

知识在视觉中不仅仅是真信念

郑文博¹, 王飞跃²

(1. 武汉理工大学计算机与人工智能学院, 湖北 武汉 430070;
2. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190)

摘要: 计算知识视觉是一个新的观点或研究领域。计算知识视觉在视觉感知模型注入先验知识, 使模型获得理解能力。模型在一定程度上像普通成年人一样进行有效的学习和推理。计算知识视觉不仅为人类先验知识与感知视觉机器提供了“桥梁”, 还能系统地适配“桥梁”上不同层次的视觉任务。另外, 提出了视觉层次及其与知识的联系, 指出知识是一种合理的真信念。最后总结了近期的工作, 并提出了一个新的研究方向, 即知识在视觉方法研究中可以是一个思维框架。

关键词: 视觉; 知识; 计算知识视觉

中图分类号: TP301.4

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202418

Knowledge is far more than true belief in vision

ZHENG Wenbo¹, WANG Fei-Yue²

1. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China
2. State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Computational knowledge vision is emphasized as a novel perspective or field in this paper. Computational knowledge vision injects a priori knowledge into a visual perception model so that the model gains comprehension, which enables the model to learn and reason effectively to a certain extent like an ordinary adult. Computational knowledge vision not only provides a “bridge” between human priori knowledge and perceptual visual machines, but also systematically adapts to different levels of visual tasks on this “bridge”. In addition, this paper proposes the visual hierarchy and its connection to knowledge, stating that knowledge is a justified true belief. Finally, our recent work is summarized and a new direction is proposed, which suggests that knowledge is also a thought framework in vision.

Key words: vision, knowledge, computational knowledge vision

0 引言

“看见”是一种来自大自然的古老能力, 最早可以追溯到5.3亿年前^[1-2]。生命通过视觉感知广阔世界的存在, 并主动用视觉理解世界^[3]。从此, 生物神经系统开始进化; 生物的视觉转化为感知, 进化为对世界的理解和推理, 最终产生智能^[4]。

视觉是人工智能 (artificial intelligence, AI) 的重要组成部分, 其目标是让机器像人类一样, 通过视觉对周围的世界进行学习、理解和推理^[5]。直到1982年大卫·马尔的《视觉》^[6]一书问世, 计算机视觉才成为一门独立的学科。马尔提出, 人工智能研究需要独立考虑包括可计算理论、算法及其实现的3层视觉理解以及他提出的独特的视觉计算理

收稿日期: 2024-04-28; 修回日期: 2024-05-31

通信作者: 郑文博, zwb2022@whut.edu.cn; 王飞跃, feiyue.wang@ia.ac.cn

基金项目: 国家自然科学基金项目 (No.62303361)

Foundation Item: The National Natural Science Foundation of China(No.62303361)

论,其中,视觉理解层次可能是最有影响力和最基础的。近十年来计算机视觉的巨大进步,尤其是元宇宙^[7]和大模型(如GPT-4V^[8]、GPT-4o)的爆发和繁荣,本质上也与理解层次有关。

然而,马尔在《视觉》中并没有研究机器学习的作用,也没有探讨数据或经验的兴起。过去十年,数据对于计算机视觉的进步至关重要。在计算机视觉中利用(人类)知识或经验的研究层出不穷^[9-11],尤其是大模型中的思维链^[12]。而研究者利用深度学习算法,可以将数据转化为知识^[7]。那么,知识对视觉意味着什么?

1 知识是一种真信念

什么是知识?这是一个经久不衰的话题。知识来源于信念,特别是文献[13]提出“我们的信念构成了我们对世界认识的很大一部分。我们对事物、文化、过去和未来都有信念。”一般来说,许多学者(如勒内·笛卡尔^[14])认为知识是一种合理的真信念^[15-16]。知识似乎无所不能、无所不知,因此,知识在一般的理解中被当作计算机视觉的超完备数据库(如多重知识表达法^[17],侧重于不同抽象层次、不同来源和不同方面的知识融合)或强大的监督源(如视觉知识^[18],以表示法抽象和多重补充为指导),而知识可以不与具体的视觉任务或场景关联,也可以不考虑具体的视觉任务或场景。简而言之,知识本身与视觉无关。但事实真的如此吗?

如图1所示,给定一幅图像,向一个强大的人工智能模型(即用超完备数据和类人监督训练的GPT-4o)提问“这些人在做什么?危险吗?(What are the people doing? Is it dangerous?)”,提示指令是“回答不超过10个字”。GPT-4o输出的回答为“人们在海滩上赶海。这并不危险。(People are clamming at the beach. It's not dangerous.)”。事实上,这是暴风雨前的宁静时刻。日落时分,人们确实在享受落日的余晖和海风的吹拂。然而,此时很可能正值涨潮,在海边玩耍可能会有危险。拥有丰富知识的强人工智能模型即使能够感知人们及其行为,但也无法理解在这种情况下是否存在风险。相比之下,我们的研究成果KM^{4[19]}则能感知到一定程度的危险,但KM⁴并不比GPT-4o强大,其训练所需的知识并不比GPT-4o多。不过,KM⁴所需的知识与特定的视觉场景或任务相关。因此,在视觉

领域,知识不仅可以作为类似超完备字典的知识库或强大的监督源,还可以作为经过良好校准并与特定视觉场景或任务相关联的信息源。



Question: What are the people doing? Is it dangerous?
GPT-4o: People are clamming at the beach. It's not dangerous.
KM⁴: People are playing around. It is a certain danger.

图1 GPT-4o和KM^{4[19]}的视觉推理结果

2 知识还是一种思维框架

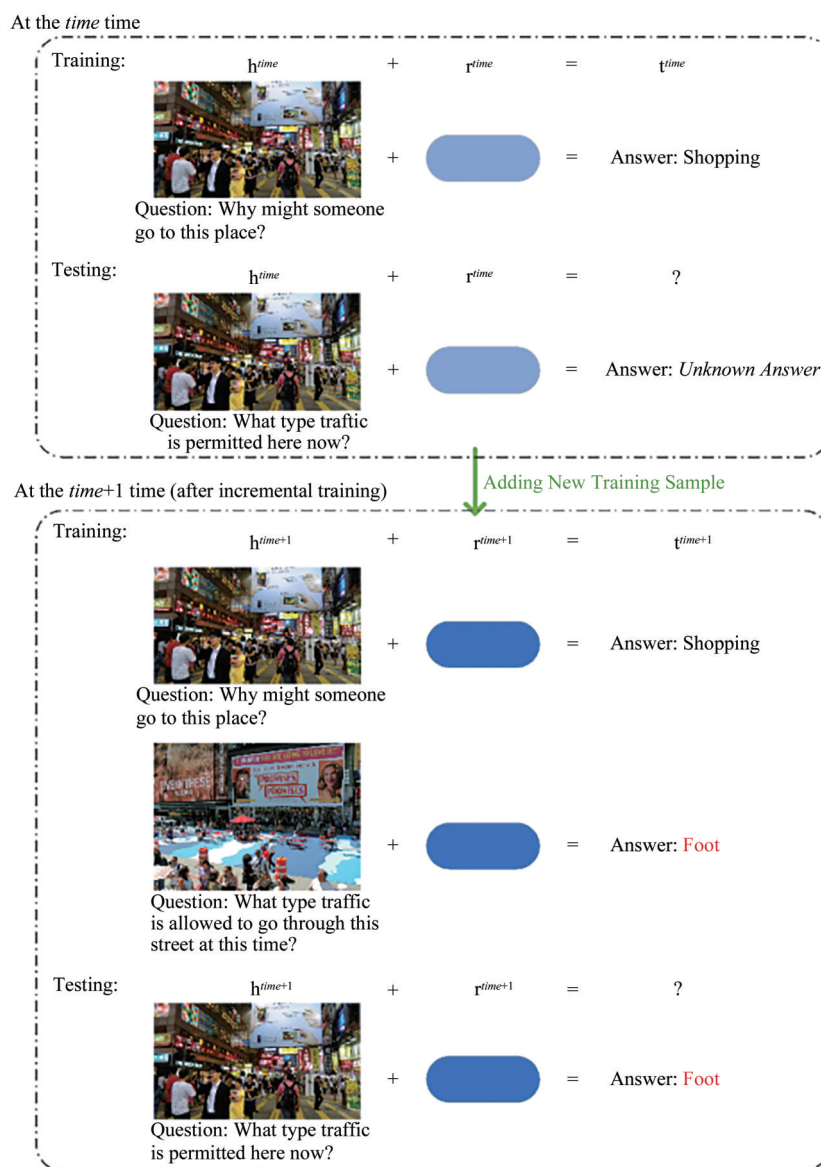
若把图1的夕阳抹去,GPT-4o的回答比KM⁴更合理一些,因为对于人类而言,快乐和享受的概念在大脑中主要与无危险相关联。这恰恰对应了文献[20]提出的“我们能够思考事物的唯一方式就在于那些使事物能够被我们思考的途径中”,这一点也与伊曼努尔·康德的观点^[21]相似,知识是我们对事物进行科学或日常体验和思考的唯一可能框架^[20-21]。正如我们之前的工作^[22-23],我们对于视觉任务采用如下解决策略:将与特定视觉场景或任务相关的知识嵌入视觉解决方案中,并允许其随着模型的发展而变化(演化)。

例如,我们提出了一种新式视觉推理模型^[23],如图2所示,通过将知识以计算三元组的形式纳入视觉模型,将知识作为思维框架,为视觉推理模型赋予认识未知事物的能力。这种方式为解决视觉问题或任务提供参考。另外,根据可计算性理论^[24],无论什么知识,都必须是计算性的^[25],让我们迈向计算知识视觉吧!

3 迈向计算知识视觉

计算机视觉是人工智能的关键途径之一,也是现代智能系统的重要组成部分。目前,计算机视觉系统进行视觉推理和因果推理的能力非常有限,而且高度专业化。因此,模型需要针对特定任务在大量训练数据上进行训练。然而,这样训练出来的模型的理解能力仍然无法达到普通成年人的常识水平,无法描述物理世界和社会世界的运作。人类能够透彻地理解视觉场景,是因为他们掌握了许多领域的先验知识,并能根据这些知识进行有效的学习和推理。

为了解决这些问题,我们提出了计算机视觉的

图2 新式视觉推理模型^[23]

新观点——计算知识视觉^[25-27]。向视觉感知模型注入先验知识，使模型获得一定的理解力，从而像普通成年人一样进行有效的学习和推理。具体来说，计算知识视觉包括两个模型，即视觉感知模型和结构化知识模型。其中，结构化知识模型可以通过总结人类经验或发现人类数据中的知识来构建。计算知识视觉通过设计知识注入和条件反馈机制，实现视觉感知模型和结构化知识模型之间的交互，最终使两个模型协同训练，完成特定的视觉任务。希望我们的著作《计算知识视觉与视觉推理》^[1]能开启视觉人工智能新时代的大门，也希望更多的学者能为这条新“路”而努力。

参考文献：

- [1] 郑文博, 王飞跃. 计算知识视觉和视觉推理[M]. 北京: 清华大学出版社, 2024.
ZHENG W B, WANG F Y. Computational knowledge vision and visual reasoning[M]. Beijing: Tsinghua University Press, 2024.
- [2] SCHOENEMANN B, PÄRNASTE H, CLARKSON E N K. Structure and function of a compound eye, more than half a billion years old[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(51): 13489-13494.
- [3] LI Z P. Understanding vision: theory, models, and data[M]. Oxford: Oxford University Press, 2014.
- [4] GAZZANIGA M, IVRY R B, MANGUN P H D G R. Cognitive neuroscience: the biology of the mind[M]. New York: W. W. Norton & Company, 2008.
- [5] TORRALBA A, ISOLA P, FREEMAN W T. Foundations of computer

- vision[M]. Cambridge: MIT Press, 2024.
- [6] MARR D. Vision: a computational investigation into the human representation and processing of visual information[M]. Cambridge: MIT Press, 2024.
- [7] ZHENG W B, YAN L, ZHANG W W, et al. D→K→I: data-knowledge-driven group intelligence framework for smart service in education metaverse[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(4): 2056-2061.
- [8] ACHIAM J, ADLER S, AGARWAL S, et al. GPT-4 technical report[EB]. arXiv preprint, 2023, arXiv:2303.08774.
- [9] ZHU S C, WU Y N. Computer vision: statistical models for Marr's paradigm[M]. Cham: Springer, 2021.
- [10] ZHU S C, HUANG S Y. Computer vision: stochastic grammars for parsing objects, scenes, and events[M]. Cham: Springer, 2021.
- [11] ZHU S C, ZHU Y X. Cognitive models for visual commonsense reasoning[M]. Cham: Springer, 2021.
- [12] WEI J, WANG X Z, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM, 2022: 24824-24837.
- [13] NILSSON N J. Understanding beliefs[M]. Cambridge: MIT Press, 2014.
- [14] DESCARTES R. Meditationes de prima philosophia[M]. Cambridge: Cambridge University Press, 1642.
- [15] RUSSELL B. History of western philosophy[M]. London: Routledge, 2009.
- [16] COHEN H, LEFEBVRE C. Handbook of categorization in cognitive science[M]. Cambridge: Elsevier, 2017.
- [17] YANG Y, ZHUANG Y T, PAN Y H. Multiple knowledge representation for big data artificial intelligence: framework, applications, and case studies[J]. Frontiers of Information Technology & Electronic Engineering, 2021, 22(12): 1551-1558.
- [18] PAN Y H. On visual knowledge[J]. Frontiers of Information Technology & Electronic Engineering, 2019, 20(8): 1021-1025.
- [19] ZHENG W B, YAN L, GOU C, et al. KM4: visual reasoning via knowledge embedding memory model with mutual modulation[J]. Information Fusion, 2021, 67: 14-28.
- [20] MARTIN R M. Epistemology: a beginner's guide[M]. London: One-world Publications, 2014.
- [21] KANT I, GREGOR M J. Critique of practical reason[M]. Cambridge: Cambridge University Press, 1997.
- [22] ZHENG W, YAN L, CHEN L, et al. Knowledge-embedded mutual guidance for visual reasoning[J]. IEEE Transactions on Cybernetics, 2024, 54(4): 2579-2591.
- [23] ZHENG W, YAN L, WANG F Y. Knowledge is power: open-world knowledge representation learning for knowledge-based visual reasoning[J]. Artificial Intelligence, 2024, 333: 104147.
- [24] TURING A M. On computable numbers, with an application to the Entscheidungsproblem[J]. J. of Math, 1936, 58(345-363): 5.
- [25] ZHENG W, WANG F Y. Computational knowledge vision[M]. Cambridge: Elsevier, 2024.
- [26] ZHENG W B, YAN L, GOU C, et al. Computational knowledge vision: paradigmatic knowledge based prescriptive learning and reasoning for perception and vision[J]. Artificial Intelligence Review, 2022, 55(8): 5917-5952.
- [27] 王飞跃. 计算知识视觉: 迈向智能视觉推理的新范式[J]. 智能系统学报, 2021, 16(3): 393.
- WANG F Y. Computing knowledge vision: a new paradigm for intelligent visual reasoning[J]. CAAI Transactions on Intelligent Systems, 2021, 16(3): 393.

[作者简介]



郑文博 (1995-), 男, 博士, 武汉理工大学计算机与人工智能学院副研究员, 主要研究方向为计算机视觉、机器学习。



王飞跃 (1961-), 男, 博士, 中国科学院自动化研究所研究员, 复杂系统管理与控制国家重点实验室主任, 澳门科技大学特聘教授, 主要研究方向为平行系统的方法与应用、知识自动化。