JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
12
VOL.25

ISSN1006-8961
CN11-3758/TB

数字浮雕建模 *P2647*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第12期（总第296期）

2020年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

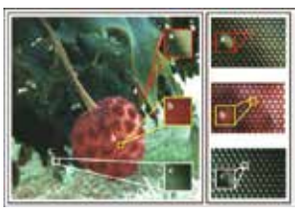
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

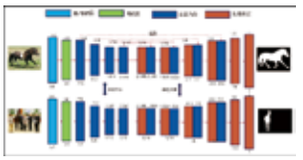
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年12月 第25卷 第12期（总第296期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

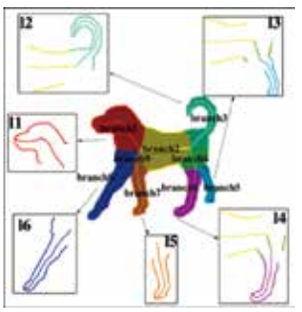
目次



光场显著性检测研究综述(第2465页)



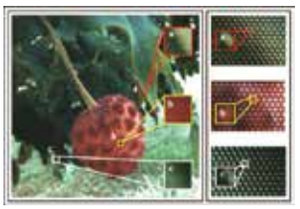
融合注意力机制与知识蒸馏的孪生网络压缩(第2563页)



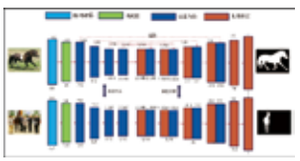
面向可解释性的物体拓扑结构骨架表征方法(第2587页)

Volume 25, Number 12
Published December 16, 2020

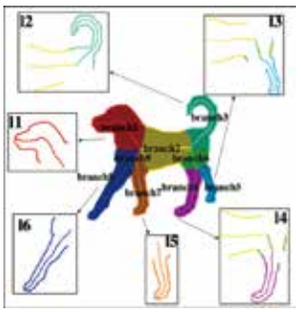
CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Review of saliency detection on light fields(P2465)



Combining attention mechanism and knowledge distillation for Siamese network compression (P2563)



Skeleton characterization of object topology toward explainability (P2587)

中图法分类号:TP18　　**文献标识码**:A　　**文章编号**:1006-8961(2020)12-2587-16

论文引用格式:Wei H and Yu L P. 2020. Skeleton characterization of object topology toward explainability. Journal of Image and Graphics,25(12):2587-2602(危辉,余莉萍. 2020. 面向可解释性的物体拓扑结构骨架表征方法. 中国图象图形学报,25(12):2587-2602)[DOI:10.11834/jig.190661]

面向可解释性的物体拓扑结构骨架表征方法

危辉,余莉萍

1. 复旦大学计算机科学技术学院认知算法模型实验室,上海　201203;2. 上海市数据科学重点实验室,上海　201203

摘　要:**目的**　模式识别中,通常使用大量标注数据和有效的机器学习算法训练分类器应对不确定性问题。然而,这一过程缺乏知识表征和可解释性。认知心理学和实验心理学的研究表明,人类往往不使用代价如此巨大的机制,而是使用表征、归纳、推理、解释和约束传播等与符号主义人工智能方法类似的手段来应对物体识别中的不确定性并提供可解释性。因此本文旨在从传统的符号计算出发,利用骨架拓扑结构表征提供一种可解释性的思路。**方法**　以骨架树为基本手段来形成物体拓扑结构特征和几何特征的形式化表征,并基于泛化框架对少量同类表征进行知识抽取来形成关于物体类别的知识概括显式化表征。**结果**　在形成物体类别的概括表征实验中,通过路径重建直观展示了同类属物体上得到的最一般表征的几何物理意义。在可解释性验证实验中,通过跨数据的拓扑应用展示了新测试样本相对于概括表征的特定差异,表明该表征具有良好的可解释性。最后在形状补全的不确定性推理实验中,不仅可以得到识别结论,而且清晰展示了识别背后做出的判断依据,进一步验证了该表征的可解释性。**结论**　实验表明一般化的形式表征能够应对尺寸、颜色和形状等不确定性问题,本文方法避免了基于纹理特征所带来的不确定性,适用于任意基于基元的表征方式,具有更好的鲁棒性、普适性和可解释性,计算代价更小。

关键词:视觉任务;骨架;不确定性;拓扑知识表征;可解释性

Skeleton characterization of object topology toward explainability

Wei Hui, Yu Liping

1. Laboratory of Cognitive Modeling and Algorithms, School of Computer Science and Technology, Fudan University, Shanghai 201203, China; 2. Shanghai Key Laboratory of Data Science, Shanghai 201203, China

Abstract: **Objective** Understanding the shape and structure of objects is extremely important in object recognition. The most commonly utilized pattern recognition method is machine learning, which often requires a large number of training data. However, this object-oriented learning method lacks a priori knowledge, uses a large amount of training data and complex computations, and is unable to extract explicit knowledge after learning (i.e., "knowing how without knowing why"). Great uncertainties are encountered in object recognition tasks due to changes in size, color, illumination, position, and environmental background. To deal with such uncertainties, a large number of samples should be trained and powerful machine learning algorithms should be used to generate a classifier. Despite achieving a favorable recognition accuracy in some standard datasets, these models lack explainability, and recent studies have shown that these purely data-driven mod-

收稿日期:2019-12-19;**修回日期**:2020-04-30;**预印本日期**:2020-05-07

基金项目:国家自然科学基金项目(61771146,61375122);上海市科学技术发展基金项目(13dz2260200,13511504300)

Supported by:National Natural Science Foundation of China (61771146, 61375122)

els are vulnerable. These models also often ignore knowledge representation and even consider this aspect redundant. However, cognitive and experimental psychology research suggests that humans do not adopt such mechanism. Similar symbolic artificial intelligence methods, such as representation, induction, reasoning, interpretation, and constraint propagation have also been used to deal with the uncertainties in object recognition. In vision tasks, improving explainability is considered more important than improving accuracy. Such is the goal of interpretable artificial intelligence. Accordingly, this paper aims to provide an interpretable way of thinking from the traditional symbolic computing idea and adopts the skeleton topology representation. **Method** Psychological research reveals that humans show strong topological and geometric preferences when engaged in visual tasks. To explicitly characterize geometric and topological features, the proposed method adopts skeleton descriptors with excellent topological geometric characteristics. First, an object was decomposed into several connected components based on a skeleton graph, and each component was represented by a skeleton branch. Second, the statistical parameter of these skeleton branches were obtained, including their area ratio, path length, and skeletal mass ratio distribution. Third, the skeleton radius path was used to describe the contour of the object. Fourth, to form a robust spatial topology constraint, the spine-like axis (SPA) was used to describe the spatial distribution of shape components. Finally, a skeleton tree was used to represent the topological structure (RTS) of objects. A similarity measure based on RTS was also proposed, and the optimal subsequence bijection (OSB) was used for the elastic matching of object shapes. A multi-level generalization framework was then built to extract knowledge from a small number of similar representations and to subsequently form a generalized explicit representation (GRTS) of the object categories. The uncertainty reasoning and explainability were verified based on certainty factor theory. **Result** The proposed model illustrates the process of generating GRTS on the Kimia99 and Tari56 datasets and presents the physical meanings of most general representations obtained from homogeneous objects. The skeletal path of an object of the same category was used for reconstruction to clearly describe the object meaning of each part of GRTS. In the explainability verification experiment, the GRTS of several categories obtained from the Tari56 dataset was used to apply the topological character to a sample of Kimia99's closest category to discover the specific differences of the new test sample relative to the GRTS. Results show that the representation has good explainability. Meanwhile, in the shape complementing experiments, RTS was initially extracted from incomplete shapes to gather evidence, and the uncertainty reasoning was validated with the rule set (Tari56) that was established according to GRTS. The proposed model only provided are cognition conclusion and showed specific judgment basis, thereby further verifying the explainability of the representation. **Conclusion** A skeleton tree was used as the basic means for generating a formal representation of the topological and geometric features of an object. Based on the generalization framework, the knowledge extracted from a small number of similar representations was used to form a generalized explicit representation of the knowledge about an object category. The knowledge representation results were then used to conduct uncertainty reasoning experiments. This work presents a new perspective toward explainability and helps build trust-based relationships between models and people. Experiment results show that the generalized formal representation can cope with uncertainties in size, color, and shape. This representation also has strong robustness and universality, can prevent uncertainties arising from texture features, and is suitable for any primitive-based representation. The proposed approach significantly outperforms the mainstream machine learning methods in terms of explainability and computational cost.

Key words: vision task; skeleton; uncertainty; topological knowledge representation; explainability

0 引 言

物体识别的最大挑战是不确定性,物体尺寸、颜色、视角、光照以及背景变化造成了多种可能性。主流做法是用大量带标签的样本数据训练一个由神经网络或支持向量机(Adankon 和 Cheriet 等,2014)构建的分类器。但是,深度神经网络本质上与生物神经系统实现的认知过程不同,通常不需要知识和知识表征,存在着黑盒方法的可解释性差、泛化能力差以及调参困难等一系列缺点。研究表明,深度学习对图像细微变化过于敏感(Yuille 和 Liu,2018;Geirhos 等,2018)。例如,在一幅“丛林中的猴子”的图像中,在猴子前添加自行车,会导致深度神经网络识别错误,将自行车把识别成鸟,将猴子识别成人(Yuille 和 Liu,2018)。深度神经网络作为一种数

据驱动的方法，缺乏可解释性，让人们无法轻易相信模型的预测，因此提高人工智能的可解释性十分必要。

人们基于少量样本就能顺利应对物体识别时的不确定性，是因为使用了超越样本刺激本身的属于更高层面的关于物体结构的知识信息。这暗示有必要研究一种能够解决物体识别时可变因素的不确定性知识表征方法。在符号主义人工智能框架下已存在不确定性知识表征方法，如概率逻辑（Catherine 和 Cohen，2016）、模糊逻辑（Rastogi 等，2015）以及置信度理论（Liu，2010）等，但源于纯粹形式逻辑和符号化的本质，它们所能做到的抽象化语义表达无助于对物体的拓扑与几何特征进行精细地刻画。人工智能中最困难的语义鸿沟难题就是由于形式符号难以对语义进行全面描述导致的，那么增加中间表达层次是否有助于缓解这个矛盾呢？

著名的心理学家 Biederman 于 1987 年提出，人们观察物体时会首先识别一些基本几何形状，并以此识别物体（Biederman，1987；Hubel 和 Wiesel，1959）。物体的几何特征是最为直观且易于理解的，通过几何特征识别物体，不仅实现简单，而且有更高的效率。Biederman 的理论指出，物体识别依赖于边缘信息而不是表面信息（例如颜色、纹理）（Biederman，1987；Eysenck 和 Keane，2005）。此外，Sanocki 等人（1988）指出，当物体在自身情境而不是其他物体的情境中呈现时，边缘抽取加工更会引导正确的物体识别加工。同时，对于物体的拓扑结构，基于知识驱动的物体检测以及知识表征的明确性和不确定性可用于推进知识表征的研究，以促进基于概念驱动的物体识别，例如可以潜在弥补语义鸿沟（Li 等，2015；Andreopoulos 和 Tsotsos，2013；Crevier 和 Lepage，1997）。上述事例证明物体识别本质上需要利用到几何和拓扑知识。

形状作为一种具有恒常性的拓扑几何特征，具有相较于颜色、纹理特征等更加优越的普适性和鲁棒性的特点（Kriegeskorte，2015；Kubilius 等，2016；Ritter 等，2017；Brendel 和 Bethge，2019；Geirhos 等，2018）。皮亚杰的拓扑首位理论认为，儿童在认知发展阶段，首先发展拓扑几何的空间概念，之后再发展欧氏几何的空间概念（Piaget，1957）。形状作为一种欧氏几何特征，稳定性远高于颜色、纹理等低级特征，人类在识别任务中对其表现出较强的拓扑偏好和几何偏好。研究表明，随着年龄增长，儿童逐渐从拓扑偏好过渡到形状偏好，成人则表现出较强的形状偏好（Graham 和 Diesendruck，2010；Hupp，2008；Sera 和 Millett，2011）。拓扑性质的恒常性最佳，具有较低的不确定性和最小的性质范围，稳定性最高（钟师鹏，2007）。这些事例表明人们具有稳定的形状偏好（Landau 等，1988）。

基于大规模有标签学习的深度学习在模式识别方面取得了长足进步（LeCun 等，2015；He 等，2016；Gao 等，2017；Ren 等，2017；周敏 等，2017；李玺 等，2019），但依然仅能提供有限的分类信息，并且几乎不使用显式知识，在很大程度上限制了它的泛化性能和可解释性（Goodfellow 等，2014；LeCun 等，2015）。研究发现，由于数据集存在知识偏差，极易导致深度学习出现决策偏差（Zhang 等，2017）。比如很难得知深度学习模型在正确识别一只熊猫时，是基于熊猫的本质特征还是因为背景中的竹子作出的决策。此外，对在正常测试样本上表现良好的神经网络模型，仅将少量恶意噪声叠加在测试样本上，会导致神经网络出错（Hayes 和 Danezis，2017；Nguyen 等，2015；Yuille 和 Liu，2018），因此出现了大量针对深度神经网络的攻击方法（Hayes 和 Danezis，2017；Nguyen 等，2015；Moosavi-Dezfooli 等，2016；Goodfellow 等，2014；Li 和 Li，2017；Chen 等，2018）。当人工智能系统不能真正理解它的任务时（不可解释），会导致严重后果，即使系统环境中最细小的意外偏差也可能使其出错，严重限制了目前人工智能模型的发展。因此提高识别方法的准确率不再是最主要问题，可解释性才是症结所在，而且是一个共性前沿问题，也是人工智能的目标。

学术界和工业界针对可解释性工作进行了广泛深入的研究，提出了一系列机器学习模型的可解释性方法。一类是针对深度学习模型出现的问题提出的以可视化为主的工作，另一类针对模型本身提出更加符合人类认知的决策模型，从根本上提升模型的可解释性。在第 1 类工作中，出现了多种探索卷积神经网络（convolutional neural networks，CNN）内部隐藏语义（Zhang 等，2017；Zhang 和 Zhu，2018）的方法，提出了大量统计方法（Szegedy 等，2013；Yosinski 等，2014；Aubry 和 Russell，2015）分析 CNN 功能的特征。CNN 中滤波器的可视化（Bau 等，2017）是探索隐藏在神经单元内部模式的最直接方法。上

卷积网络(Dosovitskiy 和 Brox,2016)将学习到的特征映射转化为图像,基于梯度的可视化(Zeiler 和 Fergus,2014;Mahendran 和 Vedaldi,2015;Simonyan 等,2013)生成能够使得给定单元最大化类别置信度的图像,更接近理解深度神经网络(deep neural networks,DNN)的内部机制。Zintgraf 等人(2017)通过对 DNN 决策贡献最大的区域可视化,提供视觉解释性。CAM(class activation mapping)(Zhou 等,2016)利用 GAP(global average pooling)的作用,在保留空间信息的同时,达到定位的目的。但是由于 GAP 的限制,需在新网络的结构上重新训练模型,在实际应用中受限。Grad-CAM(Selvaraju 等,2017)和 CAM 的基本思路一致,区别在于获取每个特征图的权重时,采用梯度的全局平均计算权重,可以达到与 CAM 一样的可解释性效果,并且不受限于网络结构。第 1 类方法几乎都集中在针对深度学习中预训练网络的可解释性工作上,第 2 类方法尝试对中间层进行语义解释,但这方面的研究较少。Wu 等人(2017)利用有向无环与或图(and-or graph,AOG)模型中的层次和组合语法模型模拟物体部位标记来探索感兴趣区域(region-of-interest,RoI),并采用最佳解析树进行解释。Kindermans 等人(2017)提出胶囊网络,每个胶囊由一组神经元构建,并施以具体的实际参数意义,最终由协议路由机制进行特定语义概念的编码。

尽管以上工作在一定程度上给出了针对深度学习的特定可解释性方案,但增加了巨大的计算开销。此外,虽然这类方法定位了重要特征,但是却很难提供特征之前的相关关系对决策结果的影响。句法结构模式识别(Gonzalez 和 Thomas,1978;McMahon 和 Oommen,2017)很适合做一些解释性工作,其将对象的结构视为基元,并建立基于符号的规则集(规则,语法)来描述基元之间的关系进行知识表征,还可以提供模式的知识表示(Iwańska 等,1999;Hofmann 等,1989;Li 和 Yu,2014),从而为模式识别提供了一个简单有限的模式基元和语法规则集来描述复杂模式的可能性,缺点是难以通过提取基元来获得与模式类兼容的语法。由此可见,收集足够多的训练模式样本,并通过原始基元来获取相应的语法,在实际应用中仍然存在一定困难。

本文方法属于第二类,旨在从模型本身提供一种可解释的机制,可以提供特征之间的关系建模,从而具有更具体的语义信息。由于骨架可以反映物体的结构和拓扑特征并形成清晰的形状表征,因此具有高度的灵活性、容错性和可解释性(Blum,1973;Bai 等,2009;Shen 等,2016;Papadopoulos 等,2019;Cho 等,2019)。将骨架枝作为物体的形状基元,分层抽取物体的结构特征,从而利于从有限样本中归纳学习到物体本质的结构信息。对此,研究者提出了一些基于骨架的方法,以期望从同类样例中学习得到该类的骨架化抽象模式。Demirci 等人(2009)基于骨架节点间的多对多映射关系建立一个同类的骨架原型,在每一步成对的抽象模式中选择平均策略,但是这种方式对后面加入的新模式具有偏好,不能很好地评估同类物体结构变化。Torsello 和 Hancock(2006)通过最小编码准则找到对同类样例有最优解释能力的联合属性树模型,但是这个最优化过程容易收敛于局部最优。Shen 等人(2013)提出了一种基于接合点的共同骨架图的多层次聚类算法,可以很好地捕捉类内变化,但是其目的在于改进聚类算法,根据聚类得到的共同结构骨架图,获得更鲁棒的形状聚类间的相似度,最终得到图结构,没有上升到语义解释性。由于自然环境中存在复杂的场景,忽略颜色、纹理等低级特征,本文方法仅探究了二值图像中的物体形状表征,试图给出一种对同类属物体高层次特征的语义归纳,从而进行一些解释性的工作。这一研究有助于推动物体识别方法由经验主义向理性主义方向发展。使用骨架树表征物体的拓扑和几何特征,并基于多层级泛化框架提取知识,形成有关物体类别知识的泛化表征,可以解决诸如大小、颜色、形状之类的不确定性,避免了基于纹理特征所带来的不确定性,适用于任意基于基元的表征方式,具有更好的鲁棒性与普适性,在可解释性、计算代价等方面明显优于当前主流的机器学习方法。

本文旨在借助基于骨架拓扑结构表征提供一种可解释性的思路。受认知科学的启发,人们无需像深度学习数以亿级数量的训练集达到学习的目的(Brenden 等,2015),仅需要几个样本便可以通过归纳得到知识,并且利用这些知识去识别类似的图像,还可以给出做出判断的依据。本文以二值图像为例,利用骨架划分基元,通过多层次框架进行归纳,得到泛化表征将低级的视觉特征转化为高阶的语义知识,与基于骨架结构的工作方法不同,结合确定性

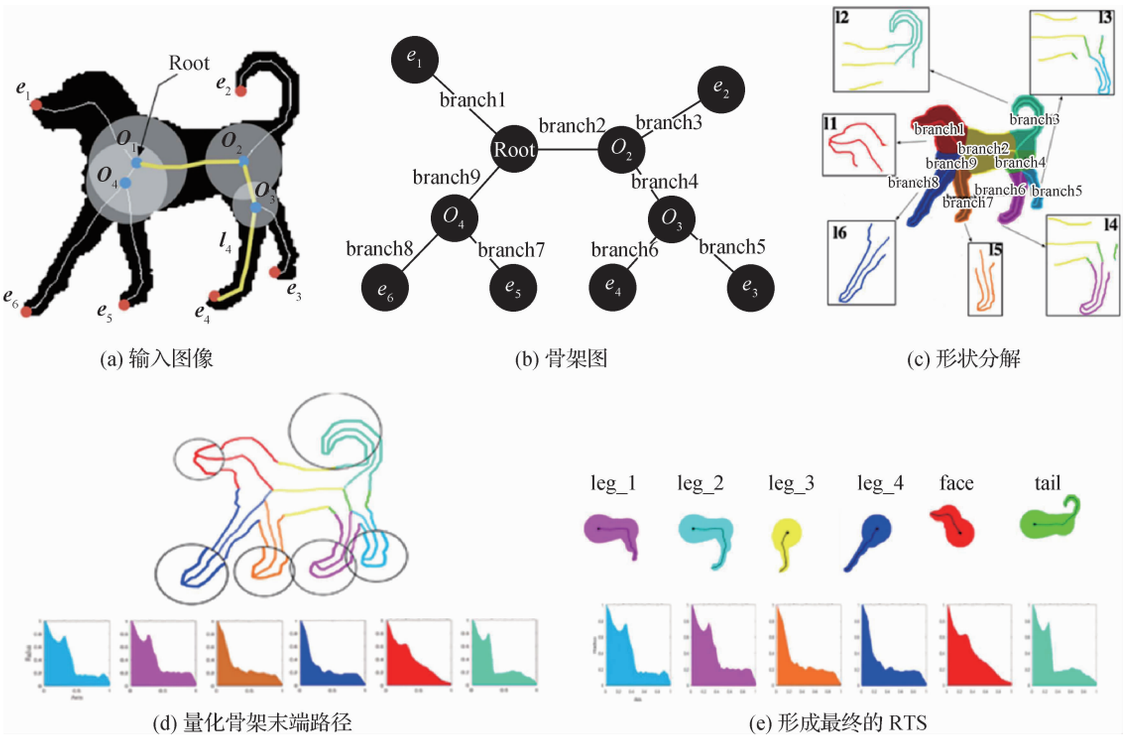
因子理论提供了一种可解释性的思路。

本文的主要贡献如下：1）提出了基于物体骨架的拓扑结构的形式化表征（representation of topological structure，RTS），详细展示了特征向量的构造过程，并引入类脊柱轴（spine-like axis，SPA）增加对表征的空间分布描述。2）提出了基于多层级归纳学习框架的 GRTS（generalized representation of topological structure）表征，利用归纳学习逐层进行学习，得到具有鲁棒性和稀疏性的泛化特征。在 Tari56 以及 Kimia99 数据集上的实验结果验证了该表征的泛化性和可解释性。3）提出了基于确定性因子理论的不确定性推理。基于 GRTS 进行语义规则导出和不确定性计算，在得到分类概率的同时，给出了显式的判断依据，提供了一种新的可解释性思路。

1　基于骨架树的表征（RTS）以及多层级泛化表征（GRTS）

1.1　基于物体骨架的结构特征的形式化表征

本文提出了一种基于骨架图的物体结构特征的形式化表征（RTS），主要思想是依据认知心理学的理论（Piaget 等，1957；Landau 等，1988），在描述物体时强调它的结构特性。首先基于骨架图将物体分解为若干个彼此连通的组成部件，每个部件由一段骨架枝表示，然后基于骨架枝的统计特征进行量化表征。RTS 的构建过程如图 1 所示。



(a) 输入图像　(b) 骨架图　(c) 形状分解

(d) 量化骨架末端路径　(e) 形成最终的 RTS

图 1　RTS 的构建过程

Fig. 1　RTS establishment

((a) input image; (b) skeleton graph; (c) shape decomposition; (d) vectorizing the skeleton end path; (e) forming the final RTS)

1.1.1　骨架

中轴（Blum，1973）或骨架 S 是平面 D 的最大内切圆心的轨迹（点集），每个骨架点 $s \in \{S\}$ 存储其最大内切圆半径 $r(s)$，具有最大骨架半径的接合点称为端节点。在图 1（a）中，e_1、e_2、e_3、e_4、e_5、e_6 是端点，O_1、O_2、O_3、O_4 是结合点，因为 O_1 是骨架半径最大的交汇点，所以它是根节点。O_1O_2、O_2O_3、O_3e_4 是骨架分支。

从根节点到骨架树末端节点的最短路径称为骨架末端路径。根据根节点 R 和末端节点 e_i，可以将一个完整的形状分解为 N 个部件，其中 N 是末端节点的数量，每个部件都由骨架末端路径 l_i：$R\,e_i$ 表示，如图 1（c）所示。

1.1.2　构造特征向量

使用骨架图寻求相对鲁棒的表征，由于骨架的根节点相对稳定，因此使用根点和端点（骨架末端

路径)将形状分解为几个子模式,如图1(c)所示,对骨架末端路径进行量化即可提取特征向量。

目标形状由一些直方图描述,这些直方图可以反映该形状的骨架结构和统计信息。直方图横坐标表示从相应根节点到末端节点的路径长度,纵坐标表示与骨架点对应的最大内切圆半径。

本文提出一种非均匀采样的方法对骨架路径进行量化。如图1(d)所示,将根节点 R 到末端节点 e_i 的骨架路径 l_i 输入到直方图中,对于骨架的末端,对骨架路径的前80%均匀采样25个数据点(稀疏采样),对最后20%(末端部分)均匀采样25个数据点(密集采样)。因为越靠近端节点的位置,骨架枝越能刻画物体的形状细节。

骨架路径为 $l_i:\{r_1,r_2,\cdots,r_{50}\}$,其中 r_i 是骨架路径的第 i 个采样点处的骨架半径。记录骨架路径上所有骨架点的骨架半径,骨架枝质量与骨架枝长度的关系为

$$M_i = \sum_{s\in l_i} r(s) \tag{1}$$

式中,M_i 表示骨架枝质量。

为了确保RTS具有尺度不变性,进行归一化操作,具体为

$$\hat{r}(s) = \frac{r(s)}{r^*},\ r^* = \max_{s\in S} r(s)$$

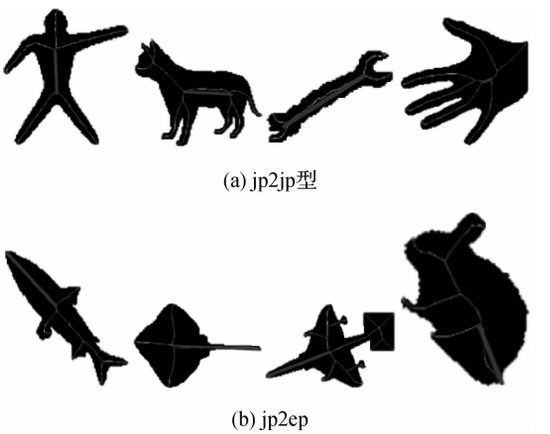
$$\hat{M}_i = \frac{M_i}{M},\ M = \sum_{s\in S} r(s)$$

$$\hat{L}_i = \frac{L_i}{L},\ L = \sum_{\text{branch}_k\in S} L_{\text{branch}_k} \tag{2}$$

式中,L_{branch_k} 表示第 k 个骨架枝的长度,L_i 表示骨架枝的长度。

为了进一步约束末端路径的分布,本文引入末端节点相对于根节点的空间角度。几乎所有物体都有相对稳定的轴,不同于生物意义上的脊椎,本文考察的是刚性的类脊柱轴(SPA),SPA的出现很好地稳定了形状的分布。根据第2个交叉点的类型,定义了两种SPA类型,对骨架图(图1(b))中的所有骨架分支,如果骨架分支的两端均为接合点,则骨架分支为jp2jp型,否则为jp2ep型,如图2所示,图中粗灰线是SPA。

为了约束端节点的空间分布,通过骨架根节点 R 到骨架端节点 e_i 的向量 $\boldsymbol{Vec}_i$ 与 $\boldsymbol{Axis}$ 的余弦相似度的绝对值(SPA的方向不受约束,因此取绝对值)来度量该节点相对于SPA的空间分布。寻找SPA



(a) jp2jp型

(b) jp2ep

图2 SPA类型

Fig. 2 Types of SPA ((a) jp2jp; (b) jp2ep)

的算法具体步骤如下:

1) procedure FIND-SPINEAxis($\boldsymbol{S},\beta,threshold$)

2) $\boldsymbol{S}$ 是骨架剪枝后的形状。

3) $branch_i$: $(\hat{r}_1,\hat{r}_2,\cdots,\hat{r}_N)$, 表示第 i 个骨架枝。N 是骨架枝上的骨架点数量。

4) M_i 和L_i 分别是该骨架枝的质量和长度。

5) $\boldsymbol{EP}$ 和 $\boldsymbol{JP}$ 分别是 $\boldsymbol{S}$ 上的端节点集合和接合点集合。R 表示根节点。

6) for each $branch_i \in \boldsymbol{S}$ do

7) if $branch_i \in$ jp2jp-type then

8) $jp2\,jp_{ml_i} \leftarrow \alpha_1 \times \hat{M}_i + (1-\alpha_1)\times \hat{L}_i$

9) $jp2\,jp_{\text{dense}_i} \leftarrow \dfrac{\hat{M}_i}{\hat{L}_i}$

10) $jp2\,jp_{\text{smooth}_i} \leftarrow \text{std}(branch_i)$

11) else

12) $jp2\,ep_{\text{ml}_i} \leftarrow \alpha_1 \times \hat{M}_i + (1-\alpha_1)\times \hat{L}_i$

13) end if

14) end for

15) $jp2\,ep_{\text{cost}_i} \leftarrow \min(jp2\,jp_{ml_i}) + \text{std}(jp2\,jp_{ml_i})$

16) $jp2\,jp_i \leftarrow \beta \times jp2\,jp_{dense_i} + (1-\beta)\times 10 \times jp2\,jp_{ml\,i}$

$jp2\,ep_i \leftarrow \beta \times jp2\,ep_{cost_i} + (1-\beta)\times 10 \times jp2\,ep_{ml\,i}$

17) $JP2P \leftarrow jp2jp \cup jp2ep$

18) 将 $\boldsymbol{S}$ 视做无向图 $\boldsymbol{G}(\boldsymbol{V},\boldsymbol{E})$, $\boldsymbol{V}$ 是点集合, $\boldsymbol{V} = \boldsymbol{EP} \cup \boldsymbol{JP}$, $\boldsymbol{E}$ 是骨架枝的综合得分。

19) 对于所有的与 R 直接相连的点 P

20）　初始化：$crossingpoint_2 \leftarrow \text{argmax}_{p_i \in P} JP2JP$

21）　while

$crossingpoint_2 \in \boldsymbol{JP}$ and $Jp2j$

$p_{\text{smooth}_{\text{crossingpoint}_2}} > threshold$ do

22）　$JP2P \leftarrow JP2P \backslash crossingpoint_2$

23）　$crossingpoint_2 \leftarrow \underset{p_i \in P}{\text{argmax}} JP2P$

24）　end while

25）　$crossingpoint_1 \leftarrow R$

26）　return $crossingpoint_1 - crossingpoint_2$

27）　end procedure

找到 SPA 和 ***Axis*** 后，则

$$V_i = \left\| \frac{\boldsymbol{Axis} \cdot \boldsymbol{Vec}_i}{\|\boldsymbol{Axis}\| \|\boldsymbol{Vec}_i\|} \right\| \tag{3}$$

式中，V_i 表示骨架根节 R 到骨架端节点 e_i 的向量 $\boldsymbol{Vec}_i$ 与 ***Axis*** 的余弦相似度的绝对值（SPA 的方向不受约束，因此取绝对值）来度量该节点相对于 SPA 的空间分布。

对于一个具有 N 个端节点的特定形状 S，根据其在轮廓上分布的顺时针顺序 $\{\boldsymbol{E}_1, \boldsymbol{E}_2, \cdots, \boldsymbol{E}_N\}$，最终的 RTS 表征为

$$\boldsymbol{E}_i = \{\hat{l}_i : \{\hat{r}_1, \hat{r}_2, \cdots, \hat{r}_{50}\}, \hat{M}_i, \hat{L}_i, V_i\}_{i=1}^{N} \tag{4}$$

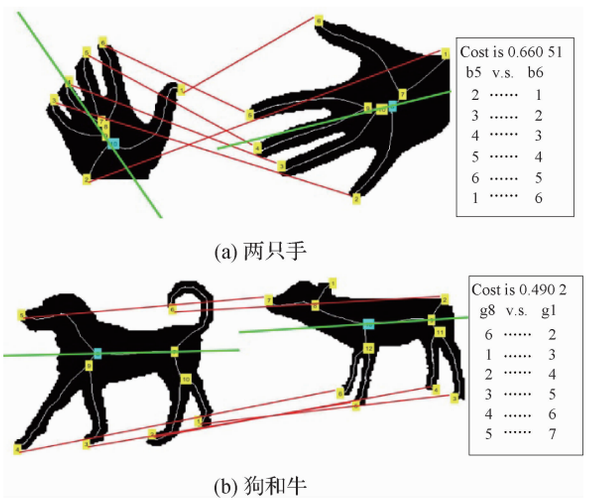
1.2　相似性度量

基于RTS的表征，可以通过骨架量化后的骨架路径 $\hat{I}: \{\hat{r}_1, \hat{r}_2, \cdots, \hat{r}_{50}\}$ 、归一化质量 $\hat{M}$ 和长度 $\hat{L}$ 、相对于 SPA 的空间分布 V 来度量目标形状的每个骨架末端路径。

Eiter 和 Mannila（1994）将量化的骨架路径描述为一条曲线，并使用 Fréchet 距离计算两条骨架路径之间的相似度。为了获得两个形状之间的匹配关系，Latecki 等人（2007）使用 OSB（optimal subsequence bijection）算法处理长度不等的两个序列之间的匹配问题。序列之间可能包含一些异常元素，该算法可以跳过一些不匹配的元素，从而实现了弹性匹配。图 3 是匹配结果的两个示例，两个形状之间的相应端节点用线连接，穿越线表示通过寻找 SPA 的算法获得的 SPA，图 3（a）和图 3（b）右侧分别为形状的匹配代价与端点的对应关系。

1.3　基于多层级泛化框架建立泛化的 RTS（GRTS）

本文提出了一种基于 RTS 的多层级泛化框架，归纳出相似形状的通用表征 GRTS，具体过程如图 4 所示。从建立的 RTS 可以看出，相似的对象具有相似的结构。基于第 1.2 节中的相似性度量算法，将每个形状初始化为一个具体类别，然后将两个最相似的实例泛化为一个新实例，直到数目减少为 1。每一个泛化的实例都可以从上级实例中提取到它们公有的相似的形状信息，用于后续迭代更一般的表征。随着泛化的逐步进行，相似的公共信息将逐步沉淀，最终得到这些类别中最具泛化能力的知识表征。



(a) 两只手

(b) 狗和牛

图 3　匹配结果的两个示例

Fig. 3　Matching results for two examples

((a) two hands; (b) dog and cattle)

GRTS 从一个逐级泛化的框架中建立，每个骨架末端路径由泛化得到它的所有 RTS 的互相匹配的骨架末端路径组成。对具有数量优势的实例，一种自然的做法是根据实例数为 GRTS 中的节点分配权重。若 GRTS 由 λ 个形状泛化得到，而泛化得到其上级的某个节点有 μ 个形状，那么该节点的权重为 $\omega = \mu/\lambda$。假设待泛化的两个节点 $GRTS_x$ 和 $GRTS_y$ 分别是泛化了 ν 和 μ 个实例形状得到的，设 $GRTS_x$ 包含 N 个端节点 $\boldsymbol{E}_x = \{e_{x_1}, \cdots, e_{x_N}\}$，$GRTS_y$ 包含 M 个端节点 $\boldsymbol{E}_y = \{e_{y_1}, \cdots, e_{y_M}\}$。对 $GRTS_x$ 和 $GRTS_y$ 的泛化结果 $GRTS$，考察 $GRTS_x$ 和 $GRTS_y$ 的形状匹配结果，找到两个形状间的骨架端节点对应关系为 $\varphi: \{e_{x_1}, \cdots, e_{x_N}\} \rightarrow \{e_{y_1}, \cdots, e_{y_M}\}$，$(e_{x_{a_1}}, e_{y_{b_1}})$，$(e_{x_{a_2}}, e_{y_{b_2}}), \cdots, (e_{x_{a_m}}, e_{y_{b_m}})$，$m \leqslant \min(N, M)$，即生成的 GRTS 具有 m 个骨架端节点，通过匹配 $(e_{x_{a_i}}, e_{y_{b_i}})$ 获得端节点 e_i，具体为

$$e_{x_{a_i}}:\{l_{x_{a_i}}\rightarrow\{r_{x_1},\cdots,r_{x_{50}}\},M_{x_{a_i}},L_{x_{a_i}},V_{x_{a_i}}\}$$
$$e_{y_{b_i}}:\{l_{y_{b_i}}\rightarrow\{r_{y1},\cdots,r_{y_{50}}\},M_{y_{b_i}},L_{y_{b_i}},V_{y_{b_i}}\}$$
$$e_i:\{l_i\rightarrow\{r_1,\cdots,r_{50}\},M_i,L_i,V_i\} \tag{5}$$

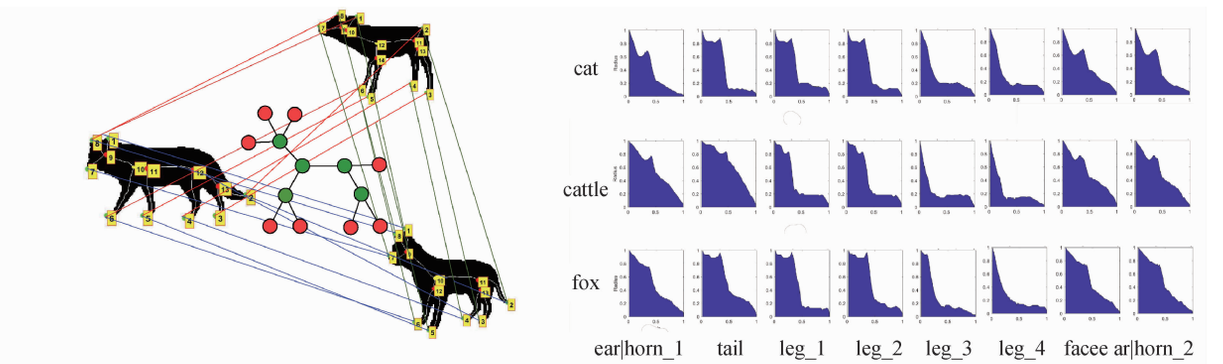
最终,泛化策略为

$$l_i=\frac{\nu}{\nu+\mu}l_{x_{a_i}}+\frac{\mu}{\nu+\mu}l_{y_{b_i}} \tag{6}$$

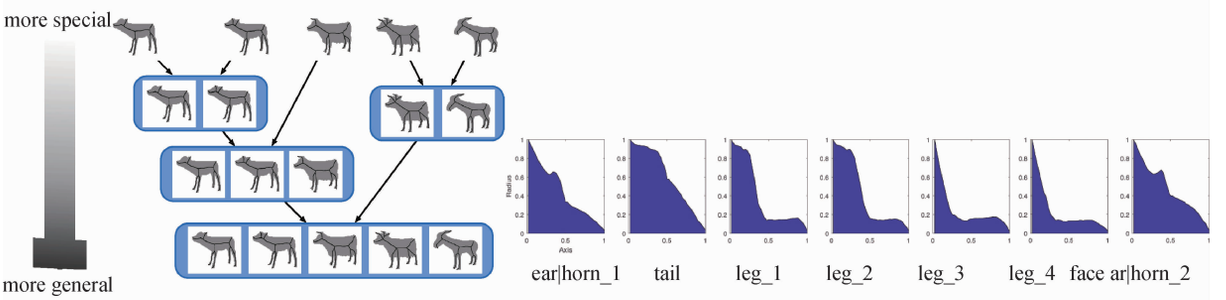
$$M_i=\frac{\nu}{\nu+\mu}M_{x_{a_i}}+\frac{\mu}{\nu+\mu}M_{y_{b_i}} \tag{7}$$

$$L_i=\frac{\nu}{\nu+\mu}L_{x_{a_i}}+\frac{\mu}{\nu+\mu}L_{y_{b_i}} \tag{8}$$

$$V_i=\frac{\nu}{\nu+\mu}V_{x_{a_i}}+\frac{\mu}{\nu+\mu}V_{y_{b_i}} \tag{9}$$



(a) 同类别的形状具有相同的拓扑结构和相似的RTS



(b) 多层级泛化得到GRTS的过程

图 4　多层级泛化框架获得 GRTS 的过程

Fig. 4　Process to obtain GRTS based on multilevel generalization

((a) common structures of shapes in the same category; (b) multilevel generalization process to obtain GRTS)

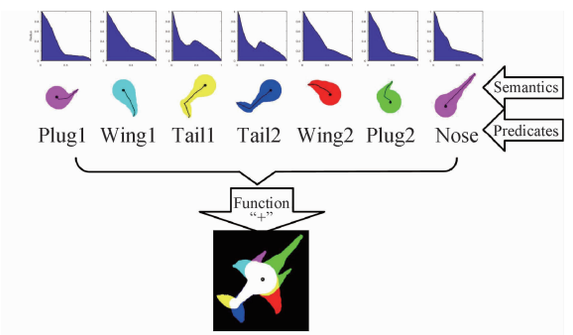
2　GRTS 可以进行不确定性推理

人工智能创始人之一 John MaCarthy 早期在关于推理不需要再编程的设想中,提出仅通过更新显式化的知识就能生成新的推理系统。理论上,这一思想特别适合用于物体识别,但事与愿违,当前的物体识别系统都需要根据具体物体的种类重新写代码,完全不能像专家系统那样做到通用。人们熟知飞机的一个典型特征就是有翅膀,很自然用一条产生式规则陈述

IF has_Wing(x) **THEN** is_Plane(x)

但是如此形式的规则是缺乏可操作性的,存在诸多不确定因素。首先,谓词 has_Wing( ) 的语义是什么? 谓词中的变元 x 如何从真实图像中提取? x 被实例化后的命题的真值怎么判断? 其次,除了 has_Wing( ),判定飞机还涉及其他谓词,如 has_Tail( )。而 Wing 和 Tail 之间是存在空间位置约束的,那么这两个谓词之间的这种关联性如何体现? 最后,飞机是多样性的,例如 Wing 就不止有唯一的形态和尺寸,那么这种不确定性仅用谓词 has_Wing( ) 是否足够? 基于这些原因,传统的符号推理方法很难用于物体识别。从本质上说,这是人工智能中的传统问题——语义落地难题在“如何将谓词语义落实到图像”上的再现。形式化相对宏观的谓词是容易的,而具体化相对微观的操作语义(operational semantic)是困难的。然而,基于得到的 GRTS,计算机能够很容易生成一组关于物体组成结构的判定法

则。图 5 是一组关于判定飞机的法则，完全基于 GRTS 中每个骨架分支派生出来的谓词及其语义定义，并且由 GRTS 到谓词符号的选择和产生式规则的构成都是非常直接的。图中符号“ + ”表示一个函数运算，就是将若干骨架分支装配起来。置信度值体现了物体识别中的不确定性。这个 GRTS 具有可操性，有以下几点原因：1）它的每条骨架分支不但能够反映飞机的一个部件，而且包含了这些部件间的装配关系；2）它的每条骨架分支都能够用一个足够概念化的谓词符号表达，语义界限很明确；3）骨架分支的数值化形态表征将谓词符号与带模板监督的图像操作连接起来，使得谓词符号的意义不再是空洞的；4）推理中的不确定性（Wei 等，2016）可以得到形状相似性比较的支持；5）一旦生成了这样的产生式规则，后续基于 Bayesian 推理和 Fuzzy 推理等其他处理不确定性的手段就可以顺利引入。



GRTS of the “plane” category
Rules can be naturally derived from a GRTS

Rule index	IF （antecedent）	THEN （consequent）	Centainty factor
1	Nose	is_Plane	0.9
2	Wing1 **AND** Wing2	has_Wing	1
3	is_Wing	is_Plane	0.9
4	Plug1 **OR** Plug2	has_Plug	0.9
5	is_Plug	is_Plane	0.7
6	Tail1 **AND** Tail2	has_Tail	0.9
7	is_Tail	is_Plane	0.8

图 5　GRTS 派生出的具有足够可操作性的语义规则

Fig. 5　A set of production rules, with sufficient operability, the rule set generated by GRTS

图 6 是使用 GRTS 生成的规则集进行不确定性推理的过程。其中，图 6（a）是飞机的形状，图 6（b）是由 GRTS 派生的规则集组成的规则树，图 6（c）由 RTS 收集的证据及相应的确定性因子，CF（certainty factor）（is_Plane，ALL_Evidences） = 0.999 347，图 6（d）是紫色掩码，即规则应用的结果。

基于图 6（b）的规则树，计算图 6（a）所有证据的可信度，每个证据都来自 RTS 的骨架分支。具体为

$$
\begin{aligned}
&CF(is_{Plane},E_1') = CFB(Plug_1 \text{ or } Plug_2,E_1') = \\
&\max(CF(Plug_1,E_1'),CF(Plug_2,E_1')) \times 0.7 = 0.28 \\
&CF(is_{Plane},E_6') = CF(Plug_1 \text{ or } Plug_2,E_6') = \\
&\max(CF(Plug_1,E_6'),CF(Plug_6,E_6')) \times 0.7 = 0.28 \\
&CF(is_{Plane},E_7') = CF(Wing_1 \text{ and } Wing_2,E'_7) = \\
&\min(CF(Wing_1,E_7'),CF(Wing_2,E_7')) \times 0.9 = 0.81 \\
&CF(is_{Plane},E_{11}') = CF(Wing_1 \text{ and } Wing_2,E_{11}') = \\
&\min(CF(Wing_1,E_{11}'),CF(Wing_2,E_{11}')) \times 0.9 = 0.81 \\
&CF(is_{Plane},E_8') = CF(Plug_1 \text{ or } Plug_2,E_8') = \\
&\max(CF(Plug_1,E_8'),CF(Plug_2,E_8')) \times 0.7 = 0.49 \\
&CF(is_{Plane},E_{10}') = CF(Plug_1 \text{ or } Plug_2,E_{10}') = \\
&\max(CF(Plug_1,E_{10}'),CF(Plug_2,E_{10}')) \times 0.7 = 0.49 \\
&CF(is\_Plane,E_9') = CF(Nose,E_9') \times 0.9 = 0.81 \\
&CF(is_{Plane},E_2') = CF(Tail_1 \text{ and } Tail_2,E_2') = \\
&\min(CF(Tail_1,E_2'),CF(Tail_2,E_2')) \times 0.8 = 0.4 \\
&CF(is_{Plane},E_5') = CF(Tail_1 \text{ and } Tail_2,E_5') = \\
&\min(CF(Tail_1,E_5'),CF(Tail_2,E_5')) \times 0.8 = 0.32 \\
&CF(is_{Plane},E_3') = CF(Tail_1 \text{ and } Tail_2,E_3') = \\
&\min(CF(Tail_1,E_3'),CF(Tail_2,E_3')) \times 0.8 = -0.24 \\
&CF(is_{Plane},E_4') = CF(Tail_1 \text{ and } Tail_2,E_4') = \\
&\min(CF(Tail_1,E_4'),CF(Tail_2,E_4')) \times 0.8 = -0.24
\end{aligned}
\tag{10}
$$

假设知识库由以下规则组成：

Rule_1：IF e_1 THEN H

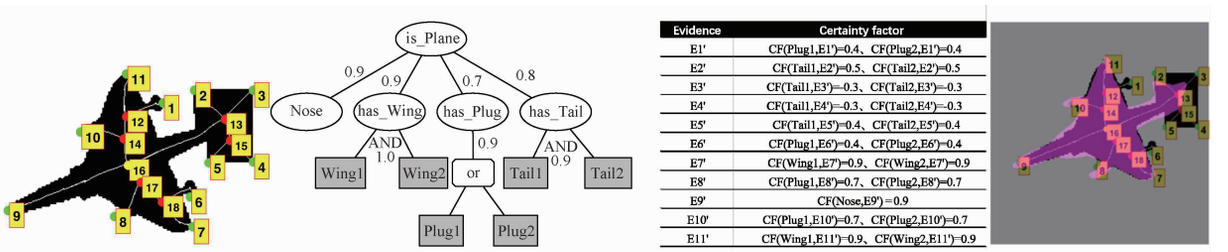
Rule_2：IF e_2 THEN H

则组合的确定性因子的计算式为

$$
CF(H,e_1 \& e_2) = \begin{cases}
CF(H,e_1) + CF(H,e_2) - CF(H,e_1) \times CF(H,e_2) \\
\qquad CF(H,e_1) > 0 \text{ 且 } CF(H,e_2) > 0 \\
CF(H,e_1) + CF(H,e_2) + CF(H,e_1) \times CF(H,e_2) \\
\qquad CF(H,e_1) < 0 \text{ 且 } CF(H,e_2) < 0 \\
\dfrac{CF(H,e_1) + CF(H,e_2)}{1 - \min\{\|CF(H,e_1)\|,\|CF(H,e_2)\|\}} \\
\qquad CF(H,e_1) \times CF(H,e_2) < 0
\end{cases}
\tag{11}
$$

式中，$\|CF(H,e_1)\|$ 和 $\|CF(H,e_2)\|$ 分别是 $CF(H,e_1)$ 和 $CF(H,e_2)$ 的绝对值。根据式（10）和式（11），可得

$$
CF(is\_plane,E_1' \& E_2' \& \cdots \& E_{11}') = 0.999\ 347 \tag{12}
$$



Evidence	Certainty factor
E1'	CF(Plug1,E1')=0.4、CF(Plug2,E1')=0.4
E2'	CF(Tail1,E2')=0.5、CF(Tail2,E2')=0.5
E3'	CF(Tail1,E3')=-0.3、CF(Tail2,E3')=-0.3
E4'	CF(Tail1,E4')=-0.3、CF(Tail2,E4')=-0.3
E5'	CF(Tail1,E5')=0.4、CF(Tail2,E5')=0.4
E6'	CF(Plug1,E6')=0.4、CF(Plug2,E6')=0.4
E7'	CF(Wing1,E7')=0.9、CF(Wing2,E7')=0.9
E8'	CF(Plug1,E8')=0.7、CF(Plug2,E8')=0.7
E9'	CF(Nose,E9')=0.9
E10'	CF(Plug1,E10')=0.7、CF(Plug2,E10')=0.7
E11'	CF(Wing1,E11')=0.9、CF(Wing2,E11')=0.9

(a) 飞机形状　(b) 规则树　(c) 确定性因子　(d) 紫色掩码

图 6　使用 GRTS 生成的规则集进行不确定性推理的过程

Fig. 6　The process of uncertainty reasoning using the rule set generated by GRTS

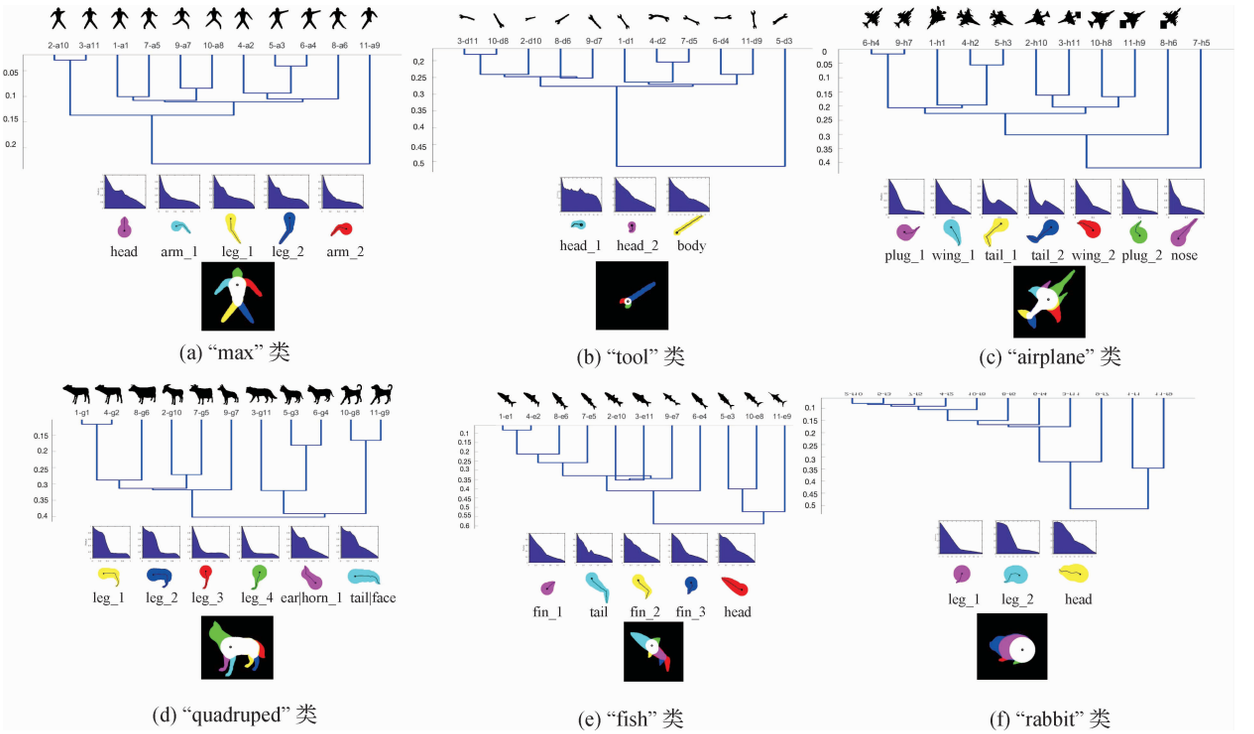
((a) shape of plane; (b) rule-tree; (c) certainty factors; (d) purple mask)

3　实　验

3.1　GRTS 的生成实验

由于本文方法是在二值形状上进行操作，对视角较为敏感(3 维不受限)，因此实验选取的数据集一般是面向形状表征的。本文选取最经典的 Kimia99(Sebastian 等，2004)以及 Tari56 形状数据集(Asian 和 Tari，2005)进行实验。

图 7 和图 8 分别是 GRTS 在 Kimia99 和 Tari56 数据集上的生成过程和结果，从形式上展示了同类属物体最一般表征的物理意义，可以看出同类属物体的泛化过程以及最终得到的 GRTS。为了直观看出 GRTS 中直方图的具体含义，使用同类的某个物体的骨架路径进行重建，可以看出 GRTS 各个部位的物理意义。



(a) “max” 类　(b) “tool” 类　(c) “airplane” 类

(d) “quadruped” 类　(e) “fish” 类　(f) “rabbit” 类

图 7　GRTS 在 Kimia99 数据集上的生成结果

Fig. 7　Results by GRTS on Kimia99 database ((a) “man” class; (b) “tool” class; (c) “airplane” class; (d) “quadruped” class; (e) “fish” class; (f) “rabbit” class)

从图 7 可以看出，GRTS 可以表征同类属物体的一般结构特征。如在 Kimia99 数据集上的 GRTS 实验中，图 7(c)展示的物体是形式各异的战斗机，挂件数量各异，尾翼形状不一，然而最终的 GRTS 结果忽略了各个实例的个异化结构，能够综合各部件的小差异，如有的尾翼较大，有的较小。真正做到了

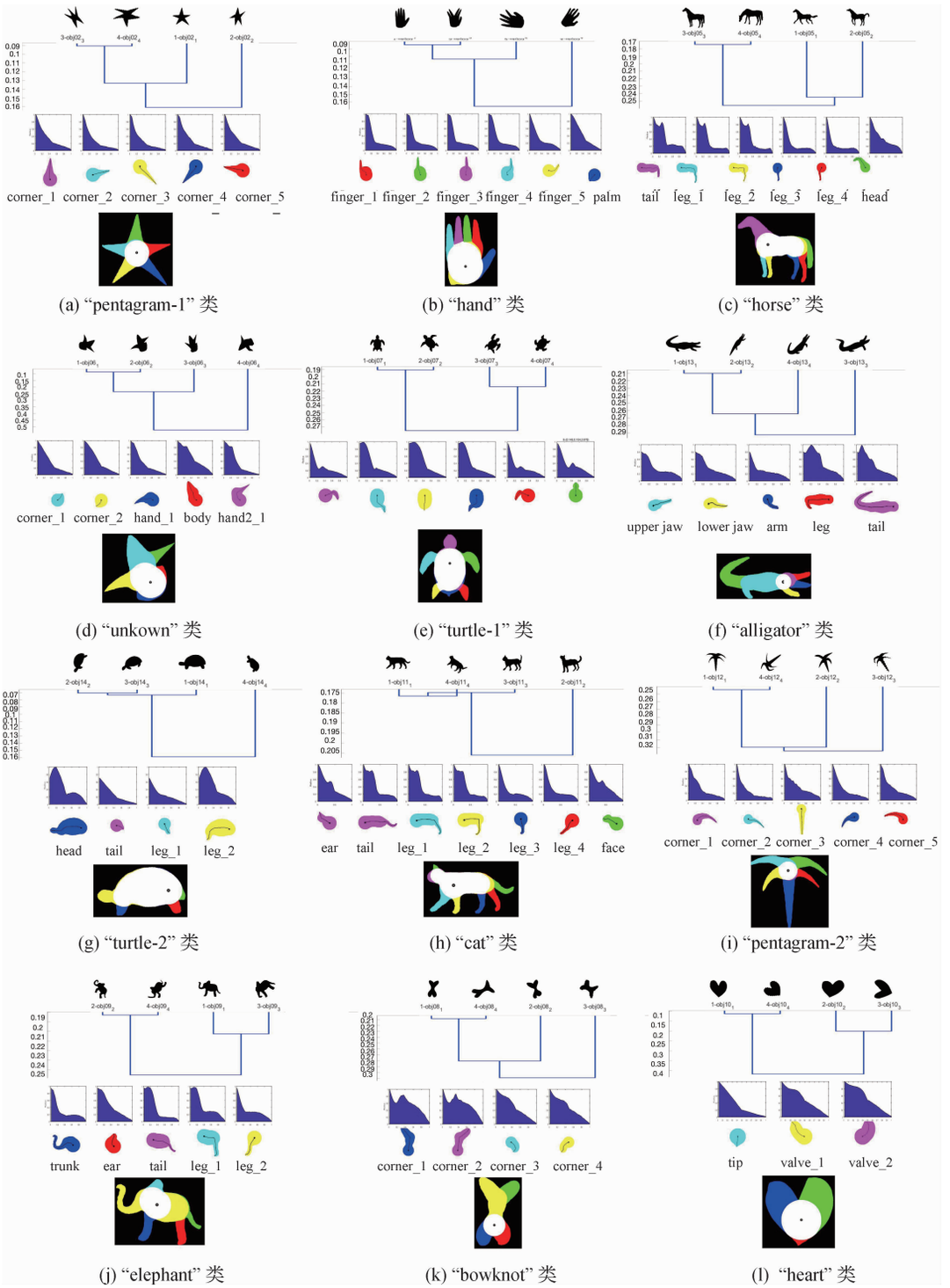


图 8　在 Tari56 数据集上 GRTS 的生成结果

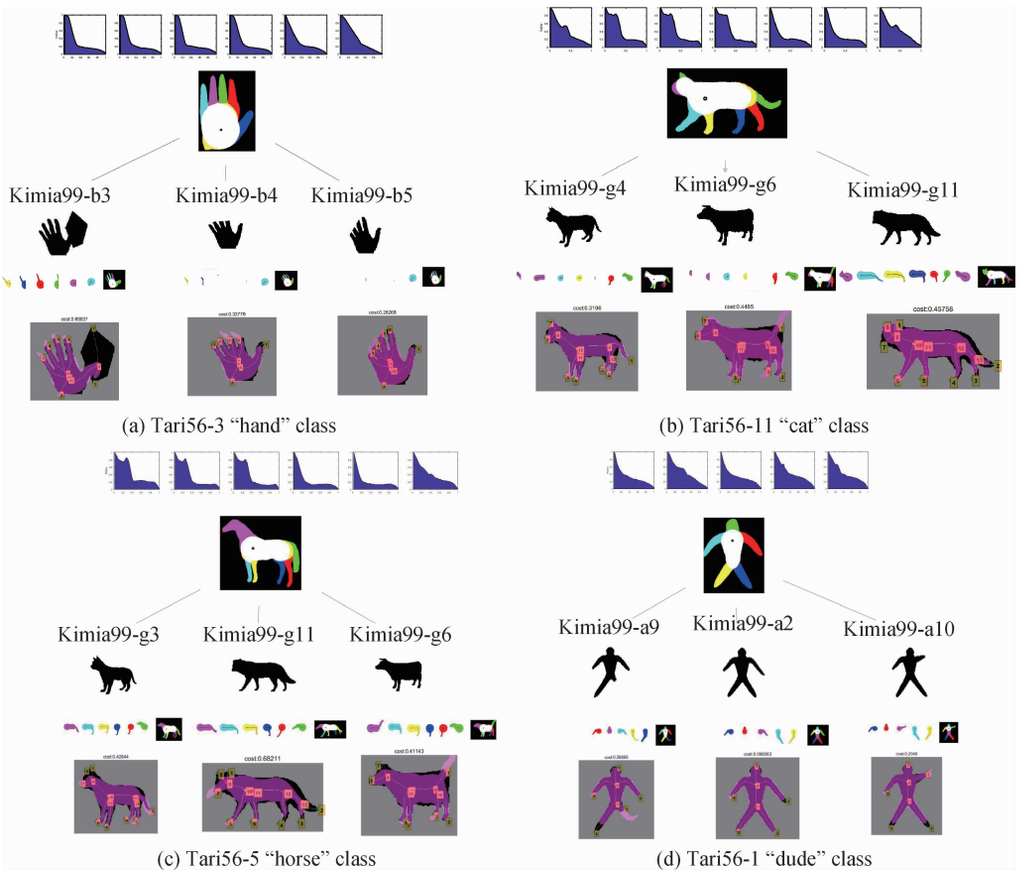
Fig. 8　Results by GRTS on Tari56 database ((a) “pentagram-1” class; (b) “hand” class; (c) “horse” class; (d) “unkown” class; (e) “turtle-1” class; (f) “alligator” class; (g) “turtle-2” class; (h) “cat” class; (i) “pentagram-2” class; (j) “elephant” class; (k) “bowknot” class; (l) “heart” class)

“求大同，存小异”，表征了战斗机的最一般结构，包括两个机翼、两个尾翼、机头以及两个外挂部件。图 7(d)展示的物体是牛、羊、猫、狐狸、狗等具有较大类内差异的动物，得到的 GRTS 仍归纳出该类物体的最一般本质特征，如都有四肢结构、至少一个角状物(或耳朵)和尾巴等。以上实验说明，GRTS 可以通过基于 RTS 的多层级泛化框架逐步归纳，得到最本质的构型特征。

3.2 可解释性验证

图9是使用从Tari56数据集获得的几类GRTS，将拓扑特征应用于Kimia99数据集中最接近类别的样本，以发现新测试样本相对于GRTS的特定差异，从而验证该表征的可解释性。

图9(a)显示了Kimia99中Tari56的“手”类别中几个示例的拓扑特征应用结果。Kimia99-b3是一只手具有巨大的遮挡，在应用“手”类别GRTS之后，可以清楚地看到两只手之间的区别：原图比紫色掩码(紫色掩码是应用GRTS生成的图，黑色底图是原图)多出了大拇指处的遮挡物。从Kimia99-b4和Kimia99-b5的应用结果可以看到，紫色掩码在原图缺失的部分手指上方生成了完整的手指。此外，图的顶部定量展示了GRTS识别的相似度。这表明GRTS不仅可以识别手指，而且具有很强的可解释性，可以直观展示被遮挡和丢失的细节。



(a) Tari56-3 “hand” class　(b) Tari56-11 “cat” class

(c) Tari56-5 “horse” class　(d) Tari56-1 “dude” class

图9　使用Tari56的GRTS在Kimia99部分数据上进行拓扑特征应用的实验结果

Fig. 9　Topological character application experiments on the Kimia99 database using Tari56 GRTS

((a) Tari56-3 “hand” class; (b) Tari56-11 “cat” class; (c) Tari56-5 “horse” class; (d) Tari56-1 “dude” class)

3.3 模型的可解释性实验

子结构的形状相似性视为属于GRTS的对象的置信度。基于第1.2节中的相似性度量算法，可以找到丢失的形状与具有最高置信度的对象之间的子结构匹配关系。应用相似的变换(仅需要两对匹配点)可以获得相似的变换矩阵以重建缺失的形状。

图10是几个形状补全示例的实验结果及其置信度。首先从不完整的形状中提取RTS以收集证据(Wei等，2016)，然后使用从Tari56数据集获得GRTS建立的规则集进行不确定性推理。从图10可以看出，收集的证据对某些类别具有很高的置信度，可以完成其余形状的重建。表1是基于图10的识别结果，通过表1不仅可以验证形状补全的结果，而且可以进一步补充不可见部分。

以图10(a)为例，该形状仅有头部和前肢，但是与马的局部匹配具有非常高的置信度，从而可以顺利地完成其余部分的重建。该模型不仅得到了这种不完整的形状属于马的结论，而且更重要的是清楚地得到了模型所基于的观察条件(马头，前肢)。如图10所示，从形状和马各部分的置信度可以探索内

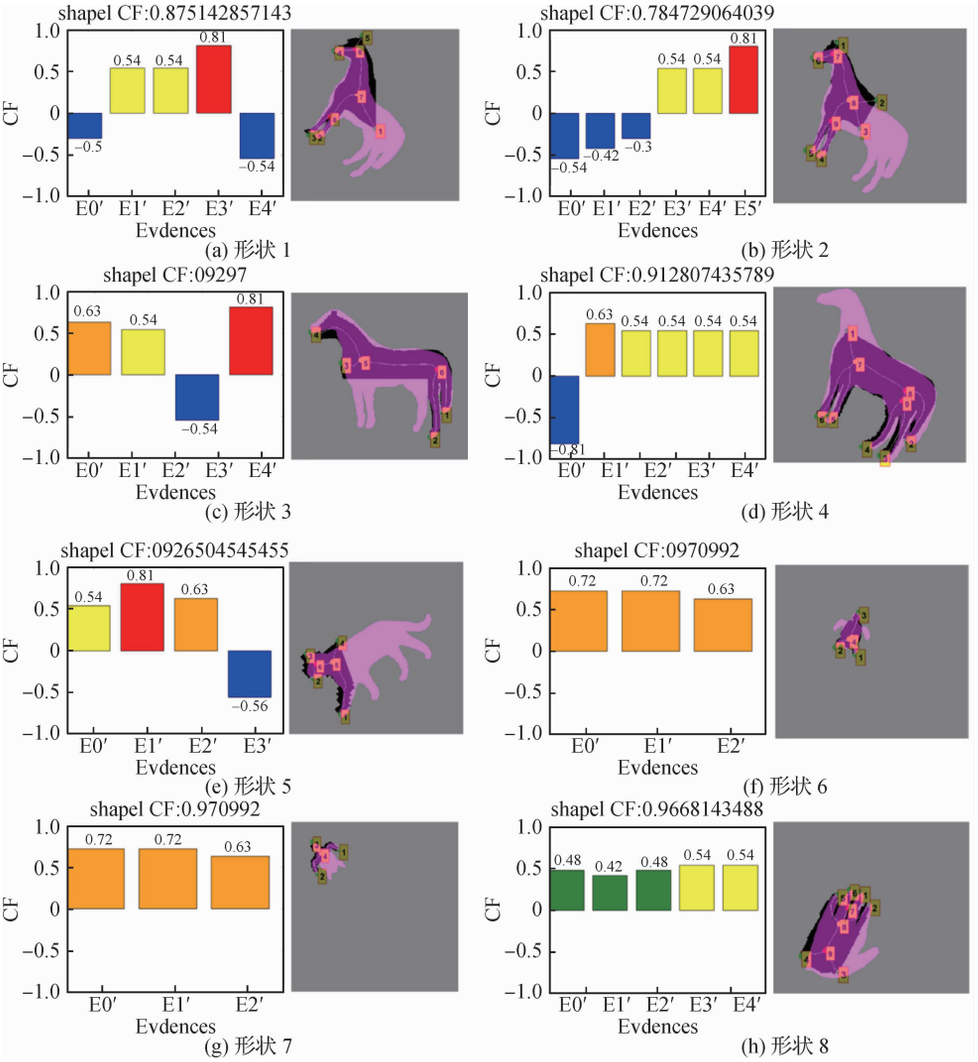


图 10　形状补全的实验结果及其置信度

Fig. 10　Experimental results and confidence in each evidence

((a) shape1; (b) shape2; (c) shape3; (d) shape4; (e) shape5; (f) shape6; (g) shape7; (h) shape8)

表 1　使用 Tari56 数据集得到的 GRTS 对 8 个形状的识别置信度

Table 1　The confidence of each shape by GRTSs of Tari56 dataset

类别	形状 1	形状 2	形状 3	形状 4	形状 5	形状 6	形状 7	形状 8
dude	-0.75	-0.82	-0.72	-0.56	-0.86	0.47	0.51	0.25
hand	-0.82	-0.75	-0.89	0.37	-0.75	-0.43	-0.43	**0.97**
horse	**0.88**	**0.79**	**0.91**	**0.93**	-0.04	-0.40	-0.40	-0.40
turtle1	-0.34	-0.56	-0.42	-0.50	-0.90	**0.97**	**0.97**	-0.40
turtle2	-0.34	-0.43	-0.36	-0.50	-0.14	0.76	0.76	-0.74
alligator	-0.87	-0.91	-0.93	-0.76	-0.89	-0.76	-0.76	-0.89
cat	-0.11	0.01	0.66	0.89	**0.93**	-0.40	-0.40	-0.48
pentagram1	0.45	0.19	-0.22	0.19	-0.29	0.74	0.74	0.75
pentagram2	0.45	0.19	-0.22	0.19	-0.29	0.74	0.74	0.75
lephant	-0.56	-0.72	-0.75	0.86	0.43	-0.11	-0.11	-0.72
bowknot	-0.89	-0.91	-0.90	-0.78	-0.89	-0.77	-0.77	-0.89
heart	-0.80	-0.87	-0.91	-0.93	-0.76	-0.76	-0.76	-0.78

注：加粗字体表示各列最优结果。

部结构的差异,而不是简单地得到它们是否相似的结论。这种直观的解释有助于人们更好地理解模型预测背后的依据,从而增加模型的可解释性。

4 结 论

本文方法以骨架树为基础,形成了基于泛化学习的物体拓扑特征的形式化表征方式,并进行知识抽取。针对人工智能基础的低阶结构不连续假设,给出了一种将不同阶表征连贯起来的可能途径。基于泛化框架,仅利用少量数据(Kimia99 数据集需要 11 个,Tari56 数据集需要 4 个)就可以形成有关物体类别的知识泛化表征 GRTS。这种泛化表征的结果可以用于进行不确定性推理实验。该方法避免了基于纹理特征带来的不确定性,适用于任意基于基元的表征方式,具有更好的鲁棒性与普适性。

在 Kimia99 和 Tari56 形状数据集上执行不确定性推理,不仅给出了识别的置信度,而且还给出了模型识别背后依据的证据。这项工作给出了一种提供解释性的新思路,有助于在模型与人之间建立信任关系,促进可解释性工作的进一步开展。

在未来工作中,将融合基于统计机器学习的方法和基于本文符号计算的方法,使之既具备很强的学习能力,又在很大程度上可理解、可支配以及可调整。

参考文献(References)

Adankon M M and Cheriet M. 2014. Support Vector Machine. New York: Springer: 1-28 [DOI: 10.1007/978-3-642-27733-7_299-3]

Andreopoulos A and Tsotsos J K. 2013. 50 years of object recognition: directions forward. Computer Vision and Image Understanding, 117(8): 827-891 [DOI: 10.1016/j.cviu.2013.04.005]

Asian C and Tari S. 2005. An axis-based representation for recognition//Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05). Beijing, China: IEEE: 1339-1346 [DOI: 10.1109/ICCV.2005.32]

Aubry M and Russell B C. 2015. Understanding deep features with computer-generated imagery//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 2875-2883 [DOI: 10.1109/ICCV.2015.329]

Bai X, Wang X G, Latecki L J, Liu W Y and Tu Z W. 2009. Active skeleton for non-rigid object detection//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE: 575-582 [DOI: 10.1109/ICCV.2009.5459188]

Bau D, Zhou B L, Khosla A, Oliva A and Torralba A. 2017. Network dissection: quantifying interpretability of deep visual representations//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE: 6541-6549 [DOI: 10.1109/CVPR.2017.354]

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. Psychological Review, 94(2): 115-147 [DOI: 10.1037/0033-295X.94.2.115]

Blum H. 1973. Biological shape and visual science (part I). Journal of Theoretical Biology, 38(2): 205-287 [DOI: 10.1016/0022-5193(73)90175-6]

Brendel W and Bethge M. 2019. Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet. https://arxiv.org/pdf/1904.00760.pdf

Catherine R and Cohen W. 2016. Personalized recommendations using knowledge graphs: a probabilistic logic programming approach//Proceedings of the 10th ACM Conference on Recommender Systems. Boston, USA: ACM: 325-332 [DOI: 10.1145/2959100.2959131]

Chen P Y, Sharma Y, Zhang H, Yi J F and Hsieh C J. 2018. EAD: elastic-net attacks to deep neural networks via adversarial examples//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA: AAAI

Cho S, Maqbool M H, Liu F and Foroosh H. 2019. Self-attention network for skeleton-based human action recognition [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1912.08435.pdf

Crevier D and Lepage R. 1997. Knowledge-based image understanding systems: a survey. Computer Vision and Image Understanding, 67(2): 161-185 [DOI: 10.1006/cviu.1996.0520]

Demirci M F, Shokoufandeh A and Dickinson S J. 2009. Skeletal shape abstraction from examples. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(5): 944-952 [DOI: 10.1109/TPAMI.2008.267]

Dosovitskiy A and Brox T. 2016. Inverting visual representations with convolutional networks//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 4829-4837 [DOI: 10.1109/CVPR.2016.522]

Eiter T and Mannila H. 1994. Computing Discrete Fréchet Distance. Technical Report CD-TR 94/64, Technische Universitat Wien

Eysenck M W and Keane M T. 2005. Cognitive Psychology: A Student's Handbook. 5th ed. [s.l.]: Psychology Press

Fu K S. 1977. Syntactic Pattern Recognition Applications. New York: Springer-Verlag

Gao H, Zhuang L, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE: 4700-4708 [DOI: 10.1109/CVPR.2017.243]

Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann F A and Brendel W. 2018. ImageNet-trained CNNs are biased towards tex-

ture; increasing shape bias improves accuracy and robustness [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1811.12231.pdf

Gonzalez R C and Thomas M G. 1978. Syntactic pattern recognition: an introduction. Reading: Addison-Wesley

Goodfellow I J, Shlens J and Szegedy C. 2014. Explaining and harnessing adversarial examples [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1412.6572.pdf

Graham S A and Diesendruck G. 2010. Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. Cognitive Development, 25(2): 111-123 [DOI: 10.1016/j.cogdev.2009.06.002]

Hayes J and Danezis G. 2017. Machine learning as an adversarial service: learning black-box adversarial example[EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1708.05207.pdf

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hofmann M, Bourne J and Brodersen A. 1989. Knowledge representation for reasoning about devices//Proceedings of the 21st Southeastern Symposium on System Theory. Tallahassee: IEEE: 446-449 [DOI: 10.1109/SSST.1989.72508]

Hubel D H and Wiesel T N. 1959. Receptive fields of single neurones in the cat's striate cortex. The Journal of Physiology, 148(3): 574-591 [DOI: 10.1113/jphysiol.1959.sp006308]

Hupp J M. 2008. Demonstration of the shape bias without label extension. Infant Behavior and Development, 31(3): 511-517 [DOI: 10.1016/j.infbeh.2008.04.002]

Iwańska L, Mata N and Kruger K. 1999. Fully automatic acquisition of taxonomic knowledge from large corpora of texts: limited syntax knowledge representation system based on natural language//Proceedings of the 11th International Symposium on Methodologies for Intelligent Systems. Poland: Springer: 430-438 [DOI: 10.1007/BFb0095130]

Kindermans P J, Schütt K T, Alber M, Müller K R, Erhan D, Kim B and Dähne S. 2017. Learning how to explain neural networks: patternnet and patternattribution [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1705.05598.pdf

Kriegeskorte N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. Annual Review of Vision Science, 1: 417-446 [DOI: 10.1146/annurev-vision-082114-035447]

Kubilius J, Bracci S and de Beeck H P O. 2016. Deep neural networks as a computational model for human shape sensitivity. PLoS Computational Biology, 12(4): e1004896 [DOI: 10.1371/journal.pcbi.1004896]

Landau B, Smith L B and Jones S S. 1988. The importance of shape in early lexical learning. Cognitive Development, 3(3): 299-321 [DOI: 10.1016/0885-2014(88)90014-7]

Latecki L J, Wang Q, Koknar-Tezel S and Megalooikonomou V. 2007. Optimal subsequence bijection//Proceedings of the 7th IEEE International Conference on Data Mining. Omaha, NE, USA: IEEE: 565-570 [DOI: 10.1109/ICDM.2007.47]

LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. Nature, 521(7553): 436-444 [DOI: 10.1038/nature14539]

Li H B and Yu J P. 2014. Knowledge representation and discovery for the interaction between syntax and semantics: a case study of must//Proceedings of 2014 International Conference on Progress in Informatics and Computing. Shanghai: IEEE: 153-157 [DOI: 10.1109/PIC.2014.6972315]

Li H J, Liu L J, Sun F M, Yu B and Liu C X. 2015. Multi-level feature representations for video semantic concept detection. Neurocomputing, 172: 64-70 [DOI: 10.1016/j.neucom.2014.09.096]

Li X and Li F X. 2017. Adversarial examples detection in deep networks with convolutional filter statistics//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 5764-5772 [DOI: 10.1109/iccv.2017.615]

Li X, Zha Y F, Zhang T Z, Cui Z, Zuo W M, Hou Z Q, Lu H C and Wang H Z. 2019. Survey of visual object tracking algorithms based on deep learning. Journal of Image and Graphics, 24(12): 2057-2080 (李玺, 查宇飞, 张天柱, 崔振, 左旺孟, 侯志强, 卢湖川, 王菡子. 2019. 深度学习的目标跟踪算法综述. 中国图象图形学报, 24(12): 2057-2080) [DOI: 10.11834/jig.190372]

Liu B. 2010. Uncertainty Theory: A Branch of Mathematics for Modeling Human Uncertainty. LZsT of FZGURES. 85(13.4)

Mahendran A and Vedaldi A. 2015. Understanding deep image representations by inverting them//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE: 5188-5196 [DOI: 10.1109/CVPR.2015.7299155]

McMahon T and Oommen B J. 2017. Enhancing English-Japanese translation using syntactic pattern recognition methods//Proceedings of the 10th International Conference on Computer Recognition Systems. Cham: Springer: 33-42 [DOI: 10.1007/978-3-319-59162-9_4]

Moosavi-Dezfooli S M, Fawzi A and Frossard P. 2016. DeepFool: a simple and accurate method to fool deep neural networks//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 2574-2582 [DOI: 10.1109/CVPR.2016.282]

Nguyen A, Yosinski J and Clune J. 2015. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE: 427-436 [DOI: 10.1109/CVPR.2015.7298640]

Papadopoulos K, Ghorbel E, Aouada D and Ottersten B. 2019. Vertex feature encoding and hierarchical temporal modeling in a spatial-temporal graph convolutional network for action recognition[EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1912.09745.pdf

Piaget J, Inhelder B, Langdon F J and Lunzer J L. 1957. The child's conception of space. British Journal of Educational Studies, 5(2): 187-189

Rastogi A, Arora R and Sharma S. 2015. Leaf disease detection and grading using computer vision technology and fuzzy logic//Proceedings of the 2nd International Conference on Signal Processing and Integrated Networks (SPIN). Noida, India: IEEE: #7095350 [DOI: 10.1109/SPIN.2015.7095350]

Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Ritter S, Barrett D G T, Santoro A and Botvinick M M. 2017. Cognitive psychology for deep neural networks: a shape bias case study//Proceedings of the 34th International Conference on Machine Learning. Sydney, NSW, Australia: ACM: 2940-2949

Sanocki T, Bowyer K W, Heath M D and Sarkar S. 1998. Are edges sufficient for object recognition? Journal of Experimental Psychology Human Perception and Performance, 24(1): 340-349 [DOI: 10.1037/0096-1523.24.1.340]

Sebastian T B, Klein P N and Kimia B B. 2004. Recognition of shapes by editing their shock graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(5): 550-571 [DOI: 10.1109/TPAMI.2004.1273924]

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//Proceedings of 2017 IEEE International Conference on Computer Vision: Venice, Italy: IEEE: 618-626 [DOI: 10.1007/s11263-019-01228-7]

Sera M D and Millett K G. 2011. Developmental differences in shape processing. Cognitive Development, 26(1): 40-56 [DOI: 10.1016/j.cogdev.2010.07.003]

Shen W, Wang Y, Bai X, Wang H Y and Latecki L J. 2013. Shape clustering: common structure discovery. Pattern Recognition, 46(2): 539-550 [DOI: 10.1016/j.patcog.2012.07.023]

Shen W, Zhao K, Jiang Y, Wang Y, Zhang Z J and Bai X. 2016. Object skeleton extraction in natural images by fusing scale-associated deep side outputs//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 222-230 [DOI: 10.1109/CVPR.2016.31]

Simonyan K, Vedaldi A and Zisserman A. 2013. Deep inside convolutional networks: visualising image classification models and saliency maps [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1312.6034.pdf

Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I and Fergus R. 2013. Intriguing properties of neural networks [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1312.6199.pdf

Torsello A and Hancock E R. 2006. Learning shape-classes using a mixture of tree-unions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(6): 954-967 [DOI: 10.1109/TPAMI.2006.125]

Wei H, Yu Q and Yang C Z. 2016. Shape-based object recognition via evidence accumulation inference. Pattern Recognition Letters, 77: 42-49 [DOI: 10.1016/j.patrec.2016.03.022]

Wu T F, Sun W, Li X L, Song X and Li B. 2017. Towards interpretable R-CNN by unfolding latent structures [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1711.05226.pdf

Yosinski J, Clune J, Bengio Y and Lipson H. 2014. How transferable are features in deep neural networks? //Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Quebec, Canada: ACM: 3320-3328

Yuille A L and Liu C X. 2018. Deep Nets: what have they ever done for Vision? [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1805.04025.pdf

Zeiler M D and Fergus R. 2014. Visualizing and understanding convolutional networks//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 818-833 [DOI: 10.1007/978-3-319-10590-1_53]

Zhang Q S, Cao R M, Zhang S M, Redmonds M, Wu Y N and Zhu S C. 2017. Interactively transferring CNN patterns for part localization. https://arxiv.org/pdf/1708.01783.pdf

Zhang Q S and Zhu S C. 2018. Visual interpretability for deep learning: a survey. Frontiers of Information Technology and Electronic Engineering, 19(1): 27-39 [DOI: 10.1631/FITEE.1700808]

Zhong S P. 2007. Transformation group and geometry. Science and Technology Information, (34): 507, 540 (钟师鹏. 2007. 变换群与几何学. 科技信息(科学教研), (34): 507, 540)

Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE: 2921-2929 [DOI: 10.1109/CVPR.2016.319]

Zhou M, Shi Z W and Ding H P. 2017. Aircraft classification in remote-sensing images using convolutional neural networks. Journal of Image and Graphics, 22(5): 702-708 (周敏, 史振威, 丁火平. 2017. 遥感图像飞机目标分类的卷积神经网络方法. 中国图象图形学报, 22(5): 702-708) [DOI: 10.11834/jig.160595]

Zintgraf L M, Cohen T S, Adel T and Welling M. 2017. Visualizing deep neural network decisions: prediction difference analysis [EB/OL]. [2019-12-19]. https://arxiv.org/pdf/1702.04595.pdf

作者简介

危辉,1971 年生,男,教授,博士生导师,主要研究方向为人工智能与认知科学。
E-mail: weihui@fudan.edu.cn

余莉萍,女,硕士研究生,主要研究方向为人工智能与认知科学。E-mail: 1070776909@qq.com