Views & Comments

# Multiple Knowledge Representation of Artificial Intelligence

Yunhe Pan [a,b]

[a] *Institute of Artificial Intelligence, Zhejiang University, Hangzhou 310027, China*
[b] *Zhejiang Lab, Hangzhou 311121, China*

In the 1970s, cognitive psychology recognized that the information in the long-term memory is scene and semantic [1] and may be encoded in parallel as verbal and mental imagery [2]. In 1991, I pointed out that not all verbal propositions can be derived from the verbal system, and that many can only be transformed from the imagery system [3]. I have proposed the concept of visual knowledge, which consists of visual concepts, visual propositions, and visual narratives [4]. Visual knowledge can simulate the various spatiotemporal operations that a person can perform on a mental imagery in his/her brain, such as the design process [5].

Moreover, existing computing technologies have provided relevant technical support for expressing and deducing visual knowledge. To this end, artificial intelligence (AI) researchers need to expand their horizons from traditional AI fields (including deep learning) to closely related technologies such as computer graphics and computer vision. Hence, researchers in AI, computer vision, and computer graphics in particular must jointly study visual knowledge. Those original verbal propositions that cannot be inferenced from the verbal system alone might be transformed with the help of visual knowledge. Therefore, by depending on verbal knowledge and visual knowledge, we can more comprehensively describe the surrounding world and solve more complex problems. Hence, the expression and deduction of visual knowledge is an important technology for AI 2.0 [6].

Based on visual knowledge, a total of three kinds of methods represent knowledge in AI 2.0, as follows:

(1) **Verbal representation of knowledge.** Verbal knowledge is represented by symbolic data, and its structure is explicit, its semantics are understandable, and its knowledge can be reasoned. Typical verbal knowledge includes the semantic network and the knowledge graph.

(2) **Knowledge representation by deep neural network.** This kind of knowledge is suitable for the tasks of classification and recognition for unstructured data such as images, videos, and audios. However, it is difficult to interpret this kind of knowledge. Typical examples include deep neural networks (DNNs) and convolutional neural networks (CNNs).

(3) **Visual representation of knowledge.** This kind of knowledge can feasibly be dealt with using graphs, animation, and three-dimensional (3D) objects. Its structure (i.e., spatial-temporal structure) is explicit, its semantics are interpretable, and its knowledge can be deduced. A typical example is visual knowledge.

The relationship between above three knowledge representations is fundamentally different from those between various other former knowledge representations that have appeared in traditional AI, such as rules, frameworks, and semantic networks. The three kinds of knowledge representation correspond to three different aspects of human memory, as follows:

(1) The knowledge graph corresponds to semantic memory content. It is suitable for the retrieval and reasoning of symbolic data.

(2) Visual knowledge corresponds to the scene memory content. It is suitable for the deduction and visualization of spatiotemporal data.

(3) The DNN corresponds to the perception memory content. It is suitable for layer-wise abstraction of input data for classification.

Parts 1 and 2 correspond to the encoded information of verbal and mental imagery in human long-term memory. Part 3 corresponds to the perceptual information in human short-term memory. Thus, these three kinds of knowledge representation are complementary and must be utilized as a whole in working.

Another important property of these three kinds of knowledge representation is that they are interconnected and mutually supportive. Visual knowledge can transform 3D graphics or animation into an image or video via projection. Image or video information can also be converted to 3D graphics or animation through 3D reconstruction techniques.

Since the semantics of visual knowledge are clearly expressed, we can align visual knowledge and the knowledge graph. Therefore, visual knowledge and the knowledge graph can be transformed into each other via symbolic retrieval and matching. That is, the connection between scene information and semantic information in visual knowledge and the knowledge graph can be realized by a structural model. Moreover, visual knowledge and the image and video sample data used in a DNN can be connected via reconstruction and transformation.

Taking the knowledge of the cat as an example, Fig. 1 illustrates how to connect the three kinds of knowledge representation.

In Fig. 1, the knowledge graph expresses the cat's species relationship; the visual knowledge expresses the spatiotemporal characteristics of the cat, including its form, structure, and movement; and the DNN expresses an abstraction of cat images, including both positive and negative sample images.

In fact, visual knowledge of the cat can be reconstructed from cat images from different perspectives of the cat. By means of
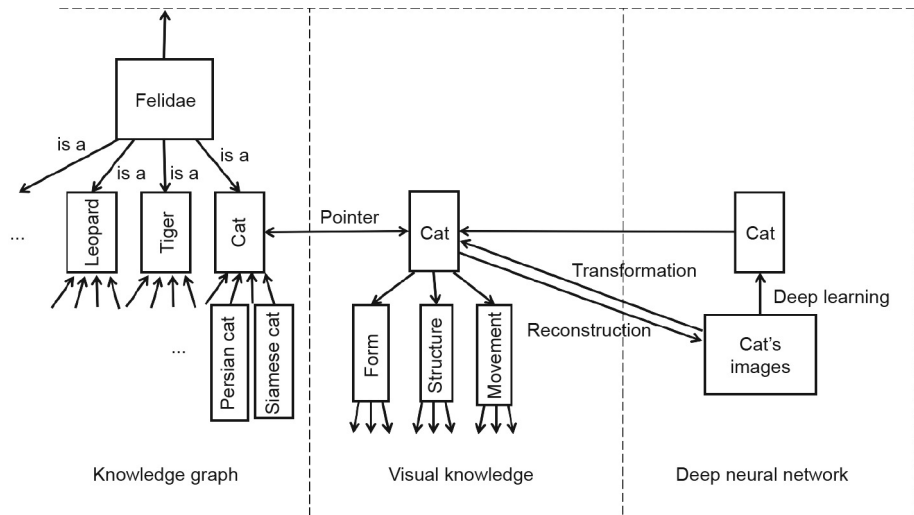
**Fig. 1.** Three kinds of knowledge representation with respect to the cat.

transformation (i.e., geometry projection and motion transformation), visual knowledge can generate many of the cat's images, which are helpful for DNN learning. Through the connection between visual knowledge and the knowledge graph, one can infer that cats, tigers, and leopards would share similar shapes, structures, and movements since they belong to the same family, as shown in Fig. 1. Therefore, visual knowledge of tigers and leopards can be captured through appropriate modifications of the cat's visual knowledge. In this way, we can realize the transfer of learning and find a way to learn a model when only small samples are available (such as zero-shot learning or few-shot learning).

In this paper, I proposes the multiple knowledge representation of AI, which consists of the knowledge graph, visual knowledge, and DNN. A knowledge graph and visual knowledge are capable of dealing with the textual descriptions and visual content, respectively, for a given concept, while a DNN is desirable to disentangle the hierarchical abstraction of visual information and therefore is similar to the information-processing mechanism in the long-term and short-term memory of the human brain. Multiple knowledge representation via a combination of knowledge graphs, visual knowledge, and DNNs will be beneficial to interpretable, evolutional, and transferable models for knowledge representation and inference.

## 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 Y. 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 Y. On visual knowledge. Front Inf Technol Electron Eng 2019;20(8):1021–5.
[6] Pan Y. Heading toward artificial intelligence 2.0. Engineering 2016;2(4):409–13.