大模型时代的视觉知识：回顾与展望

王文冠，杨易，潘云鹤
浙江大学计算机科学与技术学院，中国杭州市，310027

摘要：视觉知识是一种新型知识表达形式，其理论之根深植于认知科学；视觉知识旨在为视觉智能的核心要素——如视觉概念、视觉关系、视觉操作和视觉推理——提供统一、全面且可解释的理论框架和建模方法。认知科学的研究实证了视觉相关知识在人类认知过程和智能行为中扮演着不可或缺的角色，由此可以推断，视觉知识的表达与学习将对发展视觉智能和机器智能起到重要作用。近年来，人工智能不断取得进步，尤其是人工智能大模型涌现出超越传统模型的智能水平，大模型能够自动从海量数据中发现普遍性规律，并将这些规律编码进超大规模神经网络的参数之中，实现了大规模知识自动提取和隐式知识参数化存储。这场由大模型驱动的新一轮人工智能技术革命，将为构建具备视觉知识的先进智能体带来新的机遇和挑战。对此，本文深入剖析视觉知识的理论基础，系统性回顾近年来视觉知识相关领域的发展状况。同时，针对大模型时代下视觉知识的发展方向以及其可能发挥的关键作用，提出前瞻性观点和展望。

关键词：视觉知识；人工智能；基础模型；深度学习
https://doi.org/10.1631/FITEE.2400250