Computer Science and Application 计算机科学与应用, 2020, 10(5), 841-850
Published Online May 2020 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2020.105087

Hans汉斯

Fine-Grained Image Classification Algorithm Based on Grad-CAM and B-CNN

Shaowei Deng, Boquan Zhang

Department of Computer, Guangdong University of Technology, Guangzhou Guangdong
Email: shaowei.deng@foxmail.com

Received: Apr. 16th, 2020; accepted: May 1st, 2020; published: May 8th, 2020

Abstract

Fine-grained images are characterized by small differences between classes and large differences within classes. The differences between images mainly exist in subtle local areas, and local area localization and its representative feature extraction have become one of the main research issues in fine-grained image classification. In this paper, the fine-grained categorization method is studied based on the Grad-CAM and the Bilinear Convolution Neural Networks B-CNN. It uses the Grad-CAM model to locate the salient region in the original image, and crops the salient region image as the input of the bilinear CNN, fusing the global and local features to complete the classification. Experiments on the three datasets of CUB-200-2011, Stanford Dogs and Stanford Cars show that compared with the traditional model, this method can more accurately locate areas with significant image features and have better classification effects.

Keywords

Fine-Grained Categorization, Bilinear Convolution Neural Networks, Grad-CAM, Salient Regions

基于Grad-CAM与B-CNN的细粒度图像分类方法研究

邓绍伟，张伯泉

广东工业大学计算机学院，广东 广州
Email: shaowei.deng@foxmail.com

收稿日期：2020年4月16日；录用日期：2020年5月1日；发布日期：2020年5月8日

摘　要

细粒度图像具有类间差异小，类内差异大的特点。图像之间的差异主要存在于细微的局部区域，局部区域定位及其代表性特征提取成为细粒度图像分类的主要研究问题之一。本文基于Grad-CAM和双线性卷

文章引用: 邓绍伟, 张伯泉. 基于 Grad-CAM 与 B-CNN 的细粒度图像分类方法研究[J]. 计算机科学与应用, 2020, 10(5): 841-850. DOI: 10.12677/csa.2020.105087

积神经网络B-CNN模型对细粒度图像分类方法进行研究，它利用Grad-CAM模型定位原图像中的显著区域，并裁剪出显著性区域图像作为双线性CNN的输入，融合全局和局部的特征，从而完成分类。在CUB-200-2011、Stanford Dogs和Stanford Cars三个数据集上的实验表明，相较于传统模型，该方法能够更加准确定位图像特征显著区域，具有更好的分类效果。

关键词

细粒度图像分类，双线性卷积神经网络，Grad-CAM，显著性区域

1. 引言

近年来，随着图像数据的大规模增长，人们对图像分类提出了迫切需求，图像分类成为了热门研究领域[1]。图像分类一般分为对象级分类和细粒度图像分类(Fine-grained image categorization) [2]，细粒度图像分类又被称为子类别图像分类(Sub-category recognition) [3]，细粒度图像分类是对粗粒度图像中的更小的子类别进行分类，如飞机型号、鸟类品种、服装款式与菜肴样式等。由于同一个粗粒度图像下的各子类图像在几何结构上十分相似，造成细粒度图像类间差异小；而同一个细粒度类别下的图像，从物体的形状、姿态、背景等角度来看都有可能产生极大的差异，致使子类内图像区分难度大。细粒度图像在工业界与学术界都应用广泛，比如在道路交通管理上，可以识别不同车型的数量，计算实时的交通状况；在生物学领域，可以帮助研究人员快速识别不同种类的物种，而不用受太大专业知识的限制。因此细粒度图像分类成为研究热点和难点[4]。

细粒度图像分类主要分为强监督的分类方法和弱监督的分类方法。强监督细粒度分类方法除了需要图像的类别标签之外，还需要标注框、局部位置等额外的人工标注信息，而这些人工的标注信息往往是需要丰富的专业知识才能够获得，所以这一类方法的代价较高。文献[5]提出的 Part-based R-CNN，该算法采用 R-CNN 对图像生成大量的候选区域，然后再对这些候选区域检测，给出每一个局部区域的评分值，根据评分值确定最后的定位检测结果。结合全局特征(物体级别特征)和判别行更强的局部特征进行分类取得了不错的效果，但是需要额外的人工标注开销，并且由于 R-CNN 算法产生大量的候选区域会大大增加计算复杂度，造成检测速度较慢。文献[6]提出的 Part-Stacked CNN，它与文献[5]类似，也是分为两个步骤，由定位网络与分类网络两部分组成，在定位网络中用到了 FCN (Fully Convolutional Network)提高了分类准确率，并且加快了算法的效率，但同样是需要对象与部位级的标签。文献[7]提出了一种新颖的局部区域检测模型，在细粒度图像分类中引入了协同分割，提出了一种无需借助局部区域标注信息，只需要标注框就可以完成分割与对齐操作，分类准确度能够达到 82%。

由于获取人工标注信息代价大，弱监督图像分类方法越来越受到重视。文献[8]提出两级注意力(Two level attention)算法，它关注对象级和局部级两个不同层次的特征。但是利用聚类算法得到的局部区域并不十分准确，所以分类准确率有限。文献[9]提出基于双线性卷积神经网络(Bilinear Convolution Neural Networks, B-CNN)的弱监督分类模型，它由两路 VGGNet 构成。该模型将两个网络提取的卷积特征进行双线性操作，以提高图像特征表达能力，实现了一个端到端训练的弱监督分类网络。但是该模型利用 VGGNet 作为特征提取网络，没有能充分关注物体判别性区域对分类准确率的影响。

综合以上，影响细粒度图像分类准确率的两个主要因素，一是对图像局部显著性区域的关注，二是对局部区域特征的提取和表达。本文在 B-CNN 的基础上，提出一种基于 Grad-CAM [10]与 B-CNN 的细粒度图像分类方法，该方法首先利用 Grad-CAM 模型提高对象级显著性区域检测结果，聚焦判别性区域，去除无关区域对分类结果的影响，并且改用更加有效的特征提取函数，采用 B-CNN 模型对判别性区域进行特征提取与分类，从而提高细粒度图像的分类准确率。这种方法不需要标注框和局部位置等人工标记信息，也能够减少背景区域的干扰，在理论上将讲这种方法是一种有效的细粒度图像分类方法。

2. 相关理论

2.1. 类别激活映射

2.1.1. CAM

具有区别性的显著性区域特征是分类的关键。类别激活映射(Class Activation Mapping, CAM) [11]提供了一种解释图像分类结果的方法，它采用全局池化层(global average pooling, GAP)，解决全连接层参数过多、网络不易训练和容易过拟合等问题。该方法以 HeatMap 来映射图像中与该类别的最相近的即显著性区域，使得模型有更强的解释性。是一种寻找图像中显著性区域的更好的方法。CAM 模型如图 1 所示。

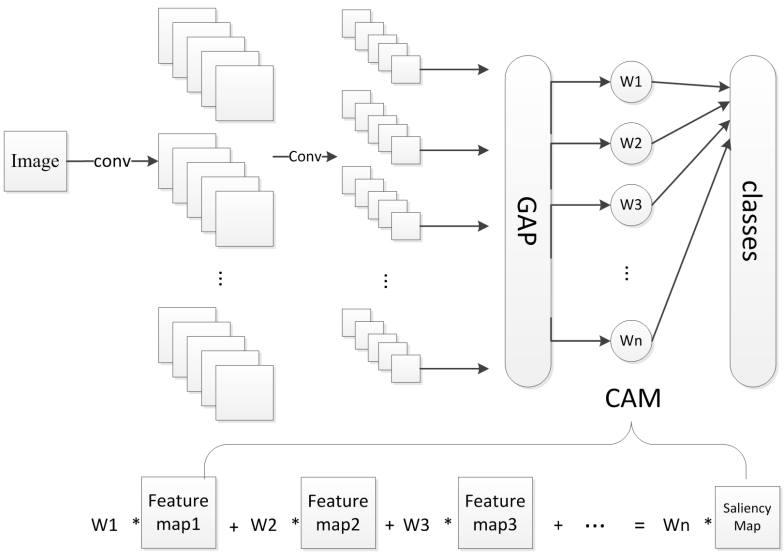

Figure 1. CAM model
图 1. CAM 模型

CAM 方法用 GAP 层替换全连接层来接收最后一层卷积的结果，输出每一个特征图在每一个通道的平均值，接着是全连接层接收这些平均值生成最后的预测值。选择最终全连接层节点的权重作为最后一层卷积层的特征图的权值，并对特征图像按通道加权形成最后的类别激活映射图。

2.1.2. Grad-CAM

CAM 能够反映特定类别的显著性区域，实现分类解释。但是 CAM 需要改变原模型结构，并且重新训练，这大大限制了 CAM 的应用场景。Grad-CAM 与 CAM 的基本思路是一样，都是通过得到每一个通道特征图的权重，最后加权求和。但是，与 CAM 不同的是 Grad-CAM 不需要改变原模型结构，只需要通过梯度的全局平均求取通道映射为类别的权重,这样可以保留卷积之后的全连接层，并且经过数学推导证明 Grad-CAM 与 CAM 得到的通道特征图的权重是一致的。Grad-CAM 模型如图 2 所示。

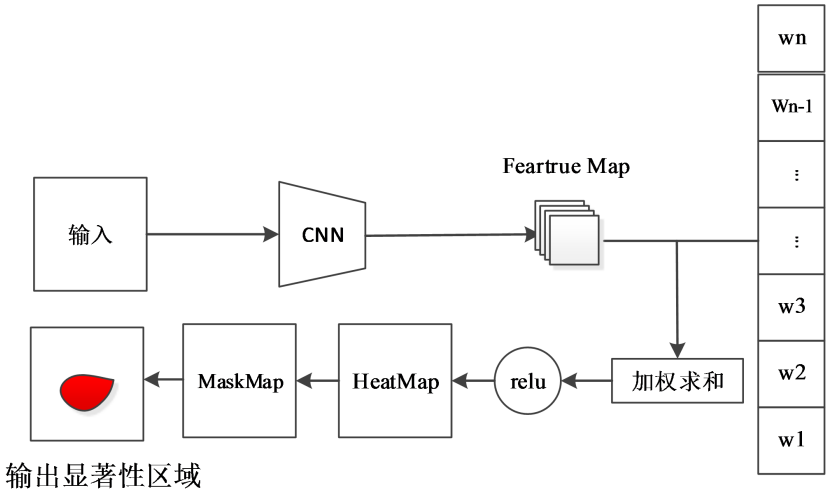

Figure 2. Grad-CAM model
图 2. Grad-CAM 模型

参考图 1、图 2 模型，设 $f_k(x,y)$ 为最后一个卷积层输出的特征图中的第 k 个通道的 (x,y) 位置上的激活值， F^k 为通道 k 的特征图，则：

$$F^k = \sum_{x,y} f_k(x,y) \tag{1}$$

对于某一个特定类别标签 c， S_c 表示的是 softmax 层的输入，则：

$$S_c = \sum_k w_k^c F^k \tag{2}$$

其中 w_k^c 是通道 k 映射为类别 c 的权重。所以根据式(1)、式(2)：则：

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y) \tag{3}$$

用 M_c 表示类别 c 对应的激活映射， $M_c(x,y)$ 表示激活映射图 (x,y) 位置上的激活值，则：

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \tag{4}$$

第 k 个通道对应的类别 c 的权重值 w_k^c 为：

$$w_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \tag{5}$$

其中 Z 为特征图中像素的个数， y^c 是对应类别 c 的分数(在代码中一般用 logits 表示，是输入 softmax 层之前的值)， $A_{i,j}^k$ 表示第 k 个特征图中， (x,y) 位置处的像素值。

类别映射图反映了原图中各个显著性区域与特定分类分别之间的相关性。因此可以利用 Grad-CAM 来进行位置定位，用于检测相应物体在原图中的区域，并获取 HeatMap 中的最大联通区域的边界框，将边界框作为定位框。

2.2. B-CNN

双线性卷积神经网络(B-CNN)是一个典型的弱监督的细粒度图像分类算法,不需要任何的人工标记就已经在鸟、飞机、狗等数据集上达到了较好的准确率，其模型如图 3 所示：

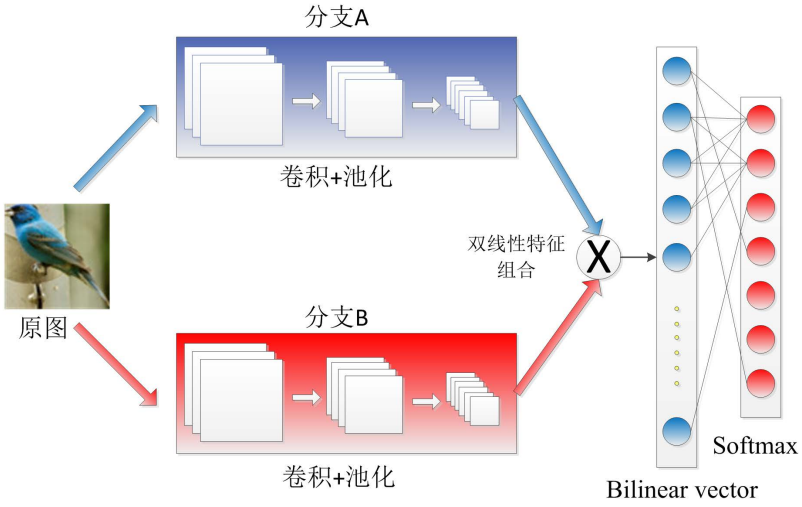

Figure 3. Structure of B-CNN model
图 3. B-CNN 模型结构

一个双线性卷积神经网络模型可由一个四元组 β 表示，$\beta=(f_A, f_B, P, C)$。其中 f_A、f_B 分别表示两个特征提取函数，P 是池化函数，C 是分类函数。在每一个位置对两个网络提取到的特征做外积，组合成双线性特征(bilinear feature)，如式(6)所示：

$$bilinear(l, I, f_A, f_B) = f_A(l, I)^{\mathrm{T}} f_B(l, I) \tag{6}$$

其中，l 表示每一个局部位置，I 表示原图。

对所有位置得到的双线性特征进行求和池化，作为原图像的特征：

$$\phi(I) = \sum_{l \in L} bilinear(l, I, f_A, f_B) \tag{7}$$

2.3. 残差模型

经典的双线性卷积神经网络是由两个 VGG 分支网络组成的，Stream A 进行图像中的目标定位，检测局部区域；Stream B 进行定位后区域的特征提取。两个网络相互协调工作，最终完成对图像的分类。通常情况下，深度卷积神经网络的层数较少时，可以增加深度来获得更好的特征提取效果；一旦网络层数过高，会使得网络产生大量的参数，也难以使得网络收敛。VGG 对于细粒度图像分类的有一定的局限性，对特征的提取和表达不能更加精确。文献[12]表明，随着网络层数的增加，网络发生了退化(degradation)的现象。因此，在文献[12]中何凯明等人提出了深度残差网络(ResNet)。其残差模块结构如图 4 所示。

其中，X 是第一层残差模块的输入，$F(X)$ 是经过第一层线性变化并激活后的输出。在第二层线性变化之后激活之前，$F(X)$ 加入了第一层输入值 X，然后激活输出。残差网络的提出解决了深层网络梯度消失的问题，提升了网络的分类准确度，相比于双线性卷积网络所使用的 VGG 网络，ResNet 有更深的网络结构，也能更加准确识别图片中的细节特征，从而实现精细化识别。

3. 本文模型结构及算法流程

结合 Grad-CAM 与 B-CNN 模型，本文采用的细粒度图像分类模型如图 5 所示。

本模型分为两个模块，原图经过 Grad-CAM 检测之后，原图中更精细区域被定位，在 Grad-CAM 中依旧采用 VGG 网络来实现检测；显著性区域检测出来为提取特征做准备，B-CNN 双线性模型中的 VGG 网络利用 ResNet50 代替，更加高层的双线性特征有利于最后的分类任务。

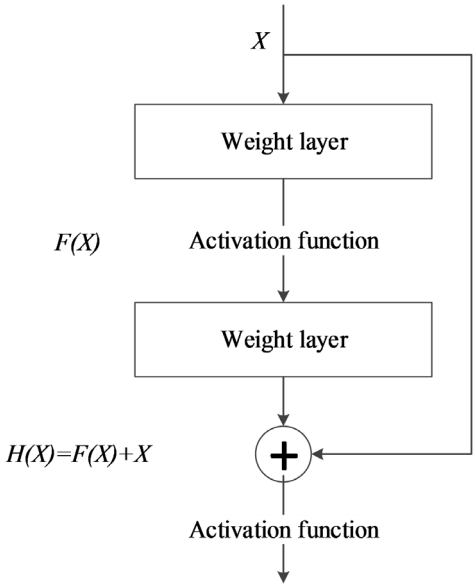

Figure 4. Structure of residual module
图 4. 残差模块结构

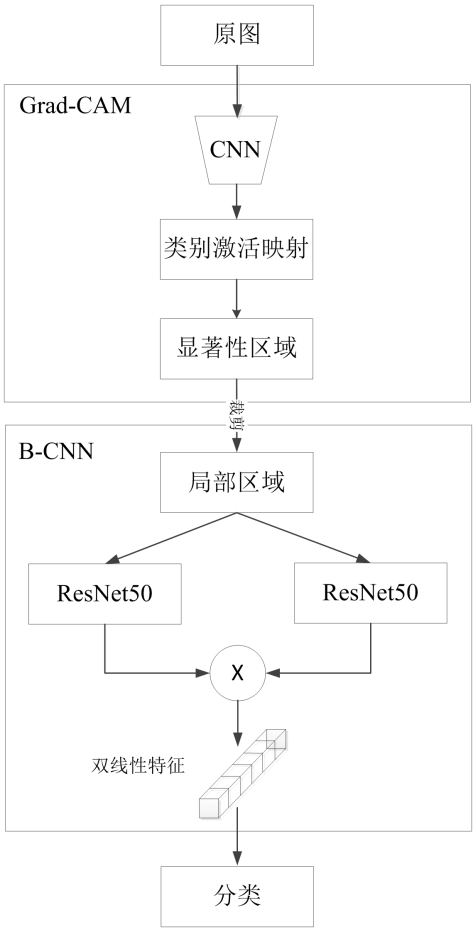

Figure 5. Fine-grained image classification model based on Grad-CAM and B-CNN
图 5. 基于 Grad-CAM 和 B-CNN 的细粒度图像分类模型

算法描述如下：

1) 输入原图，对原图进行统一尺度缩放为 $H \times W$ 的图像 I；

2) 将图像 I 进行卷积，并计算最后一层特征图 $A \in R^{w \times h \times d}$， $w \times h$ 表示 A 的空间维度，d 表示通道数量。A 的第 k 个通道对应的 c 类别的权重值为 α_k^c，分别表示每一个通道的重要程度，则：

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{i,j}^k}}_{\text{gradients via backprop}} \tag{8}$$

$A_{i,j}^k$ 表示通道 k 中位置 (i,j) 的像素值，用梯度 $\frac{\partial y^c}{\partial A^k}$ 的全局平均来计算每一个通道的权重值 α_k^c。

3) 求得所有特征图的权重之后对其加权求和并 ReLU 激活；

$$L_{Grad-CAM}^c = \text{ReLU}\underbrace{\left(\sum_k \alpha_k^c A^k\right)}_{\text{linear combination}} \tag{9}$$

A^k 表示通道 k 的卷积特征图，$L_{Grad-CAM}^c$ 表示类别 c 判别定位图，用 ReLU 线性组合所有的加权特征图，输出突出目标类别的热力图；

4) 得到图像 I 的热力图 $L_{Grad-CAM}^c$，设定一个像素阈值 β，如果热力图中 $L_{i,j}^c \geq \beta$，则令 $L_{i,j}^c = 255$；如果 $L_{i,j}^c \leq \beta$，则令 $L_{i,j}^c = 0$，通过热力图得到掩码图 M；

5) 显著图 = $L_{Grad-CAM}^c \cdot M$，通过点乘去掉不显著区域，得到显著图；

6)根据显著图中像素值为 255(即白色部分)区域，用矩形框框出显著图区域并从原图中裁剪出显著图部分的图像 I_2，并对 I_2 的尺寸归一化；

7) 双路残差网络把 I_2 映射成同一维度的特征，两个特征通过一个双线性池化操作 P 汇聚，得到一维的双线性特征 B；

8) 对所有位置的双线性组合特征 B 求和；

9) 训练网络，完成分类。

在算法开始前，将原图缩放到统一的尺寸。然后利用 Grad-CAM 对图像的显著性区域进行定位，生成显著性区域的热图。根据热图得到掩码图，并计算掩码图中的最大联通区域，该连通区域就是目标物体所在的位置。从原图中相应位置裁剪出显著性区域，并缩放图像尺寸，并输入双线性残差卷积神经网络以完成分类

4. 实验与分析

在算法开始前，将原图缩放到统一的尺寸。然后利用 Grad-CAM 对图像的显著性区域进行定位，生成显著性区域的热图。根据热图得到掩码图，并计算掩码图中的最大联通区域，该连通区域就是目标物体所在的位置。从原图中相应位置裁剪出显著性区域，并缩放图像尺寸，并输入双线性残差卷积神经网络以完成分类。

4.1. 实验设计

论文使用开源深度学习框架 Keras 作为实验平台，基于一台英伟达 GTX1070 显卡和 16G DDR4 内存的 ubuntu16.04 计算机系统上采用 Python 编程实现。

实验采用加州理工大学鸟类数据集 CUB-200-2011 [13]、斯坦福大学狗类数据集 Stanford Dogs [14]和斯坦福大学汽车数据集 Stanford Cars [15]等三个经典的细粒度图像分类数据集。

CUB-200-2011 鸟类数据集是在细粒度图像分类领域使用最为广泛的一个数据集。该数据集包括11,788 张鸟类图片，一共分为 200 个类别。其中 5994 张图片用于训练模型，5794 张图片用于测试模型。每一张图片都有详细的人工标注标签，物体标注框和局部位置标注点。Stanford Dogs 狗类数据集总共包括 20,580 张图片，其中 12,000 张图片用于训练模型，8058 张图片用于测试模型，一共分为 120 个类别。Stanford Cars 汽车数据集总共包括 16185 张图片，其中 8144 张图片用于训练模型，8041 张图片用于测试模型，一共分为 196 个类别。

因为细粒度图像分类的 3 个数据集都比较小，用于训练和测试的样本数较少，如果直接在这 3 个数据集上训练可能会导致网络无法收敛，因此，利用在 ImageNet 数据集上预训练好的参数对网络进行初始化，然后在三个细粒度图像数据集上对模型进行微调(fine-tuning)，这样会有更好的训练效果。

输入图像统一缩放成 448 * 448 的三通道的彩色图像，裁剪出来的显著性区域统一缩放为 224 * 224 * 3，学习率设置为 0.005，训练批次为 32，迭代次数为 100,000 次，使用随机梯度下降优化器来训练和优化模型。

4.2. 性能指标

对于以上所述的三个经典数据集，为了直观体现算法的性能与效果，使用分类准确度 Accuracy 作为评价指标。

$$\text{Accuracy} = \frac{N_{test}}{N} \tag{10}$$

其中，N 表示总共用于测试样本的数量 N_{test}，表示测试样本中正确预测的数量。用准确度 Accuracy 可以直观的反应算法的分类性能。

4.3. 实验结果

通过实验，计算出原图的掩码图与显著性图如图 6 所示。

Figure 6. Salient regions detection of Fine-grained images. (a) Original image; (b) HeatMap; (c) Salient region
图 6. 细粒度图像显著性区域检测。(a) 原图；(b) 热图；(c) 显著性区域

将本文提出的基于 Grad-CAM 与 B-CNN 的细粒度分类方法与 PDFR [16]、Two-level [17]、Low-rank [18]、Constellations [19]、DVAN [20]、B-CNN 等主流细粒度分类算法对比，在三个数据集上的实验结果如表 1 所示。

实验结果表明本文方法在 CUB-200-2011 数据集上的分类效果比其他的方法略有提高，比 B-CNN 模型提高了 0.6%；在 Stanford Dogs 数据集上，本文算法优于 B-CNN、Low-Rank、Two-Level 等弱监督分类模型 87.1%；与原模型 B-CNN 相比较，在 3 个数据集上的分类效果都有所提高。

Table 1. Comparison of experimental result
表 1. 实验结果对比

方法	分类准确度		
	CUB-200-2011	Stanford Dogs	Stanford Cars
PDFR [16]	80.3	79.3	82.3
Two-level [17]	82.8	83.4	85.1
Low-rank [18]	81.7	82.8	86.3
Constellations [19]	81.0	68.61	\
DVAN [20]	79.0	81.5	87.1
B-CNN [7]	84.1	\	91.3
Grad-CAM + B-CNN	84.7	87.1	91.8

5. 结束语

本文在双线性卷积神经网络的基础上提出改进的基于梯度类别激活映射与双线性残差网络的细粒度分类方法。首先基于 Grad-CAM 提取图像中的显著性区域，将显著性区域从原图中裁剪出来并预处理，用两个 ResNet50 网络作为特征提取函数，提取显著性区域更加细致的特征，然后通过结合全局与局部的特征信息进行分类。在 3 个经典数据集上实验结果表明，在不使用物体包围框以及局部位置标注点的情况下，本文方法可以提升双线性卷积神经网络的分类性能，能够优于其它分类方法对细粒度图像进行有效分类。

基金项目

本文得到广东省自然科学基金项目(No.2019A1515011056，2018A030313868)的资助。

参考文献

[1] 杨兴. 基于 B-CNN 模型的细粒度分类算法研究[D]: [硕士学位论文]. 北京: 中国地质大学, 2017.

[2] Fu, J., Zheng, H. and Mei, T. (2017) Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 4438-4446. https://doi.org/10.1109/CVPR.2017.476

[3] 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.

[4] 盛纾纬. 基于弱监督学习的细粒度图像识别技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2019.

[5] Zhang, N., Donahue, J., Girshick, R., *et al*. (2014) Part-Based RCNNs for Fine-Grained Category Detection. In: *Proceedings of the* 13*th European Conference on Computer Vision*, Springer, Zurich, 834-849. https://doi.org/10.1007/978-3-319-10590-1_54

[6] Huang, S., Xu, Z., Tao, D., *et al*. (2016) Part-Stacked CNN for Fine-Grained Visual Categorization. *Computer Vision and Pattern Recognition IEEE*, Las Vegas, 27-30 June 2016, 1173-1182. https://doi.org/10.1109/CVPR.2016.132

[7] Krause, J., Jin, H.L., Yang, J.C., *et al*. (2015) Fine-Grained Recognition without Part Annotations. *Proceedings of the* 15*th IEEE International Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 5546-5555. https://doi.org/10.1109/CVPR.2015.7299194

[8] Xiao, T.J., Xu, Y.C., Yang, K.Y., *et al*. (2015) The Application of Two-Level Attention Models in Deep Convolutional Neural Network for Fine-Grained Image Classification. *Proceedings of the* 2015 *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 842-850. https://doi.org/10.1109/CVPR.2015.7298685

[9] Lin, T.Y., Aruni, R., Maji, S., *et al*. (2015) Bilinear CNN Models for Fine-Grained Visual Recognition. *Proceedings of the* 15*th IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1449-1457. https://doi.org/10.1109/ICCV.2015.170

[10] Selvaraju, R., Cogswell, M., Das, A., *et al*. (2017) Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 618-626. https://doi.org/10.1109/ICCV.2017.74

[11] Zhou, B., Khosla, A., Lapedriza, A., *et al*. (2016) Learning Deep Features for Discriminative Localization. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 2921-2929. https://doi.org/10.1109/CVPR.2016.319

[12] He, K.M., Zhang, X.Y., Ren, S.Q., *et al*. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[13] Wah, C., Branson, S., Welinder, P., *et al*. (2011) The Caltech-UCSD Birds-200-2011 Dataset. Computation & Neural Systems Technical Report, CNS-TR, California Institute of Technology, Pasadena.

[14] Khosla, A., Jayadevaprakkash, N., Yao, B.P., *et al*. (2011) Novel Dataset for Fine-Grained Image Categorization: Stanford Dogs. *Proceedings of the* 1*st Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, 1-2.

[15] Krause, J., Stark, M., Deng, J., *et al*. (2013) 3d Object Representations for Fine-Grained Categorization. *Proceedings of the 4th IEEE Workshop on 3D Representation, IEEE International Conference on Computer Vision*, Sydney, 2-8 December 2013, 554-561. https://doi.org/10.1109/ICCVW.2013.77

[16] Zhang, X.P., Xiong, H.K., Zhou, W.G., *et al*. (2016) Picking Deep Filter Responses for Fine-Grained Image Recognition. *IEEE Conference on Computer Visio and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1134-1142. https://doi.org/10.1109/CVPR.2016.128

[17] Xiao, T., Xu, Y., Yang, K., *et al*. (2015) The Application of Two-Level Attention Models in Deep Convolutional Neural Network for Fine-Grained Image Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 842-850.

[18] Kong, S. and Fowlkes, C. (2017) Low-Rank Bilinear Pooling for Fine-Grained Classification. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 7025-7034. https://doi.org/10.1109/CVPR.2017.743

[19] Simon, M. and Rodner, E. (2015) Neural Activation Constellations: Unsupervised Part Model Discovery with Convolutional Networks. *Proceedings of the* 15*th IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1143-1151. https://doi.org/10.1109/ICCV.2015.136

[20] Zhao, B., Wu, X., Feng, J., *et al*. (2017) Diversified Visual Attention Networks for Fine-Grainde Object Classification. *IEEE Transactions on Multimedia*, **19**, 1245-1256. https://doi.org/10.1109/TMM.2017.2648498