

基于外观的复合属性学习的细粒度识别

宋凤义 胡 太 杨 明

(南京师范大学计算机科学与技术学院, 南京, 210023)

摘 要: 由于语义可理解性及共享性, 视觉属性作为刻画对象的中间特征表示在众多领域得到了广泛应用。视觉属性学习中, 大量的人工成本用于属性定义和标注, 因此难以避免地引入了主观偏见, 属性表示的类别判别性难以保证, 尤其面对对判别性要求较高的细粒度识别任务时更为明显。复合属性符合人类认知规律以及对象复杂多模分布的事实, 从刻画对象的分布入手, 以较低廉的代价建立兼具一定描述能力及较好判别能力的特征表示, 以应对细粒度识别任务对判别特征和判别模型的较高要求。在细粒度识别代表性公开数据集 CUB 上验证了所提方法的有效性。在细粒度识别任务中, 复合属性表现出比人工定义的属性以及类别判别属性更优的性能。

关键词: 属性学习; 复合属性; 分散式表示; 细粒度识别

中图分类号: TP391.4 **文献标志码:** A

Appearance-based Complex-Attributes Learning for Fine-Grained Recognition

Song Fengyi, Hu Tai, Yang Ming

(School of Computer Science and Technology, Nanjing Normal University, Nanjing, 210023, China)

Abstract: Visual attribute as an inter-mediate representation is exploited in many applications due to its semantic understandable and generalization. However, attribute learning need great manual effort for choosing attribute taxonomy and labeling attributes instance, which inevitably introduces human bias and leads to weak discriminant ability of attributes, especially for fine-grained recognition where high discriminant ability is crucial for recognizing subtle distinctness. Motivated by human cognition and the fact of multi-mode distribution of object, the proposed complex attributes try to model the distribution of object explaining various appearance factors and finally forms a distributed representation with better describable and discriminant ability, which will competent for handling high discriminant requirement for fine-grained recognition. Some experiments are conducted on publicly available fine-grained datasets CUB. Results show it has better performance than the handcrafted attributes and also holds simple category discriminant attributes.

Key words: attribute learning; complex attribute; distributed representation; fine-grained recognition

引 言

视觉属性作为对象的一种中间特征表示, 由于具有语义可解释性和可理解性、编码高效性以及应对

复杂噪声变化的稳定描述能力,使得其在场景理解^[1]、人脸验证^[2]、对象识别^[3, 4]和动作识别^[5]等众多领域中得到广泛应用。另外,视觉属性的跨类别刻画能力,如刻画颜色、形状和纹理等信息的视觉属性在不同类别上的含义近似相同,使得其特别适用于样本奇缺情况下的视觉辨识任务,如,零样本学习^[4]、视觉属性学习吸引了来自计算机视觉领域研究者的广泛关注。视觉属性学习的关键目标是使得学习到的属性兼具较好的描述能力和判别能力。这两个目标决定了属性学习不是一件简单的事情,尤其在面临细粒度识别问题时。细粒度识别^[6]是计算机视觉领域中最近受到广泛关注的一类典型的多类别辨识任务,涉及到一个大类下的多个子类别之间的区分和辨识,比如,CUB数据集^[7]中建立了200种鸟类识别的任务,Stanford Dog^[8]中建立了120种狗的识别任务。而辨识不同种类的鸟、狗,不同于人脸识别任务中辨识不同身份的人,因为前者中相同子类别对象具有多个个体存在,而后者中同样身份的人只有一个个体。可以看出,相同的子类具有多个个体,决定了子类内的较大差异度,而不同子类隶属于一个相同的大类(如鸟、狗),使得子类间具有较高混淆度。由此,子类内的较大差异度以及子类间的较高混淆度,使得待辨识子类别难以区分。这种门类较近的子类别识别任务被抽象归结为细粒度识别问题。面向细粒度识别的属性学习中,提升属性的描述能力和判别力成为解决问题的关键。围绕属性定义、属性抽取及属性利用的一系列属性学习问题被提出和研究。属性定义涉及到如何构建属性集,即用多个属性来刻画目标对象,类似于标记学习^[9]。比如,非控制人脸识别数据集LFW^[10]上使用73个属性来刻画人脸,动物识别数据集AWA上^[4]使用了85个属性来刻画20种动物,鸟类识别数据集CUB^[7]上使用312种属性来刻画200种鸟。通常情况下,属性集由专家依据领域知识人工设定。视觉属性从不同角度、不同部位来刻画目标对象,大体上涵盖颜色、纹理和形状等信息,通常刻画对象的局部部件信息、整体环境信息以及高层抽象语义特性。以上所述属性作为一种刻画对象视觉信息的手段,如盲人摸象般从多角度、多方位来刻画目标对象的视觉信息,称之为单一属性。这种依赖领域知识由人工定义的单一属性,存在如下几个方面的不足:(1)属性定义和属性样本标注需耗费大量的人工成本。(2)容易引入人工定义的偏见性,难以保证属性的良好判别能力。(3)属性视觉证据存在模糊性,增加了属性计算的难度,同时属性之间的复杂关系增加了属性学习的复杂性。针对单一属性中存在的问题,研究者针对其中某个或几个问题提出了相应的处理方法。为了解决耗费巨大人工代价的属性定义和标注问题,可以尝试借助普遍存在的第三方信息来自动构造属性集,如从对象的自然语言描述中,利用文本挖掘技术自动构建描述对象属性集的方法^[11],利用电子商务网站中的商品描述自动建立商品对象的描述属性集的方法^[12],利用WordNet词典中的实体语义结构自动建立对象的描述属性集的方法^[13],采用交互方式降低属性标注同时提升自动定义属性语义可理解性的方法^[14, 15],针对属性视觉证据模糊所进行的特征选择方案^[2],最大化类别似然比的属性提取方案^[16],借助于多示例技术来应对属性证据出现的位置和尺度不确定的难题^[17],利用姿态及部件结构信息先验来降低属性证据寻找范围的方案^[18-20]和针对属性之间复杂关系交织提出的融入属性关系考虑的属性学习和利用方案^[21, 22]。

本文不同于以上方法,提出了基于外观的复合属性来提升细粒度识别中属性表示的描述能力和判别能力。在人类视觉认知中,当描述一个对象(尤其是陌生对象),通常不直接、简单地描述其具有哪些属性(对应单一属性),而是描述为“这只鸟(信天翁)浮在水上像鸭子,飞起来像海鸥”。其中蕴含着一些融合了多种属性特点的描述,但又不是简单地进行多属性组合,更多地体现为一种能够涵盖一堆相似样本构成的群体公共模式的特征,这构成了复合属性的内涵(见图1)。概括而言,复合属性旨在挖掘典型群体公共模式,使得学习到的属性表示具有较好的描述能力,这些公共模式不局限于解释外观变化的单一因子,如颜色、纹理和形状,而是具有更高语义抽象级的复合属性(模板),更多地反映出具有代表性的分散式外观信息。其中,群体发现可借助于无监督聚类技术实现,同时结合利用类别判别信息提升属性编码的判别能力,期望建立的复合属性表示能较好地提升细粒度识别的性能。

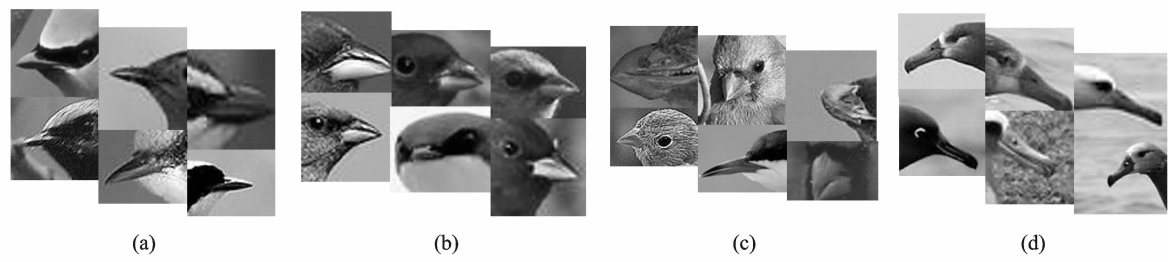


图 1 复合属性的内涵示意图(典型群体公共模式)

Fig. 1 Meaning of complex attributes (typical group common patterns)

1 属性学习和细粒度识别

1.1 属性学习

属性学习中除了依据领域知识进行人工定义属性并结合属性标注进行属性学习的常规思路之外^[2],也出现了一些尝试降低属性学习的人工代价、提升学习的判别性和规模可扩展性的思路方法^[4, 23]。典型的工作包括 Lampert^[4]提出的判别属性,尝试学习出两组类别间的视觉辨识证据,例如,在动物类别和植物类别之间学习出一种分类器作为辨识这两组类别的视觉信息编码器。文献^[23]则提出了两两类别之间的判别分类器,在人脸验证任务中取得了较好的效果,随后把这种想法细化到局部特征上建立两两类别识别器作为特征编码方式^[24],在细粒度识别任务上取得了较好的效果,但该方法依赖于较精准的特征点对齐预处理,同时在两两类别之间学习一个编码器导致最终的样本编码维度较高。相对属性^[25]则从属性标注的方式上降低详细量化属性值的难度,即,不是简单在一个样本上评测其精确属性值,而是在一组样本上评测出属性值的相对高低。这些方式很大程度上降低了属性标注所耗费的人工成本。属性的描述能力和判别能力也是属性学习的关键问题。属性表示相对于低层特征表示最大区别在于建立对象的分散式刻画,比如刻画空间上的局部外观信息,如眉毛、鼻子和嘴巴等;刻画不同特征通道上的信息,如颜色;刻画不同类型的外观信息,如形状、纹理,这些属性大多可归结为单一属性。另一方面,从刻画对象的分布规律上建立对象外观的分散式表示,具有比单一属性更好的描述能力。典型的方法有“明喻(simile)”属性^[2],即用来刻画典型个体对象的典型特征,比如“安吉丽娜·朱莉的嘴巴”以及“奥巴马的嘴巴”。这种表示方式既有较好的刻画对象分布能力,也兼具较好的语义抽象级。

1.2 细粒度识别

自然界中对象类别的划分呈现出典型的层级树形结构。父节点代表较粗概念类别,如鸚鵡;子节点代表较细概念的子类别,如虎皮鸚鵡、牡丹鸚鵡和金刚鸚鵡。子类别涵盖一类具有生物学分类或人工分类依据的多个个体而不是单个个体,这一点区分于人脸身份辨识。细粒度识别中,子类别样本表现在多个个体上具有较大的固有外观差异,复杂的非刚体变形更增加了视觉外观上的差异性。另外,由于隶属于一个同宗的大类,不同子类别间也具有较强的混淆性,这些都增加了细粒度识别的难度。本文重点讨论基于视觉信息的子类别识别,围绕具体的应用问题可以采用不同的信息来源作为辨识依据^[26]。解决细粒度识别问题的思路主要有局部特征的精确抽取、判别特征和模型的学习。前者从自身表示的角度入手,从空间几何结构和视觉外观分布上建立子类别精细特征的抽取;后者关注判别特征表示和判别模型学习,来最大化类别间区分度,最终提升细粒度识别的准确性。典型的方法包括:文献^[27]挖掘具有判别力的局部图像块,考虑在特征利用层面上提升细粒度模型的判别性;文献^[28]用有语义部件的位置信息来提升细粒度识别的性能;文献^[29]利用多种信息以强监督的方式来进行细粒度识别,其他方式包括用深度特征学习^[30]、度量学习^[31]来应对细粒度识别中高判别性的要求。

2 基于外观的复合属性学习

相对于单一属性,所提出的复合属性旨在捕获典型群体中蕴含的复杂公共模式。从刻画对象分布的角度,建立对象的分散式分布刻画,并引入类别判别信息来提升复合属性的判别性。

2.1 复合属性的内涵

如图 1 所示,复合属性定义在样本集合之上,每个集合中蕴含着一种复合属性。复合属性的内涵在于相应集合中样本的典型外观分布。给定了承载复合属性的样本集之后,复合属性的含义也清晰了。具体地,可以采用人工交互的方式获取复合属性的语义含义。图 1 通过聚类技术获取鸟头部的典型外观簇,其中簇 1(图 1(a))可以描述为“头部呈现出纯色纹理,生长在绿色植被环境中的黑色细嘴鸟”,簇 2(图 1(b))可以描述为“头部呈现蓝色纯色纹理的短嘴鸟”,簇 4(图 1(d))可以描述为“头部呈现灰色,活动在水域岩石附近具有扁长前端下钩嘴形的鸟”,等。从特征表示角度,清晰的语义含义便于人们理解特征表示的含义且有助于设计良好交互界面的图像检索系统。从最终识别任务的角度,是否明确复合属性的语义含义并不影响后续任务的识别精度。复合属性内涵具有人类认知规律上的可解释性。具体而言,人类描述对象时,除了使用单一属性之外,还会使用一些更高语义抽象级且带有复杂指代模板的描述方式,比如描述一个人具有一张娃娃脸,一个女孩具有典型的东方女性气质。这些描述传递出很强的语义含义,但不同于单一属性从单个因素来刻画对象,更多地体现出一堆典型样本群体中所蕴含的公共模式,这正是复合属性的内涵所在,考虑到可计算问题则定义在对应的承载样本集合上。复合属性能够满足刻画复杂对象分布的内在需求。计算机视觉领域中分析的视觉对象的分布通常很复杂,一方面源于自身丰富个体的内在多样性,一方面源于姿态、视角、光照和遮挡等外部复杂环境因素。为了更好地刻画对象的分布,分散式表示是一种很好的解决方法。每个复合属性都可以看作是对对象分布的一种局部刻画,这区别于单一属性刻画对象的局部特征。复合属性实现了从分布刻画的角度建立对象复杂分布的分散式刻画,同时也不失属性的语义可理解性。

2.2 复合属性表示学习

复合属性表示学习中,收集、建立具有一定相似程度的样本堆成为关键。无监督聚类技术具有以较低成本发现相似样本的能力,最适合用来构建复合属性的承载样本集合,因此将复合属性定义在聚类簇上成为一个可行的方法,如图 1 所示。其次,要考虑以何种方式来刻画聚类簇中的样本分布规律。经典的刻画方式可以分作生成模型和判别模型,其中在面临小数据量情况下,判别模型表现出较好的优势。在判别模型训练中,把聚类簇中的样本看作正类,负类可以直接设定为其他聚类簇。这种做法是完全的无监督学习方式,其判别性能难以保证,为此可以引入类别信息,对于每个聚类簇依次引入各个类别的样本作为负类,同时舍弃聚类簇中该类别的样本,由此构建多个二值分类器用来刻画簇内的样本分布特点。具体地,假设聚类个数为 M ,类别数目为 K ,则总共需要学习 $M \times K$ 个分类器,分别刻画聚类簇相对于每个类别(作为负类)的分布规律。随后,这些分类器用作编码器,可以把任意样本投影到该编码空间,即,建立了样本的新表示($M \times K$ 维),然后基于该表示训练最终的类别分类器。多类分类情况下,使用 one-vs-all 的线性支持向量机作为分类模型。算法的详细过程整理如下。

输入: N 个训练样本及类标(K 类)。

输出:复合属性提取器及类别分类性能。

- (1)对训练样本进行聚类,聚成 M 个簇 $\{C_1, C_2, \dots, C_M\}$
- (2)for $i=1:M$
- (3)for $j=1:K$
- (4)选择第 i 个簇中的样本为正类

- (5)舍弃簇中属于第 j 类的样本
- (6)选择第 j 类的所有训练样本为负类
- (7)训练线性 SVM 分类器作为编码器
- (8)end
- (9)end
- (10)在所有样本上计算 $M \times K$ 个分类器的响应值。
- (11)基于响应值的新编码表示,训练 one-vs-all 线性 SVM 类别分类器。
- (12)在测试样本上评估分类性能。

3 实 验

为了验证提出的复合属性学习方法在细粒度识别中的有效性,实验选择在典型的细粒度识别公开数据集 Caltech-UCSD Birds 200-2011(CUB)^[7]上展开。

3.1 数据集和实验协议

CUB数据集^[7]建立了鸟的子类别辨识任务,共收集了11 788张鸟的图片,涵盖200种不同种类的鸟,每张图片上给出的标注信息包括:类别、局部部件位置、属性和对象框位置。其中,类别涵盖200种鸟,包括黑足信天翁、北美红雀和灰猫鹊等,定义了15个局部部件,包括背部、眼部、喙和翅膀等,定义了312种属性来刻画15个局部部件的特性,包括颜色、形状等,并在样本级上进行了属性标注。CUB数据集提供了标准的训练数据和测试数据划分协议,共分作5 994张训练样本和5 794张测试样本。训练样本和测试样本类别上的样本数保持均衡,平均每类30个样本分别用于训练和测试。本文实验设置遵从该标准实验协议,同时确保所提方法涉及到的聚类、编码分类器训练以及鸟子类别分类器的训练过程都在训练样本上进行,最终在标准测试集上汇报细粒度识别性能。本文采用平均类别识别正确率作为评估多类别识别性能的指标。另外,考虑到所提出的复合属性重在刻画典型的分布,不依赖于图像的具体底层特征表示,因此为了提高图像表示的高效性,采用数据库提供的人工定义和标注的312种属性值来表示图像。

3.2 基准算法

与本文方法最相关的工作 Tom-vs-Peter 编码器的方法^[23]尝试在两两类别之间学习判别分类器作为图像特征的编码器,期望捕捉能够最大化任意两类分类间隔的关键视觉信息,在人脸验证和鸟类识别上均取得了较好的性能。考虑到不同的实验细节会严重影响最终的性能结果,本文依照 Tom-vs-Peter 编码器的想法,在遵从与本文相同的实验设置条件下,重新评估了模型性能,使得算法比对在相对公平的条件下进行,从而对不同算法的效果获得较明确的认识。考虑到维度问题,编码器采用了 one-vs-all 而非 Tom-vs-Peter 编码器^[23]采用的 one-vs-one 方式,最终的鸟类子类别分类器也采用 one-vs-all 的多分类器设计方案。考虑到本文提出的复合属性的落脚点在于建立一种特征表示——兼具较好的语义描述能力和类别判别能力,由此直接使用 CUB 数据集上提供的人工定义和标注的属性表示向量作为图像的特征表示,成为本文要比对的另一基准算法,类别分类器同样采用 one-vs-all 的多分类器设计方案。

3.3 实验结果

首先对比分析了所提方法与两个基准算法之间的性能差别,详细结果见图2。可以看出本文方法表现出比两种基准算法更突出的性能。在聚类簇个数设置为15的情况下,本文方法取得了45.10%的性能,相对于CUB数据集上提供的人工定义和标定的属性表示作为图像特征编码的41.34%的性能提升了约4%,相对于基于 Tom-vs-Peter 类别判别编码方案36.19%的性能提高了约9%。本文方法表现出的相对于两种典型基准算法较突出的性能验证了所提方法的有效性,表明了所提复合属性从判别能

力上优于 Tom-vs-Peter 的纯类别判别编码的后续识别性能,也优于人工定义和标注的属性编码的后续识别性能。另外,还验证了复合属性个数(聚类个数)对性能的影响,从图 3 可以看出,在较小的个数基数下,性能逐渐提升,当到达一定程度之后性能难以继续提升。因为:(1)编码长度随着复合属性个数的增强表现出倍数扩增趋势,应用到更大复合属性个数下会造成最终类别分类器的维数太大,增加了模型的训练复杂度。(2)复合属性个数增加导致每个聚类中的样本数减少(CUB 上训练样本共 5 994 个),难以充分刻画对象的分布特点,同时从模型机学习的角度,也容易造成编码器训练的过拟合,难以保证在测试样本上的编码可靠性。

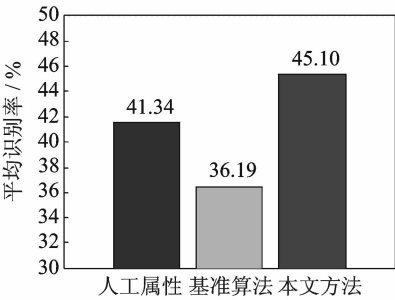


图 2 本文方法与两种典型的基准算法的性能对比

Fig. 2 Comparison of proposed method with two typical benchmarks

最后考虑到复合属性的定义及内涵依赖于具体的承载数据集,也会影响所构建的复合属性作为特征编码方案应用到后续识别任务中的性能。为了验证这一点,实验测试了一种极端情况——以随机方式划分出数据子集,然后采用复合属性的学习方法学习出图像编码器,继而在后续识别任务中验证图像编码方案的性能。需要指出的是,这里获得的编码器不再具有复合属性的语义,承载数据集集中的样本呈现出无明显规律状态。从图 4 可以看出,在不同的子集个数下(对应图 3 中的聚类个数),所获得的图像编码方案在后续识别任务的性能均低于图 3 中的复合属性性能。另外,识别性能也呈现出随着子集个数增加而增加的趋势,其得益于本文方法中的类别判别信息的引入,但此时所得编码器完全没有本文所提复合属性的含义,唯一可解释的是编码器训练过程中的负类样本(在本文中为一个纯净的子类别)成为分类器要刻画的重要内容,而且也可以看作是对对象分布的分散式刻画,其性能优于 Tom-vs-Peter 的两两类别判别编码方案 36.19% 的性能,也表明了对象分布的分散式刻画的有效性。

4 结束语

细粒度识别问题作为典型的多类别分类问题,由于其问题的复杂性,需要更具判别力的特征表示和模型学习方案。属性表示在计算机视觉任务中表现出语义可理解性优势,符合人类视觉认知习惯,但面对细粒度识别任务时

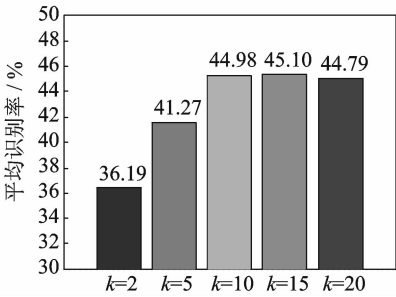


图 3 本文方法性能受复合属性个数的影响

Fig. 3 Performance of proposed methods affected by complex attributes number

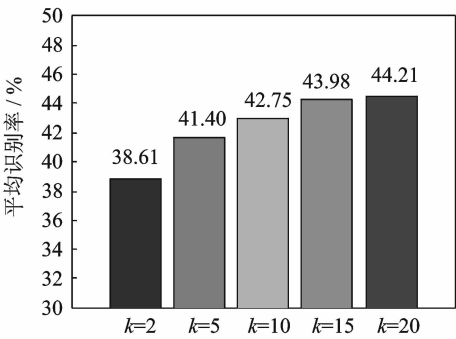


图 4 随机簇划分下的图像编码方案用于细粒度辨识的性能及其受不同簇个数的影响

Fig. 4 Performance of fine-grained recognition accuracy with image encoding learned under random clusters, and the performance tendency with increasing number of clusters

仍难以保证较好的性能。本文提出的复合属性尝试从对象分布的分散式刻画入手,区别于通常单一属性刻画对象的局部特征方式,在细粒度识别任务中表现出较好的性能。所提复合属性的学习方法关键在于融入了两个策略:(1)构建复合属性的承载数据集具有视觉语义上的可理解性,同时形成了对象分布的分散式刻画。(2)学习复合属性的方式引入了类别判别信息,同时为每个复合属性设定多个参照负类样本集,进一步强化了对象的分散式刻画。最后,细粒度识别任务上的实验也验证了所提方法的有效性。另外,考虑到细粒度识别中子类间辨识的关键在于局部特征判定上,因此,结构分析也是提升细粒度识别问题中的关键,这将是本文下一步的工作。

参考文献:

- [1] Patterson G, Xu C, Su H, et al. The sun attribute database: Beyond categories for deeper scene understanding[J]. International Journal of Computer Vision, 2014, 108(1/2): 59-81.
- [2] Kumar N, Berg A, Belhumeur P N, et al. Describable visual attributes for face verification and image search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(10): 1962-1977.
- [3] Farhadi A, Endres I, Hoiem D, et al. Describing objects by their attributes[C]//Computer Vision and Pattern Recognition. Florida: IEEE, 2009: 1778-1785.
- [4] Lampert C H, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer[C]//Computer Vision and Pattern Recognition. Florida: IEEE, 2009: 951-958.
- [5] Liu J, Kuipers B, Savarese S. Recognizing human actions by attributes[C]//Computer Vision and Pattern Recognition. Colorado: IEEE, 2011: 3337-3344.
- [6] Ling X, Weld D S. Fine-Grained Entity Recognition[C]// AAAI Conference on Artificial Intelligence. Toronto: AAAI, 2012: 94-100.
- [7] Wah C, Branson S, Welinder P, et al. The caltech-UCSD birds-200-2011 dataset[EB/OL]. <http://www.vision.caltech.edu/visipedia/cub-200-2011.html>, 2011-11-6.
- [8] Khosla A, Jayadevaprakash N, Yao B, et al. Novel dataset for fine-grained image categorization: Stanford dogs[C]//Proc CVPR Workshop on Fine-Grained Visual Categorization (FGVC). Golorado: IEEE, 2011: 2.
- [9] 张敏灵. 偏标记学习研究综述[J]. 数据采集与处理, 2015, 30(1): 77-87.
Zhang Minling. Research on partial label learning[J]. Journal of Data Acquisition and Processing, 2015, 30(1): 77-87.
- [10] Learned-Miller E, Huang G B, RoyChowdhury A, et al. Labeled faces in the wild: A survey[M]. Advances in Face Detection and Facial Image Analysis. Switzerland: Springer International Publishing, 2016: 189-248.
- [11] Wang J, Markert K, Everingham M. Learning models for object recognition from natural language descriptions[C]//British Machine Vision Conference. London: British Machine Vision Association, 2009: 2-11.
- [12] Berg T L, Berg A C, Shih J. Automatic attribute discovery and characterization from noisy web data[C]//European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2010: 663-676.
- [13] Rohrbach M, Stark M, Szarvas G, et al. What helps where and why? semantic relatedness for knowledge transfer[C]//Computer Vision and Pattern Recognition (CVPR). San Francisco: IEEE, 2010: 910-917.
- [14] Parikh D, Grauman K. Interactively building a discriminative vocabulary of nameable attributes[C]//Computer Vision and Pattern Recognition (CVPR). Colorado: IEEE, 2011: 1681-1688.
- [15] Duan K, Parikh D, Crandall D, et al. Discovering localized attributes for fine-grained recognition[C]//Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2012: 3474-3481.
- [16] Ferrari V, Zisserman A. Learning visual attributes[C]//Advances in Neural Information Processing Systems. Hyatt Regency Vancouver: MIT Press, 2007: 433-440.
- [17] Wang G, Forsyth D. Joint learning of visual attributes, object classes and visual saliency[C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 537-544.
- [18] Bourdev L, Maji S, Malik J. Describing people: A poselet-based approach to attribute classification[C]//2011 International Conference on Computer Vision. [S.l.]: IEEE, 2011: 1543-1550.
- [19] Chen H, Gallagher A, Girod B. Describing clothing by semantic attributes[C]//European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012: 609-623.
- [20] Livne M, Sigal L, Troje N F, et al. Human attributes from 3D pose tracking[J]. Computer Vision and Image Understanding, 2012, 116(5): 648-660.
- [21] Wang Y, Mori G. A discriminative latent model of object classes and attributes[C]//European Conference on Computer Vi-

sion. Berlin, Heidelberg: Springer, 2010: 155-168.

[22] Song F, Tan X, Chen S. Exploiting relationship between attributes for improved face verification[J]. Computer Vision and Image Understanding, 2014, 122: 143-154.

[23] Berg T, Belhumeur P N. Tom-vs-Pete classifiers and identity-preserving alignment for face verification[C]//British Machine Vision Conference. Guildford: British Machine Vision Association, 2012: 1-7.

[24] Berg T, Belhumeur P. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 955-962.

[25] Parikh D, Grauman K. Relative attributes[C]//2011 International Conference on Computer Vision. [S.l.]: IEEE, 2011: 503-510.

[26] 孙斌, 万鹏威, 陶达, 等. 基于自适应最优核时频分布的鸟类识别[J]. 数据采集与处理, 2015, 30(6): 1187-1195.
Sun Bin, Wan Pengwei, Tao Da, et al. Identification of birds based on adaptive optimal kernel time-frequency distribution [J]. Journal of Data Acquisition and Processing, 2015, 30(6): 1187-1195.

[27] Wang Y, Choi J, Morariu V I, et al. Mining discriminative triplets of patches for fine-grained classification[C]//Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1163-1172.

[28] Zhang Han, Xu Tao, Elhoseiny M, et al. SPDA-CNN: Unifying semantic part detection and abstraction for fine-grained recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1143-1152.

[29] Akata Z, Malinowski M, Fritz M, et al. Multi-cue zero-shot learning with strong supervision[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 59-68.

[30] Reed S, Akata Z, Schiele B, et al. Learning deep representations of fine-grained visual descriptions[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 49-58.

[31] Cui Y, Zhou F, Lin Y, et al. Fine-grained categorization and dataset bootstrapping using deep metric learning with humans in the loop[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1153-1162.

作者简介:



宋凤义 (1982-), 男, 博士, 讲师, 研究方向: 图像理解、对象识别、属性学习和零样本学习, E-mail: f. song @ njnu. edu. cn.



胡太 (1993-), 男, 硕士研究生, 研究方向: 图像理解。



杨明 (1964-), 男, 教授, 博士生导师, 研究方向: 机器学习, 代价敏感学习。

