基于 RPN 与 B-CNN 的细粒度图像分类算法研究

赵浩如 张 永 刘国柱
(青岛科技大学信息科学技术学院 山东 青岛 266000)

摘 要 随着大数据和硬件的快速发展,细粒度分类任务应运而生,其目的是对粗粒度的大类别进行子类分类。为利用类间细微差异,提出基于 RPN(Region Proposal Network)与 B-CNN(Bilinear CNN)的细粒度图像分类算法。利用 OHEM(Online Hard Example Mine)筛选出对识别结果影响大的图像,防止过拟合;将筛选后的图像输入到由 soft-nms(Soft Non Maximum Suppression)改进的 RPN 网络中,得到对象级标注的图像,同时减少假阴性概率;将带有对象级标注信息的图像输入到改进后的 B-CNN 中,改进后的 B-CNN 可以融合不同层特征并加强空间联系。实验结果表明,在 CUB200-2011 和 Standford Dogs 数据集平均识别精度分别达到 85.50% 和 90.10%。

关键词 细粒度分类 类间差异 双向卷积网络 非极大值抑制 特征融合

中图分类号 TP391.41 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.03.038

FINE-GRANTED IMAGE CALSSIFICATION ALGORITHM BASED ON RPN AND B-CNN

Zhao Haoru Zhang Yong Liu Guozhu
(*College of Information Science and Technology*, *Qingdao University of Science and Technology*, *Qingdao 266000*, *Shandong*, *China*)

Abstract With the rapid development of big data and hardware, fine-grained classification has emerged. Its purpose is to classify the coarse-granted categories into subclasses. In order to use the subtle differences between similarities, we proposed a fine-granted classification algorithm based on RPN and B-CNN. The online hard example mine (OHEM) algorithm was used to screen out the images which had a great impact on the recognition results to prevent the over-fitting. Then, the selected image was input into the RPN network improved by soft non maximum suppression(soft-nms). The false negative probability was reduced, and the image with object-level annotation was obtained. The image with object-level annotation was input the improved B-CNN. The improved B-CNN could fuse features of different layers and enhanced their spatial connection. The experimental results demonstrate that the average recognition accuracy of CUB200-2011 and Stanford Dogs datasets is 85.50% and 90.10%.

Keywords Fine-granted classification Interclass difference B-CNN Non-maximum suppression Feature fused

0 引 言

作为计算机视觉的重要研究方向,图像分类[1]问题一直备受学者关注。图像分类又包括对象级分类,如对猫和狗进行分类。还包括细粒度分类[2],如对狗的不同品种进行分类。由于细微的类内差异,往往只能借助微小的局部差异才能分出不同的子类别,使得细粒度分类十分具有挑战性。细粒度分类的方法主要包括两种:一种是基于强监督的分类模型,如 Part-based R-CNN[3]不仅需要物体级标注,还需要局部区域的标注,这大大限制了在实际场景的应用;另一种是基于弱监督的分类模型,如 B-CNN[4]仅仅需要图像级别的标注,不需要局部信息的标注。因此,基于弱监督的分类模型在识别精度上要比基于强监督的分类模型差一些。Huang 等[5]提出了 Part-Stacked CNN 进行细粒度分类。这个网络需要提供对象及部位级标签,它分为定位网络和分类网络两个子网络,采用经典的 Alex-

收稿日期:2018 - 09 - 30。国家自然科学基金项目(61472196, 61672305, 61702295);山东省自然科学基金项目(ZR2014FM015)。**赵浩如**,硕士生,主研领域:计算机视觉与图像工程。**张永**,硕士生。**刘国柱**,教授。

Net 网络结构作为整个网络的基本结构。Shen 等[6]提出一种迭代的传递策略来优化目标框,借助对象及部分级标注框进行细粒度分类。Yao 等[7]提出了多级的由粗到细的目标描述方法进行细粒度分类,不需借助标注框,但识别率不如最前沿的算法。Liu 等[8]提出了基于全连接的注意力机制的网络结构进行细粒度分类,未考虑各层特征间的联系。Murabito 等[9]提出显著性特征分类网络(SalClassNet)。它包括两个子网络,网络 A 计算输入图片的显著性特征,网络 B 计算网络 A 输出的显著性特征进行细粒度分类,计算显著性特征首先要计算图像像素对应正确分类标准化分数梯度的绝对值,然后取三个颜色通道的最大值,因此,计算成本太高。综上,为避免人工标注部位级标签花费的巨大时间,以及减少计算成本。本文提出利用 soft-nms 和 OHEM 优化 RPN 算法得到更精确的对象级标注,以防止背景的干扰,同时改进 B-CNN 网络,加强不同层特征间的空间联系,提高识别精度。

1　算法描述

为利用细微的类内差异,本文采用 OHEM[10]筛选出对识别结果影响大的数据,可以有效防止无关信息的干扰。然后,利用 soft-nms[11]优化 RPN[12]网络,选择出置信度更高的目标所在区域。最后,改进 B-CNN 网络结构对目标区域进行细粒度分类,具体的算法流程如图 1 所示。

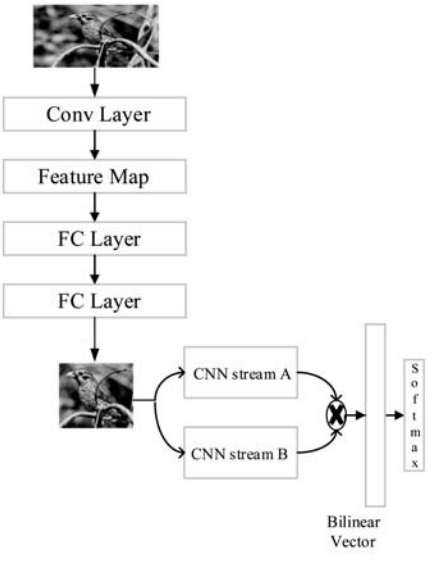

图 1　算法流程图

1.1　获取目标区域

RPN 网络的作用是输入一张图像,输出置信度排名前 N 个目标可能在的区域。本文利用 OHEM 筛选出对最终识别结果影响大的样本,并用筛选后样本进行随机梯度下降。去除了对识别结果影响小的样本后,有效防止过拟合,具体算法流程如图 2 所示。

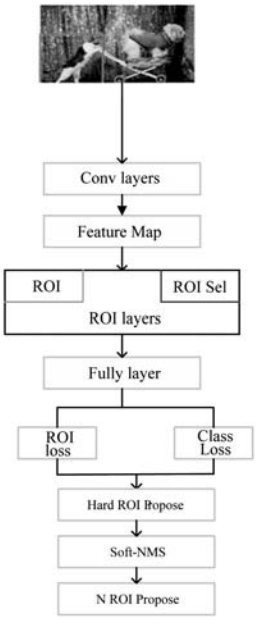

图 2　获取目标区域算法描述图

图 2 中,OHEM 有两个不同的 ROI 网络。左边的 ROI 网络只负责前向传播计算误差,右边的 ROI 网络从左边的 ROI 网络中通过对误差排序,选出误差大的样本作为右边 ROI 网络的输入。RPN 网络输出的矩形目标框 D_i,其得分 f_i 的计算如下:

$$f_i = \frac{S_i}{S} \tag{1}$$

式中:S_i 是重叠框的交集的面积;S 是重叠框的并集的面积。

NMS(Non-maximum suppression)是 RPN 中重要的组成部分。RPN 输出一系列的检测框 D_i 以及对应的分数 f_i。NMS 设置常数阈值 τ,当检测框的得分大于阈值 τ,将其放入最终的检测结果集合 D。同时,集合 D 中任何与检测框 M 的重叠部分大于重叠阈值 τ 的检测框,被强制归零并移除。非最大抑制算法中的最大问题就是将相邻检测框的分数均强制归零后,如果真实的物体在重叠区域出现,则将导致对该物体的检测失败并降低了算法的平均检测率。soft-nms 不将大于阈值 τ 的相邻目标框得分重置为 0,而是乘以一个衰减函数。选取所有的目标框中得分最高的 N 个,这样可以有效减少假阴性的概率,提高平均识别率。具体计算如下:

$$f_i = \begin{cases} f_i & f_i \geqslant \tau \\ f_i \times (1 - f_i) & f_i < \tau \end{cases} \tag{2}$$

1.2　基于深度学习进行细粒度分类

Bilinear CNN 模型包括 Stream A 和 Stream B,Stream A 和 Stream B 的网络结构都是采用的 VGGNet。Stream A

的作用是对物体的局部信息进行定位,而 Stream B 则是对 Stream A 检测到的局部信息进行特征提取。两个网络相互协调作用,完成了细粒度图像分类过程中两个最重要的任务:物体、局部区域的检测与特征提取。本文在 B-CNN 基础上增加了两个外积操作,外积计算如下:

$$B = f_A^{\mathrm{T}} \cdot f_B \tag{3}$$

双线性特征 B_2、B_3 分别是 conv4_3 的特征与 conv5_3 的特征,conv5_1 的特征与 conv5_3 的特征进行点乘得到的。然后将双线性特征 B_2、B_3 与原有的 conv5_3 层特征与 conv5_3 层特征点乘得到的双线性特征 B_1 拼接起来,以加强不同层特征间的空间联系。最后,将拼接后的特征 B 送进全连接层,进行 softmax 分类。具体算法流程如图 3 所示。

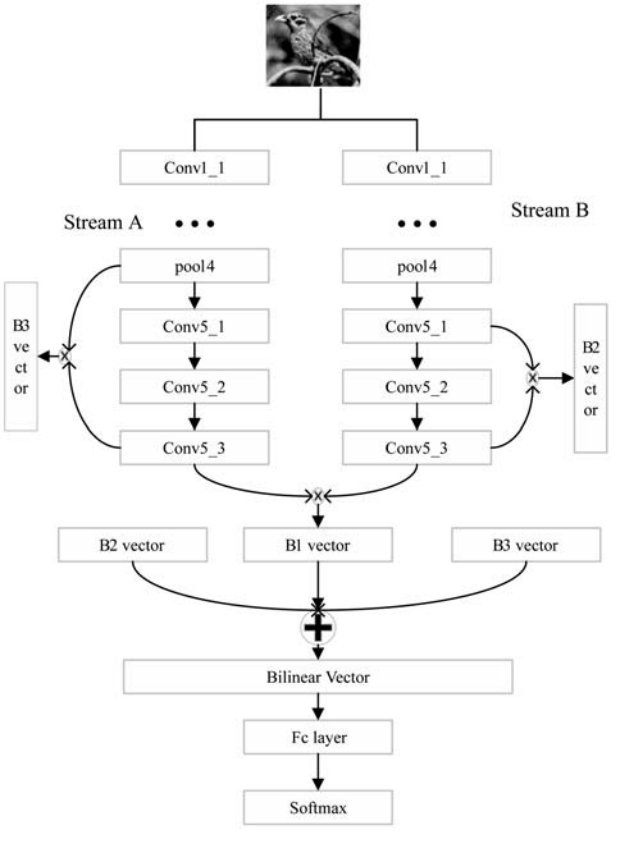

图 3 改进的 B-CNN 算法流程图

f_A 与 f_B 是不同层的特征,双线性特征 $\boldsymbol{B}_i$ 是一个 $\boldsymbol{C} \times \boldsymbol{W} \times \boldsymbol{H}$的三维矩阵,将其转化为长度为 $\boldsymbol{C}$、$\boldsymbol{W}$、$\boldsymbol{H}$ 的列向量。然后将双线性特征 B_1、B_2、B_3 拼接成一个长度为 $3\boldsymbol{CWH}$ 的列向量 $\boldsymbol{B}$,将其输入到 softmax 函数进行分类。最后,模型在端到端[13]的训练过程中,从图 3 可以看出模型的前半部分是普通的卷积层与池化层。因此,只要求得后半部分的梯度值,即可完成对整个模型的训练。假设 Stream A 与 Stream B 的特征分别是 f_a 与 f_b,则双线性特征为 $B = f_a^{\mathrm{T}} \cdot f_b$。$\frac{d_l}{d_B}$为特征 B 的梯度值,l 是损失函数。根据链式法则结合式(4)、式(5),得到两个网络的梯度值,从而完成端到端的模型训练。

$$\frac{d_l}{d_{f_a}} = f_b\left(\frac{d_l}{d_B}\right)^{\mathrm{T}} \tag{4}$$

$$\frac{d_l}{d_{f_b}} = f_a\left(\frac{d_l}{d_B}\right) \tag{5}$$

2 实验仿真

2.1 实验背景

为验证本算法的有效性,与文献[5-9]中的算法的结果进行对比。文献[5-9]分别采用 CUB200-2011 数据集[14]与 Standford Dogs 数据集[15]。因此本文也在两组数据集上进行两组实验,来证明本算法的识别精度比文献[5-9]中的结果高。第一组实验是在 CUB200-2011 数据集进行的测试和验证。该数据集是最常用和经典的细粒度分类数据集,包括 200 中不同类别,共 11 788 张不同鸟类图片,不仅提供了对象级标注框而且还提供了局部级标注框。第二组实验是在 Standford Dogs 数据集[15]进行测试和验证。该数据集包括 120 类狗的图像数据,共有 20 580 张图片,只提供对象级标注框。基于 RPN 与 B-CNN 的细粒度分类过程中所用到的主要参数如表 1 所示。

表 1 B-CNN 主要参数表

Batch size	Display	Momentum	Base_lr
10	20	0.9	0.001

2.2 实验结果与分析

本文利用 OHEM 与 soft-nms 优化 RPN,获取对象级标注,然后输入到改进的 B-CNN。在 RPN 阶段,训练集、验证集和测试集的比例是 7:2:1。采取的 Anchor 的尺度是(128,256,512),比例为(0.5,1,2),共九种。将一张图片输入到 RPN 就会产生大量的 Anchor,对这些 Anchor 进行 soft-nms,最终输出得分最高的目标框。在目标框提供的位置上剪贴图片,剪贴后的图片只含有目标对象,没有背景的干扰。B-CNN 阶段中训练集,验证集与测试集的比例是 7:1.5:1.5。在 ImageNet 中 1 000 类分类训练好的参数的基础上,在 CUB200-2011 数据集进行微调。将图片输入到 B-CNN 后,Stream A 的作用是对图像中对象的特征部位进行定位,而 Stream B 则是用来对 Stream A 检测到的特征区域进行特征提取。两个网络相互协调作用,完成了细粒度图像分类过程中两个最关键的任务。

本文采用 softmax 函数做分类函数输出一个概率

值,计算公式如下:

$$S_i = \frac{e_i}{\sum_j e_j} \tag{6}$$

式中:S_i 是第 i 个类别的概率值;e_i 是第 i 个类别的得分。

与文献[5-6]借助对象级及部位级标注框进行细粒度分类对比,本文仅仅采用了对象级标注框。与文献[7]利用迭代的方法获取对象级与部位级标注框对比,本文利用 RPN 提取目标区域,并将深度学习框架的注意力[16]全放在目标区域,防止无关信息的干扰,提高识别速度与精度。实验结果如表 2 所示。实验表明,本文的算法识别率为 85.5%,比文献[5-7]中的方法分别高了 8.90%、1.5%、3.0%。证明本文提出的基于 RPN 与 B-CNN 的细粒度分类算法,将识别的重心放在目标区域内。利用 B-CNN 优化目标区域的同时,在目标区域内提取特征,不仅不需要提供额外的部位级标注框,并且准确率有较大提高。

表 2　不同方法在 CUB200-2011 数据集的识别率

方法	Huang	Shen	Yao	Our approach
准确率	76.60%	84.00%	82.50%	85.50%

Standford Dogs 数据集是从 ImageNet 数据集中提取狗的类别组成的。本文在第一组实验获取的参数基础上进行微调,实验结果如表 3 所示。与文献[8]基于对象级与部位级标注框与注意力机制相比,虽然两者都将识别重心放在目标区域,但本文在仅仅使用对象级标注框的前提下,利用外积将 B-CNN 的 Strean A 与 Stream B 统一成一个端到端的训练模型。与文献[9]使用 SalClassNet 网络提取显著性特征,并对显著性特征进行细粒度分类相比,本文使用对象级标注框在 ROI 区域上进行特征提取。因此,识别率分别比文献[8]和文献[9]的方法高了 1.2% 和 3.9%。这表明同时对标注框与类别进行端到端的训练能有效提高识别率。

表 3　不同方法在 Standford Dogs 数据集的识别率

方法	Liu	Murabito	Our approach
准确率	88.9%	86.20%	90.10%

此外,对本文提出的算法,增加了 5 组对比实验分别为:方案一,不使用 OHEM 优化 RPN,不改变 B-CNN 网络结构;方案二,不使用 soft-nms 优化 RPN,不改变 B-CNN 网络结构;方案三,在使用 OHEM 及 soft-nms 的前提下,不增加 B-CNN 的外积操作;方案四,仅增加 B-CNN 的外积操作;方案五,使用 OHEM 及 soft-nms,同时增加 B-CNN 的外积操作。实验对比结果如表 4 所示。实验结果表明,方案五的识别率为 90.10%,比方案一、方案二、方案三、方案四分别高了 2.9%、2.3%、1.6%、1.1%。方案一仅使用 OHEM,仅有效地防止了过拟合;方案二仅使用 soft-nms,使输出的对象级标注更加准确,并减少了假阴性概率;方案三则结合了方案一与方案二,识别率有所提升;方案四仅增加 B-CNN 的外积操作,加强了不同层之间的空间联系。这表明使用 OHEM 与 soft-nms 改进 RPN,能让获得的对象级标注更加精确,既可以避免背景的干扰,减少假阴性,又能有效防止过拟合。而增加 B-CNN 的外积操作,增加了不同层特征间的空间联系。这是因为不同层关注的特征不同并且感受野大小也不同,这可以有效地提高识别率。

表 4　对比实验结果图

方法	方案一	方案二	方案三	方案四	方案五
准确率	87.20%	87.90%	88.50%	89.00%	90.10%

3　结　语

本文针对细粒度分类子类别间细微的类间差异、较大的类内差异、依赖大量人工标注信息等问题,提出了基于 RPN 与 B-CNN 的细粒度分类算法。本文的主要贡献如下:(1) 利用 RPN 网络自动输出对象级标注,不需要部位级标注,避免标注对象部位花费的精力。(2) 使用 soft-nms 和 OHEM 算法改进 RPN,输出更加精确的区域提议,可以有效防止过拟合并减少假阴性概率。(3) 改进 B-CNN 网络,增加不同层间的外积操作,以融合不同层的特征,并将双线性特征级联在一起加强空间的联系。实验结果证明,基于 RPN 与 B-CNN的细粒度分类算法能显著提高识别率。但由于增加了 RPN 网络以及 OHEM 与 soft-nms 操作,程序的运行时间相比其他算法有所增加。并且,未将 RPN 网络与 B-CNN 网络联合起来,也是本文的不足。接下来,我们的工作重心将放在使 RPN 与 B-CNN 网络联合成一个端到端的模型,并提取同类物体不同子类的差异特征,作为深度网络的输入来提高准确率。

参　考　文　献

[1] 彭晏飞, 陶进, 訾玲玲. 基于卷积神经网络和 E2LSH 的遥感图像检索研究[J]. 计算机应用与软件,2018,35(7):250-255.

(下转第 264 页)

参 考 文 献

[1] Storn R, Price K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces [J]. Global Optimization, 1997, 11(4): 341 - 359.

[2] Vesterstrom J, Thomsen R. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems[C]//Proceedings of the 2004 Congress on Evolutionary Computation. IEEE, 2004: 1980 - 1987.

[3] Rahnamayan S, Tizhoosh H R, Salama M M A. Opposition-based differential evolution[M]//Advances in Differential Evolution, 2008: 155 - 171.

[4] Brest J, Greiner S, Boskovic B, et al. Self-Adapting Control Parameters in Differential Evolution: A Comparative Study on Numerical Benchmark Problems[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(6): 646 - 657.

[5] Liu J, Lampinen J. A fuzzy adaptive differential evolution algorithm[J]. Soft Computing—A Fusion of Foundations, Methodologies and Applications, 2005,9(6): 448 - 462.

[6] Qin A K, Suganthan P N. Self-adaptive differential evolution algorithm for numerical optimization[C]//Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2005, 2 - 4 September 2005, Edinburgh, UK. IEEE, 2005.

[7] Yang Z, Tang K, Yao X. Self-adaptive differential evolution with neighborhood search[C]//2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). IEEE, 2008.

[8] Yao X, Liu Y, Lin G. Evolutionary programming made faster[J]. IEEE Transactions on Evolutionary Computation, 1999, 3(2): 82 - 102.

[9] Fan H, Lampinen J. A trigonometric mutation operation to differential evolution[J]. Journal of Global Optimization, 2003,27(1): 105 - 129.

[10] Ali M M. Population set-based global optimization algorithms: some modifications and numerical studies[J]. Computers and Operations Research,2004,31(10):1703 - 1725.

[11] Sun J, Zhang Q, Tsang E P K. DE/EDA: a new evolutionary algorithm for global optimization[J]. Information Sciences—Informatics and Computer Science, Intelligent Systems, Applications: An International Journal, 2005,169(3/4): 249 - 262.

[12] Yang Z Y, He J S, Ya X. Making a difference to differential evolution[M]//Advances in Metaheuristics for Hard Optimization, 2007: 397 - 414.

(上接第 213 页)

[2] Dasgupta R, Namboodiri A M. Leveraging multiple tasks to regularize fine-grained classification[C]//International Conference on Pattern Recognition. IEEE,2017:3476 - 3481.

[3] Sang N, Chen Y, Gao C, et al. Detection of vehicle parts based on Faster R-CNN and relative position information [C]//Pattern Recognition and Computer Vision. 2018:83.

[4] Lin T Y, Roychowdhury A, Maji S. Bilinear CNN Models for Fine-Grained Visual Recognition[C]//IEEE International Conference on Computer Vision. IEEE,2016:1449 - 1457.

[5] Huang S, Xu Z, Tao D, et al. Part-Stacked CNN for Fine-Grained Visual Categorization[C]//Computer Vision and Pattern Recognition. IEEE, 2016:1173 - 1182.

[6] Shen Z, Jiang Y G, Wang D, et al. Iterative object and part transfer for fine-grained recognition[C]//IEEE International Conference on Multimedia and Expo. IEEE,2017:1470 - 1475.

[7] Yao H, Zhang S, Zhang Y, et al. Coarse-to-Fine Description for Fine-Grained Visual Categorization[J]. IEEE Transactions on Image Processing, 2016, 25(10):4858 - 4872.

[8] Liu X, Xia T, Wang J, et al. Fully Convolutional Attention Networks for Fine-Grained Recognition[EB]. arXiv:1603.06765, 2017.

[9] Murabito F, Spampinato C, Palazzo S, et al. Top-Down Saliency Detection Driven by Visual Classification[J]. Computer Vision & Image Understanding, 2018, 40(7):1130 - 1141.

[10] Shrivastava A, Gupta A, Girshick R. Training Region-Based Object Detectors with Online Hard Example Mining[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:761 - 769.

[11] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with RegionProposal Networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems—Volume 1. MIT Press, 2015:91 - 99.

[12] 杨国亮, 王志元, 张雨, 等. 基于垂直区域回归网络的自然场景文本检测[J]. 计算机工程与科学, 2018, 40(7): 1256 - 1263.

[13] Yeung S, Russakovsky O, Mori G, et al. End-to-end learning of action detection from frame glimpses in videos[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:2678 - 2687.

[14] 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.

[15] 杨兴. 基于 B-CNN 模型的细粒度分类算法研究[D]. 北京:中国地质大学(北京), 2017.

[16] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2017:1480 - 1489.