



## Views &amp; Comments

## AI 的多重知识表达

潘云鹤<sup>a,b</sup><sup>a</sup> Institute of Artificial Intelligence, Zhejiang University, Hangzhou 310027, China<sup>b</sup> Zhejiang Lab, Hangzhou 311121, China

20世纪70年代认知心理学已经认识到, 长期记忆的内容应为语义和情景的集成[1], 并被分别编码为言语和心象的表征[2]。1991年, 笔者在论文《形象思维中的形象信息模型的研究》[3]中指出, 并非所有言语命题都能从言语系统中推理而获得, 很多只能从形象系统中转化而来。两个月前, 笔者在论文《论视觉知识》[4]中提出了视觉知识(VK)的概念, 它包括视觉概念、视觉命题和视觉叙事。视觉知识可以模拟人在大脑中对心象进行的各种时空操作, 如设计过程就充满了此类操作[5]。

而且, 令人可喜的是, 现有的计算机技术也已经提供了表达和推演视觉知识的有关技术基础。为此, 当今AI的研究者需要把视野从传统的AI领域(包括深度学习), 扩大到计算机图形学和计算机视觉等密切相关的技术领域, 这三个领域的研究者特别需要联手研究视觉知识。当视觉知识在AI中登堂入室之后, 原先那些大量无法用言语系统的推理得出的言语命题, 就可能从视觉知识转化而来。所以, 用言语知识和视觉知识的双重表达, 能更完整地描述世界的存在与发展, 解决更复杂的智能计算问题。由此观之, 视觉知识的表达与推演是AI走向2.0的重要技术[6]。

建立视觉知识之后, 在AI 2.0中的知识就有了三种表达。现列出这三种知识的表达与处理方法如下。

(1) 知识的言语表达。其特点是使用符号数据, 结构清晰, 语义可理解, 知识可推理。其典型例子如语义网络、知识图谱等。

(2) 知识的深度神经网络表达。其特点是适用于图像、音频等非结构化数据的分类与识别。缺点是语义解释困难。其典型例子如深度神经网络(deep neural network, DNN)、卷积神经网络(convolutional neural network, CNN)等。

(3) 知识的形象表达。其特点是适用于图形、动画等形状、空间、运动的数据。知识的结构清晰, 语义可解释, 知识可推演。其典型如视觉知识等。

这三种知识表达之间的关系和传统AI中已经出现过的多种知识表达技术, 如规则、框架、语义网络等之间的关系, 有着本质的不同。因为AI的这三种表达是针对人类记忆中的三种不同的内容, 现说明如下:

(1) 知识图谱——语义的记忆内容, 宜用于字符检索与推理;

(2) 视觉知识——情景的记忆内容, 宜用于时空推演与显示;

(3) 深度神经网络——感觉的记忆内容, 宜用于对原始数据作逐层抽象的分类。

其中,(1)和(2)与人类长期记忆中的两大内容——言语的和心象的编码方式相对应。其中,(3)与人类短期记忆中的感知内容相对应。因此, 上述三种知识表达具有如同人类记忆中的信息特点一样, 具有内容系统互补的特点, 它们是需要被同时使用的, 因此组成为三重知识表达。

除了信息内容与使用的互补之外, 上述三重表达的另一重要性质是它们之间的相互联系与相互支持。视觉

知识通过投影变换可以将三维图形或动画信息转化为图像或视频信息。反之，通过3D重建技术，也可以将图像或视频信息转化为3D图形或动画信息。

视觉知识的语义是很清晰的，因此，可以用符号检索与匹配等技术把视觉知识与知识图谱进行语义的对应，从而实现转化。也就是说，视觉知识与知识图谱中的情景信息与语义信息的联系，可以通过表达的结构（指针）模型来实现；视觉知识与深度神经网络所用的图像与视频数据的联系可以用重建与变换计算来实现。

上述三重知识表达之间的互相联系与支持关系，以“猫”为例，如图1所示。

图1中的知识图谱表达了猫的种属上下关系；视觉知识表达了猫的形体、结构和动作等时空特征；深度神经网络表达了对所提供的猫的正负样本图像判别的一种抽象。

其实，当猫的形象符合其某种系统要求时（如同一

猫的不同观察视角），这些图像也可以重建为视觉知识。视觉知识可以通过变换（几何、投影、动作变换等）生成各种猫的形象供DNN学习。通过视觉知识与知识图谱联系可推知，因为猫、虎、豹同属猫科动物，所以具有类似的形体、结构与动作。因此，猫的视觉知识可通过合适的修改而形成虎、豹的视觉知识，这就以容易理解的方式实现了迁移学习的功能，也为小数据下的知识学习（如少样本、零样本学习）的小数据与模型的通用化打开了一扇大门。

本文提出由知识图谱、视觉知识和深度神经网络构成的AI三重知识表达的结构。其中的知识图谱、视觉知识分别擅长于处理字符性内容和形象性内容，DNN擅长对感性数据作层次抽象，并分别对应于模拟人类大脑中对长期记忆和短期记忆中的信息加工与处理。它们彼此能相互衔接、相互支持，从而有利于知识表达与推理等智能计算的可解释性、可推演性和可迁移性的实现。

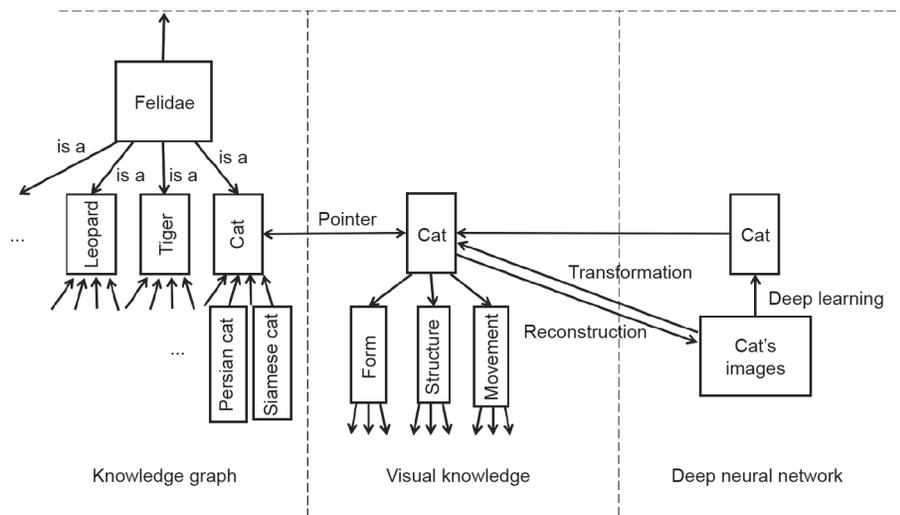


图1. 猫的三元知识表达及其关系。

## 致谢

庄越挺、吴飞、耿卫东、汤斯亮等教授为本文提供宝贵意见，特此表示感谢。

## References

[1] Tulving E. Episodic and semantic memory. In: Tulving E, Donaldson W,

editors. Organization of memory. New York: Academic Press; 1972. p. 381-403.  
 [2] Paivio A. Perceptual comparisons through the mind's eye. Mem Cognit 1975;3(6):635-47.  
 [3] Pan YH. An approach to information processing models of thinking with mental imagery. Pattern Recognit Artif Intell 1991;4(4):7-14. Chinese.  
 [4] Pan Y. Patterns of thinking process in scheme design about shapes. J Zhejiang Univ-Nat Sci 1993;27(3):363-70. Chinese.  
 [5] Pan YH. On visual knowledge. Front Inf Technol Electron Eng 2019;20(8): 1021-5.  
 [6] Pan YH. Heading toward artificial intelligence 2.0. Engineering 2016;2(4): 409-13.