细粒度图像分类

一、 选题依据(课题来源、选题依据和背景情况；课题研究目的、学术价值或实际应用价值)

最近多年来,针对大规模图像数据集上的图像分类一直是研究的热点。图像分类是我们让计算机理解图像关键步骤,这也意味着对图像分类的研究具有广泛的应用价值。传统的词袋模型(Bag-of-Features(BoF))框架如今被广泛应用于图像表示,它通过对局部特征量化,形成特征的稀疏表示,是一种基于统计的模型结构。尽管词袋模型取得了巨大的成功,但是它仍然没有解决从低层特征到高层概念之间存在的语义鸿沟问题,而且在图像对齐上的表现也不好。近年来,研究人员针对词袋模型的不足提出了很多成功的改进办法。其中包括,提取不同种类的特征,建立中间层的特征表达,空间赋权值等等。使用这些新方法的系统不断的提高图像分类的最佳表现,但是图像的语义于图像的特征表达之间的联系仍然很弱。

幸运的是,随着神经科学的不断累积发展,研究者们发现人类是通过许多局部特征的组合来识别物体的。这就告诉我们,需要建立结构性的模型来学习高层概念。然而,传统的图片集合包含了太多不相关的高层概念,这就限制了计算机视觉算法通过有限的训练数据学习到结构性的模型。这时候图像分类的一个极具前景的分支—图像细粒度分类(Fine-Grained Visual Categorization(FGVC))进入我们的视野。在图像细粒度分类问题中,待分类的图片类别之间语义很相似,从而我们可以从整个图像数据集的共同特征中学习到更好的层次结构模型。

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。

二、 文献综述(国内外研究现状、发展动态；查阅的主要文献)

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。

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%。

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[2]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。

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[3]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。

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[4]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。

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[5]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。

除了基于深度框架的细粒度方法外，Yaming Wang[6]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。

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[7]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。

references

1.Wah C, Branson S, Welinder P, Perona P, Belongie S.
The Caltech-UCSD Birds-200-2011dataset, Technical Report CNS-TR-2011-001, California Institute of Technology, USA, 2011

2. Picking Deep Filter Responses for Fine-Grained Image Recognition
Xiaopeng Zhang; Hongkai Xiong; Wengang Zhou; Weiyao Lin; Qi Tian 2016 CVPR

3. Fine-Grained Image Classification by Exploring Bipartite-Graph Labels
Feng Zhou; Yuanqing Lin 2016 CVPR

4. 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

5. Part-Stacked CNN for Fine-Grained Visual Categorization
Shaoli Huang; Zhe Xu; Dacheng Tao; Ya Zhang 2016 CVPR

6. Mining Discriminative Triplets of Patches for Fine-Grained Classification
Yaming Wang; Jonghyun Choi; Vlad I. Morariu; Larr S. Davis 2016 CVPR

7. Fine-grained recognition without part annotations

Jonathan Krause; Hailin Jin; Jianchao Yang; Li Fei-Fei 2015 CVPR

三、研究内容

1．学术构想与思路，主要研究内容及拟解决的关键问题（或技术）

针对当前细粒度图像分类的两种算法，一种是传统经典算法一种是近期新颖算法，在数据集上进行试验，将结果进行对比。由于目前该课题本身的困难性，传统的方法不得不依赖于大量的人工标注信息，严重制约了算法的实用性。因此，越来越多的算法倾向于，不再依赖人工标注信息,仅仅使用类别标签来完成分类任务，这也是该领域逐渐发展成熟的标志。其次，细粒度图像识别有别于普通图像分类任务的一个特点，是具有区分度的信息隐藏在局部区域中，所以要更有效地利用这些局部信息，高效的获取这些有用的信息，是需要解决的一部分。再有就是数据规模与精细程度都不太高,标注质量与类别数量也十分有限。由于深度学习的性能与数据库的规模呈正相性，训练图像越丰富，所能带来的性能提升越明显,实用性也越强。建更高质量的标准数据库成为了未来研究急需解决的一个问题。

因此本次打算拟解决的问题是能够自动进行局部区域检测，并采用深度卷积方法进行特征提取。

2．拟采取的研究方法、技术路线、实施方案及可行性分析

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。选择一种早期算法，发布 CUB200-2011 数据库的技术报 Wah 等人提到的。将未经处理的测试图像，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。

在早期的方法基础上，结合其他算法进行实验。选择最近的其中一种方法目前打算选择 SPDA-CNN 这种方法，特点是提出了 samll semantic parts 的生

成方法。整个网络由两个子网络组成，一个是 Detection 网络一个是 Classification 网络。

目前，准备这两种方法的研究和实验，最后至少完成一种方法的实验。

四、论文（设计）进度安排

4、5 月开始进入实验部分，对数据集的下载，算法的研究和测试

6 月整理各种方法的结果并且整理归纳对比,由此找到以往实验的不足，总结自己的心得