细粒度图像分类研究综述

卢婧宇

摘要：细粒度图像分类是近几年新出现的一个研究方向，它是对属于同一层次（例如不同的鸟种），并且具有相似的形状或视觉外观的分类对象进行分类的任务。虽然它与一般对象分类有关，但细粒度分类要求算法在通常仅由细微差别区分的高度相似的对象之间进行区分。随着近年来研究的越来越深入，目前已经有很多方法可以实现。本文首先介绍细粒度图像分类的定义以及研究意义，并再介绍其算法的发展现状。之后从一定角度分析不同算法之间的差异,并比较了这些算法在常用数据集上的性能表现。最后，对这些算法进行总结，并且讨论未来该领域的发展和进一步要研究的问题。

关键词：细粒度图像分类 计算机视觉 深度学习

A Survey on Fine-Grained Image Categorization

Abstract：Fine granularity image classification is a new research direction in recent years ,which is a task for classifying objects that belong to the same level (such as different bird species) and have similar shapes or visual appearance. Although it is related to general object classification, fine-grained classification requires that the algorithm distinguish between objects that are very similar to those that are usually distinguished by only minor differences. With the deepening of research in recent years, there are many ways to achieve. This paper first introduces the definition of fine granularity image classification and research significance, and then introduces the development status of its algorithm. Then analyzes the differences between different algorithms from a certain angle, and compares the performance of these algorithms on common data sets. Finally, the paper summarizes these algorithms, and discusses the future development and further research in this field.

Key words: Fine-grained image categorization, computer vision, deep learning

细粒度图像分类（Fine-Grained Categorization）作为一个近几年热门的研究方向，越来越受到各方面的关注。细粒度分类属于目标识别的一个子领域，它的目标是区分属于同一基本层次范畴的数百个子类。精细图像分类研究的是需要专业知识指导才能进行的物种分类问题。这类问题的一个共同点是不同类别之间的相似度较高，同一类别内的差异性较大，所以细粒度图像分类与传统的图像分类相比难度更大。

由于近些年计算机发展的很快，又有很多新的方法出现解决了以往存在的

种种方面的缺陷。以往的方法有基于标注信息的方法，这种方法需要人们对于目标特性或者关键点位置有冗长的注释，并且人工花费大量的时间在标注图像环节上，这对全自动学习来进行区分图像信息的情况构成阻碍。但随着，深度学习的兴起，基于卷积神经网络的各种新的方法大量出现，使得细粒度图像分类的准确度和便捷度有所提高。本文的主要内容是，对部分传统方法进行简单说明，并对比近期新兴方法在准确度，是否需要人工标注以及能否实现端到端的这几方面进行比较。最后，对对未来将要发展的方向进行展望和总结。

1.细粒度图像分类概述

图像信息的获取和传输是图像处理系统的重要组成部分，直接影响图像处理系统的性能。从目标的区分精细程度来看，图像分类分为两种：一种是传统的图像分类方法，属于粗糙分类；另一种是精细图像分类方法，属于细分类。而传统的图像分类是用来区分物体属于哪一物种；而精细分类是用来区分物体属于哪一类。现在，精细图像分类研究工作主要集中在花卉鸟类狗类等,并且也取得了一定的研究成果。目前，绝大多数的分类算法遵循的流程框架是：首先找到前景对象(鸟)及其局部区域(头、脚、翅膀等),之后分别对这些区域提取特征。对所得到的特征进行适当的处理之后，用来完成分类器的训练和预测。

传统的方法依赖于大量的人工标注，但是准确度也并不好还很繁琐。常用训练标注有标注框(BBox)和局部区域位置(Parts)。先利用标注框将图像中的要是别的前景对象选中，之后检测局部有用特征比如选出鸟的腿部，之后进行训练。然而，人工标注信息耗费的人力时间都很多，所以目前新兴的方法大部分偏向弱监督，无需人工标注只需要标签就可以自动训练，并且有一些算法的准确度可以达到百分之八十多。局部特征的提取也是牵制结果准确度的一方面，

类算法一般是先从图像中提取 SIFT 或者 HOG 这些局部特征。SIFT 特征不只具有尺度不变性，即使改变旋转角度，图像亮度或拍摄视角，仍然能够得到好的检测效果。Hog 特征结合 SVM 分类器已经被广泛应用于图像识别中，尤其在行人检测中获得了极大的成功。之后利用 VLAD 或者 Fisher Vector 等编码模型进行特征编码，得到最终所需要的特征表示。然而，由于人工特征的描述能力有限，导致分类效果不佳。

2．细粒度图像数据库介绍

针对目前的常用数据库进行介绍并大致说明其特点和缺陷：

- CUB200-2011[1]：CUB200-2011 是细粒度图像分类领域最经典，也是最常用的一个数据库，共包含 200 种不同类别，共 11,788 张鸟类图像数据。同时，该数据库提供了丰富的人工标注数据，每张图像包含 15 个局部区域位置，312 个二值属性，1 个标注框，以及语义分割图像。由于规定了局部区域位置的数量可能会导致图片的资源浪费。
- Stanford Dogs[2]：该数据库提供了 120 种不同种类的狗的图像数据，共有 20,580 张图，只提供标注框这一个人工标注数据。提供的人工标注少，对于一些需求大量人工标注的算法不方便，若不再额外进标注可能结果准确度会有下降。
- Oxford Flowers[3]:分为两种不同规模的数据库，分别包含 17 种类别和 102 种类别的花。其中，102 种类别的数据库比较常用，每个类别包含了 40 到 258 张图像数据，总共有 8,189 张图像。该数据库只提供语义分割图像。不包含其他额外标注信息。图像比较少，对于深度卷积来说不是很足够。
- Cars[4]：提供 196 类不同品牌不同年份不同车型的车辆图像数据，一共包含

有 16,185 张图像，只提供标注框信息。

- FGVC-Aircraft[5]：提供 102 类不同的飞机照片，每一类别含有 100 张不同的照片，整个数据库共有 10,200 张图片，只提供标注框信息。

3. 基于人工特征的早期算法简述

随着图像识别与分类的不断发展，针对传统的图像分类算法基本上准确率已经很高了，并基本成熟。而细粒度图像分类比较困难算法和准确度还是比较有限。比较经典的算法是，在发布 CUB200-2011 数据库的技术报告中 Wah[1]等人给出的基准测试的结果仅为 10.3%。算法过程是，给定一张原始的、未经过裁剪的测试图像，利用训练得到的模型完成局部区域的定位;之后，提取 RGB 颜色直方图和向量化的 SIFT 特征，经过词包(bag of words, BoW)模型进行特征编码后，输入到线性 SVM 分类器完成分类。但如果在测试时给定了标注框和局部区域位置这些标注信息的话,利用同样的方法，得到的基准测试结果为 17.3%。一方面，是由于定位不够准确,局部区域无法归一化对齐；另一方面，则是因为特征的描述能力太弱，不具备足够的区分度。这种方法明显准确率太低。之后，提出一种方法（基于部位的一对一特征，POOFs）可以自动地从一组特定领域的带有特定位置和类别标注的图片集中学习大量不同的具有高区分性的中级特征。每一个特征都能够根据对象特定位置的表观特征来区分两个不同的类。这种方法精确度最高可以达到 73.3%。

可见传统的方法准确度不高，其缺陷在于需要大量的人工标注，同时利用定位算法去确定关键点的效果不够好，因此需要进一步的改善方法。

4.细粒度图像分类现阶段国内外进展

4.1SWFV-CNN

特征表示是细粒度识别的一个关键问题.近年来,卷积神经网络(CNN)被广泛应用于特征提取。然而,细粒度表示存在两个挑战.首先,传统的 CNN 表示需要固定大小的矩形作为输入,这不可避免地包括背景信息。然而,背景不太可能对细粒度的识别起任何重要作用,因为所有的子类都有相似的背景(例如,所有的鸟通常栖息在树上或天空中飞翔)。

针对这个问题 Xiaopeng Zhang[6]等人提出了一种基于深度滤波的细粒度图像分类方法，将 CNN 的深层过滤反应作为局部描述符,并通过 SWFV-CNN 对其进行编码。第一步的目标是挑选深度过滤器,以显著和一致地响应特定的模式。第二步是通过 fisher 向量的空间加权组合来选择 CNN 滤波器。实验结果表明, SWFV 的性能优于传统的 CNN,与传统的 CNN 相辅相成,进一步提高了性能。图 4.1 为整体框架。

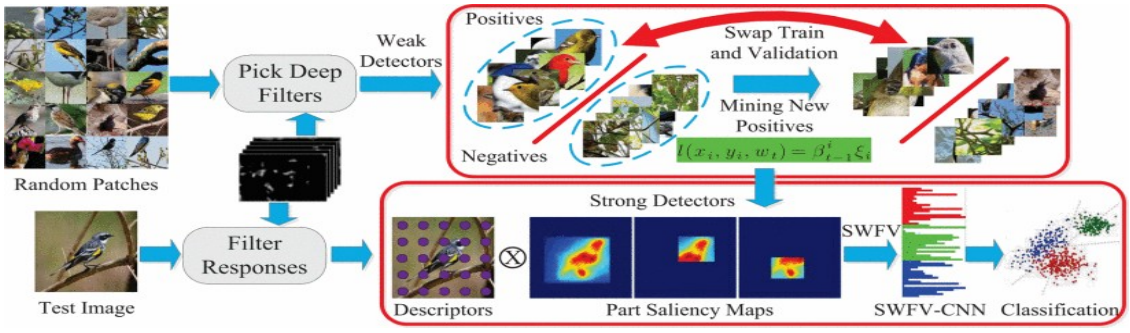


图 4.1

4.2 Bipartite-Graph Labels

大多数现有的工作都侧重于单标签分类问题，但是用标签或属性等多个标签来描述真实图像会使其更自然。根据标签结构的假设,以往关于结构标签学习的工作大致可以分为学习二进制、相对或层次属性。以前的工作大多侧重于学习二进制属性，然而，当对某些对象属性的描述是连续或模糊的时，二进制

属性是限制性的。

针对这个问题，Feng zhou[7]等人提出了一种新的方法,使用二分图标签(BGL)来建立超细粒度级之间的丰富关系。斯坦福汽车数据集上的实验结果表明，没有使用 CNN 但使用传统的基于 LLC 的表示,最好发表的结果,是 69.5%。BGL 具有预测粗标签的优点，GN-BGLm 在预测汽车图像的类型方面实现了 95.7%。总的来说，BGL 通过二分图标签联合建模细粒度和粗标签来提高传统的 softmax 损失。提出的 BGL 方法改进了以前在各种数据集上的工作。

4.3 SPDA-CNN

现有的基于 CNN 的细粒度分类方法并不侧重于对 small semantic parts 的检测和利用。所以对于鸟类幼崽的分类来说识别头部的结果总是比身体更糟糕,因为头部的尺寸更小。传统的 part-based CNN 方法对于部分网络之间的卷积过滤器的共享有困难。

来自俄罗斯大学的 Han Zhang[8]等人关注到富有语义的 parts,认识到大多数卷积神经网络缺乏模型化对象语义部分的中层。故提出了一种新的 CNN 架构, 构建了一个 End-to-End 的网络，将语义部分检测与抽象相结合进行细粒度分类。该网络有两个子网络，一个用于检测一个用于识别，该检测子网络具有一种新颖的自顶向下的生成小语义部分候选检测方法。分类子网络引入了一种新的部件层,从检测子网络检测到的部分提取特征,并结合它们进行识别。

在 CUB-2011 数据集的测试中通过直接使用经过 Oracle 局部标注训练的模型，并利用 DET-NET 检测的局部进行测试图像分类,能够达到 78.15%的准确率,仅比使用 Oracle 局部标注的测试图像的准确率低 1.31%。这种部分特定的学习为更深入地了解细粒度类别打开了门,不仅仅是识别类标签。

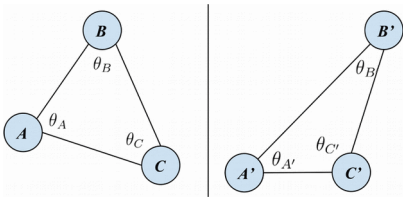
4.4 Part-Stacked CNN

细粒分类之间的细微差别主要在于对象的局部，所以一些现有的细粒度数据集提供了详细的局部注释,包括部分标记和属性。然而,它们通常与大量的对象局部相关联,这给局部检测和分类带来了沉重的计算负担。从这个角度来看,人们希望寻找一种遵循对象感知策略的方法来提供 interpretable 预测标准同时需要最小的计算努力来处理可能大量的局部。

对此 Shaoli Huang[9]等人提出了一种新颖的部分堆叠的 CNN 架构, 在一个统一的框架中对多个对象部件进行建模,提高了效率。通过从对象的部分细微差异来明确地解释有细密纹理的识别过程，但人工参与的更少。并且也是分了两个步骤。基于人工标记的 strong part 注释, 该结构由一个完全卷积网络组成,用于定位多个目标部件和一个两流分类网络,同时对对象级和部分级信号进行编码。对 CUB-200-2011 数据集的实验表明 PS 具有很高的效性和效率,特别是对引入对象部分对细化的视觉分类任务。文章所提出的 PS-CNN 架构不仅可以用于鸟类还可以使用于其他分类任务。

4.5 Mid-level patch

除了基于深度框架的细粒度方法外，Yaming Wang[10]等人提出挖掘 mid-level patch 的一种方法，这种方法只需要对象的标注框注释，属于弱注释细分类。并且他们基于 mid-level representation 的特征构建方法并结合 SVM 分类器进行分类，提出了一种基于顺序与形状条件限制的 Triplet Mining 方法。下图是 triplet 和两种限制：



本实验主要在车辆数据集(BMVC-14, Cars-196)，实验结果中表明，当没有用几何约束的方法时结果就已经优于所有的三种基线方法。当几何约束进一步增加时,这种方法不仅优于只使用显著边缘标注框，也好于使用额外的三维模型拟合的 BB-3D-G 方法。并且仅仅使用 HOG 特征的分类准确率就超过了 fine-tune 的 AlexNet。在结合深度特征的基础上，该方法取得了 state-of-the-art 的结果。

4.6 Generating parts using co-segmentation and alignment

当样本扩展到数百或数千个域时再依赖要点或 part 注解的使用，会使注释成本增大。Li[11]等人提出了一种可以在新的图像中检测局部的方法，这种方法可以学习有用的局部，并且利用 co-segmentation 对训练图像进行分割。然后，将相似的图像高密度对齐，在所有图像中进行对齐，作为这些更可靠的本地图像的组成。在 CUB-2011 的测试中，使用不同的方法进行对比后发现，他们的方法是在测试时没有使用注释的方法中最好,甚至优于所有方法。

5.未来研究方向

细粒度图像分类的应用十分广泛，所以它具有很大的房展空间。在细粒度分类问题中，现在面临很多难题，比如常见的数据集样本较少有限对于深度卷积来说是不够的。对于细粒度图像分类而言，最终的特征表示往往是由多个不同的局部区域特征组合而成。简单的特征拼接，尽管有效，但似乎并不是最佳选择。等等问题都需要进一步的研究。

References

1. Wah C, Branson S, Welinder P, Perona P, Belongie S. The Caltech-UCSD Birds-200-2011 dataset, Technical Report CNS-TR-2011-001, California Institute of Technology, USA, 2011
2. Aditya K, Nityananda J, Yao B, Li F F. Novel dataset for fine-grained image categorization. In: Proceedings of the First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition. Springs, USA: IEEE, 2011.
3. Nilsback M E, Zisserman A. Automated flower classification over a large number of classes. In: Proceedings of the 6th Indian Conference on Computer Vision, Graphics and Image Processing. Bhubaneswar, India: IEEE,2008. 722－72
4. Krause J, Stark M, Deng J, Li F F. 3D object representations for fine-grained categorization. In: Proceedings of the 4th IEEE Workshop on 3D Representation and Recognition. IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 554－561
5. Maji S, Kannala J, Rahtu E, Blaschko M, Vedaldi A. Fine-Grained Visual Classification of Aircraft [Online], available: https://arxiv.org/abs/1306.5151, Jun 21, 2013
6. Picking Deep Filter Responses for Fine-Grained Image Recognition
   Xiaopeng Zhang; Hongkai Xiong; Wengang Zhou; Weiyao Lin; Qi Tian 2016 CVPR
7. Fine-Grained Image Classification by Exploring Bipartite-Graph Labels
   Feng Zhou; Yuanqing Lin 2016 CVPR
8. SPDA-CNN: Unifying Semantic Part Detection and Abstraction for Fine-Grained Recognition
   Han Zhang; Tao Xu; Mohamed Elhoseiny; Xiaolei Huang; Shaoting Zhang; Ahmed Elgammal; Dimitris Metaxas 2016 CVPR
9. Part-Stacked CNN for Fine-Grained Visual Categorization
   Shaoli Huang; Zhe Xu; Dacheng Tao; Ya Zhang 2016 CVPR
10. Mining Discriminative Triplets of Patches for Fine-Grained Classification
    Yaming Wang; Jonghyun Choi; Vlad I. Morariu; Larr S. Davis 2016 CVPR
11. Fine-grained recognition without part annotations
    Jonathan Krause; Hailin Jin; Jianchao Yang; Li Fei-Fei 2015 CVPR