进行比较，本文方法亦优于其他方法，因为本文方法提取出有用的信息，这样的信息比较有判别力，可以更好地进行图像分类.

表4 本文提出的分类方法和其他分类方法在 CUB-200-2011 数据集上的分类结果

方 法	分类准确率/%
PDFR[5]	80.3
Two-level[7]	82.8
Multi-task[8]	81.0
Low-rank[9]	81.7
LaT Twice[11]	82.6
Full	81.3
Object	80.1
Full+Object(Ours)	85.4

表 5　本文提出的分类方法和其他分类方法在 Stanford Dogs 数据集上的分类结果

方　　法	分类准确率/%
PDFR[5]	79.3
Two-level[7]	83.4
Multi-task[8]	81.5
Low-rank[9]	82.8
LaT Twice[11]	83.6
Full	82.9
Object	81.6
Full+Object(Ours)	86.3

表 6　本文提出的分类方法和其他分类方法在 Stanford Cars 数据集上的分类结果

方　　法	分类准确率/%
PDFR[5]	82.3
Two-level[7]	85.1
Multi-task[8]	83.9
Low-rank[9]	86.3
LaT Twice[11]	89.1
Full	88.1
Object	86.6
Full+Object(Ours)	91.3

4　结　　语

本文提出了一种基于注意机制和循环卷积神经网络的细粒度图像分类算法，首先基于注意机制提取图像的显著区域，然后结合原始图像和得到的有用物体区域进行分类. 在数据集 PASCAL VOC2012 和数据集 MSCOCO 上进行定位实验，实验结果表明，本文提取的显著性物体非常准确. 在公开的、有挑战性的数据集 CUB-200-2011，Stanford Dogs 和 Stanford Cars 上进行大量的细粒度分类实验，实验结果表明本文提出的细粒度图像分类方法有效.

参考文献：

[1] KRAUSE J, STARK M, DENG J, et al. 3D Object Representations for Fine-Grained Categorization [C]//IEEE International Conference on Computer Vision Workshops. New York: IEEE, 2013.

[2] 刘鑫童，刘立波，张　鹏. 基于多重卷积神经网络跨数据集图像分类 [J]. 计算机工程与设计，2018，39(11)：3549-3554.

[3] 路　易，吴玲达，朱　江. 基于卷积神经网络的高光谱图像分类方法 [J]. 计算机工程与设计，2018，39(9)：2836-2841.

[4] 罗建豪，吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述 [J]. 自动化学报，2017，43(8)：1306-1318.

[5] ZHANG X, XIONG H, ZHOU W, et al. Picking Deep Filter Responses for Fine-Grained Image Recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016.

[6] 刘尚旺，郜　翔. 基于深度模型迁移的细粒度图像分类方法 [J]. 计算机应用，2018，38(8)：64-70.

[7] XIAO T, XU Y, YANG K, et al. The Application of Two-Level Attention Models in Deep Convolutional Neural Network for Fine-Grained Image Classification [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015.

[8] GEBRU T, HOFFMAN J, L Fei-Fei. Fine-Grained Recognition in the Wild: A Multi-Task Domain Adaptation Approach [C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017.

[9] KONG S, FOWLKES C. Low-rank Bilinear Pooling for Fine Grained Classification [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017.

[10] 赵浩如，张　永，刘国柱. 基于 RPN 与 B-CNN 的细粒度图像分类算法研究 [J]. 计算机应用与软件，2019，36(3)：210-213，264.

[11] CAO C S, LIU X M, YANG Y, et al. Look and Think Twice: Capturing Top-Down Visual Attention with Feedback Convolutional Neural Networks [C]//2015 IEEE International Conference on Computer Vision(ICCV). Santiago: IEEE, 2016.

[12] WEI X S, LUO J H, WU J X, et al. Selective Convolutional Descriptor Aggregation for Fine-Grained Image Retrieval [J]. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881.

[13] KHOSLA A, JAYADEVAPRAKASH N, YAO B, et al. Novel Dataset for Fine-Grained Image Categorization: Stanford Dogs [C]. Denver: Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC), 2011.

[14] SUN C, PALURI M, COLLOBERT R, et al. ProNet: Learning to Propose Object-Specific Boxes for Cascaded Neural Networks [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016.

[15] DURAND T, MORDAN T, THOME N, et al. WILDCAT: Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu: IEEE, 2017.

Fine-Grained Image Classification Algorithm Based on Attention Mechanism and Circular Convolutional Neural Network

WANG Wei[1], WU Fang[2]

1. *College of Information Engineering, Zhengzhou Institute of Technology, Zhengzhou 450044, China;*
2. *College of Physical and electronic Engineering, Henan Finance University, Zhengzhou 450046, China*

Abstract: Fine-grained image classification is a hot research field in computer vision. Because subcategories within a large species have similar appearances and similar colors, the differences are subtle. Therefore, fine-grained image classification is very challenging. To solve this problem, an attention-based cyclic convolutional neural network for fine-grained image classification has been proposed in this paper. Firstly, according to the attention mechanism, the region of the significant object in an image is extracted. Secondly, the original image and the significance region of each extraction are classified respectively. And finally, the score of classification layer is fused for final classification. We conduct experiments on very challenging public datasets: CUB 200—2011, Stanford Dogs and Stanford Cars. We compared our method with the state-of-the-art methods, and the experimental results show that our proposed method is very effective.

Key words: fine-grained image classification; significance detection; attention mechanism; convolutional neural network

责任编辑 夏 娟