Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

观点：

大数据人工智能下的多重知识表达：框架、应用及案例研究

杨易‡，庄越挺，潘云鹤

浙江大学计算机学院，中国杭州市，310027

E-mail: yangyics@zju.edu.cn; yzhuang@zju.edu.cn; panyh@zju.edu.

投稿日期：2021-9-30；录用日期：2021-10-7；Crosschecked：2021-11-22

摘要：提出一种多重知识表示框架，探讨了其对推动大数据人工智能技术在各个领域中发展的重要意义及深远影响。传统知识表达和现代基于深度学习的知识表达通常着眼于利用特定变换方式，将输入转换为符号编码或者向量。例如，知识图谱关注于描述各个概念之间的语义联系，而深度神经网络更像是感知原始信号输入的工具。多重知识表达是一种更为先进的人工智能表征框架，具备更完整的智能功能，比如原始信号感知、特征提取及向量化、知识符号化和逻辑推断。多重知识表达有如下两点优势：（1）与现有以深度学习为主导的人工智能技术相比，具有更强的解释性以及更好的泛化能力；（2）将多重知识表达集成于现有人工智能技术，有利于各种表征（例如原始信号感知以及符号化编码）发挥互补优势。我们希望多重知识表达相关研究以及应用能够驱动新一代人工智能蓬勃发展。

本文译自 YANG Y, ZHUANG YT, PAN YH, 2021. Multiple knowledge representation for big data artificial intelligence: framework, applications, and case studies. *Front Inform Technol Electron Eng*, 22(12):1551-1558. https://doi.org/10.1631/FITEE.2100463

1 多重知识表达

在本节中，我们将简要回顾几种典型知识表达方式，紧接着介绍多重知识表达框架（Pan, 2020）。

1.1 回顾知识表达

单一知识表达的方案通常侧重于使用特定变换方式，将输入转化为符号编码或者向量。我们首先回顾两种典型知识表达方法：传统知识表达和现代基于深度学习的知识表达。

1.1.1 传统知识表达

传统知识表达模型（例如生成式表达、一阶逻辑表达以及过程式表达）利用高度抽象化的概念作为输入，旨在建立概念之间的因果关系。在这类模型中，典型的抽象化知识/信息包括以下几类：

1. 陈述性知识（描述性知识）。这类知识通常以一些描述性语句的形式出现，而这些语句中可能包含与我们感兴趣事物相关的概念以及对事实的描述。

2. 过程性知识。以具体的任务为依据，这类知识包含与之相关的规则、策略、过程，通常用于推理。

3. 启发式知识。启发式知识包含基于专家、过往经历以及其他资源产生的一些经验法则。

4. 结构化知识。结构化知识刻画了不同概念和事物之间的关系。传统知识图谱就是表达结构化知识的一种典范。

1.1.2 深度知识表达

基于深度学习的知识表达方法首先将接收的原始信号（通常位于相对低级的抽象层，如直接采集到的视觉、音频信号）作为输入，然后利用

‡ 通讯作者

* 本文得到以下项目资助：国家重点研发计划（No. 2020AAA0108800）

ORCID: 杨易，https://orcid.org/0000-0002-0512-880X

深度神经网络（如深度卷积神经网络（CNN）（Krizhevsky et al., 2012; He et al., 2016）和Transformer模型（Vaswani et al., 2017）），将原始信号编码为特征向量。目前，这类以深度学习为范式的知识表达学习在大数据人工智能研究中占据主导地位。

在处理图像、视频、音频、文本或时间序列等非结构化数据的各类任务（如分类、预测）中，基于深度学习的知识表达表现出优异性能。与传统表达方式相比，基于深度学习的表达在从大规模数据中解构提取信息/知识的能力上更胜一筹。然而，现有基于深度神经网络的算法无法很好地将过程性知识以及结构性知识抽象化，因此限制了其推理能力。此外，深度神经网络最大劣势在于其黑盒属性，导致其输出缺乏解释性（Arrieta et al., 2020），近年来也最为学者所诟病。这一点严重制约了深度神经网络的应用场景，尤其是可信性受到关注的应用场景，例如医疗领域中涉及决策制定等相关环节。

1.1.3 讨论

在实际问题中，上文所述的两类知识表达作为单一的从输入信号中提取知识的方法，各自具有一定局限性。传统知识表达依赖于先验的符号化知识。尽管在漫长的文明史中，人类积累了大量的知识，但人类创造的符号系统与其对现实世界的综合认知之间仍然存在比较显著的差距。因此，计算机能处理的抽象化知识例如向量、符号等无法涵盖全部有用的信息，也就无法为进一步的智能计算与推理提供支撑。另一方面，虽然基于深度学习的表达能够从数据中学到隐式的知识，但是这类表达缺乏常识支撑，无法用于进一步推断，并且以目前的形式也不能表达过程性、结构化的知识。

为了全面理解一个概念，人们通常会借助于多种知识包括直觉上的感知、认知、高度抽象的知识以及逻辑先验（Pan, 2020）。在人类日常生活（如学习、决策制定）中，人们也常常借助不同来源的多类别知识之间的互补优势。这一事实表明，使用某种合理机制（Pan, 2020）来融合多种知识表达的多重知识表达（MKR）框架将是推动新一代人工智能时代向更高级智能计算发展的重要方向之一。

1.2 多重知识表达框架

多重知识表达框架旨在获取、表达、使用从不同来源或者不同方法中获得的位于不同抽象层次的知识。潘云鹤院士的开拓性工作（Pan, 2020）中描述了一些多重知识表达框架雏形。图1阐述了多重知识表达框架的主要特点及其重要组成部分。

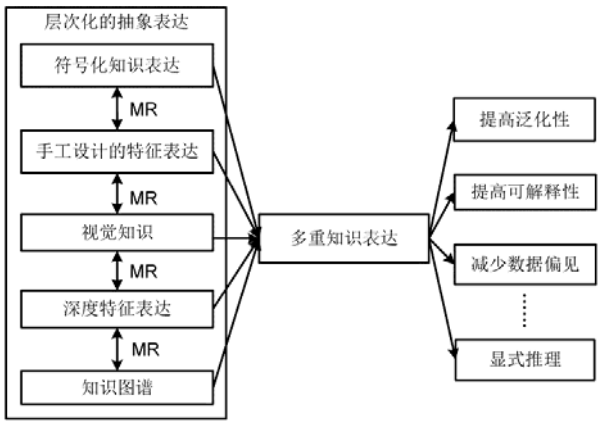

图1 多重知识表达整体框架一例（MR：相互增强）

1.2.1 多重表达融合

多重知识表达框架不仅结合了多种知识表达，同时采用某种合理机制将其融合。每一类知识都有其内在的优势。目前针对多重知识表达的研究主要包含以下几种形式：符号化知识表达、知识图谱、手工设计的特征表达以及基于深度学习的表达。绝大多数人类知识存在于上述四种知识表达中。具体而言，符号化的知识表达显式地依赖于专家预先定义的一些概念及其之间的因果关系（有些时候这些关系会变得相当复杂）。而知识图谱包含了一组互相关联的实体描述以及实体对或者实体链之间的关系。这两类表达适合于描述高度抽象化的概念以及它们之间的关系（例如逻辑或者语义之间的关系）。手工设计的特征表达以及基于深度学习的表达更倾向于从数据中获取表征，在从原始信号中提取高层语义表达方面更具优势。由于深度特征表达的学习由数据驱动，使得该表达天然具有提取丰富信息的能力，这也是现有符号系统所缺少的。正如 1.1 小节中讨论的，我们设计的多重知识表达旨在扬长避短地利用各个知识表达。在未来的研究中，我们考虑将一些新型知识表达——如潘云鹤院士提出的视觉知识（Pan，2020）——融合到多重知识表达框架。

1.2.2 多层次知识抽象

多重知识表达对不同抽象层次的知识进行编码。在本小节中，抽象层次指的是提取的表达保留重要信息并丢弃一些平凡细节的程度。

在大多数情况下，人类从感知到认知是一个由浅入深的过程。我们以人类识别动物类别为例，人们本能地倾向于先观察动物的外貌与叫声。这些通过人的感官直接获得的信息——比如动物的颜色、体型以及牙齿形状——通常包含很多细节，因此属于比较低级的抽象表征。更高级的抽象化知识包含动物的生活习性以及分类法则。在上面的例子中，低层和高层抽象知识（例如动物的外貌和习性）都是必不可少并且互为补充的。两者的融合与任一抽象知识相比，所能产生的表达更为全面。

多重知识表达对多层次的知识进行融合。近年来，大多数人工智能应用依旧着眼于处理相对低级的抽象表达，因此基于深度学习的知识表达越来越受到人工智能研究的青睐。当然，高层次的抽象如常识、逻辑推理也常常被纳入考量。多重知识表达将不同抽象层次的表征结合在一起，因而可支持具有从感知、识别到关联、推理等更多功能的人工智能系统。

1.2.3 多模态知识增强

多重知识表达能够通过不同知识表达之间的交互增强各自的表征，形成耦合性更强的深层表征。值得注意的是，多重知识表达并不是多种不同表达之间的简单组合。这里，我们以计算机视觉研究为例：在机器感知任务中，特征表达，特别是近几年来兴起的基于深度学习的深度特征表达（He et al., 2020; Sun et al., 2020），具有较强的鲁棒性。与符号化的知识相比，深度特征表达通常包含了更多针对视觉细节的描述。另一方面，符号化的知识能够弥补深度特征表达泛化性差的不足。例如，如果我们将“汽车具有不同颜色”这一符号化知识和红色小汽车的视觉特征相结合，送入 AI 系统，那么这一系统能很容易地识别出黑色小汽车。同样，基于“观察到小汽车具有不同颜色”这一视觉先验，符号系统可以确信一个手工制品同样可以被涂上不同的颜色。因此，多重知识表达研究的一个关键点在于如何使不同的表达互相增强，从而得到更全面的表征。

2 相关应用和案例研究

根据任务的目的或者采用的方法，近期一些新兴研究可被视为多重知识表达的早期尝试。

1. 视觉理解（Pan, 2021)。深度神经网络是一种非常强大的特征提取器。结构化信息通常能够提供合理的补充线索，促进系统对视觉内容的理解。结构化的表达常常被用于处理结构化的视觉理解。例如，Xu 等人（2017）提出将视觉场景表示为图的结构，包含物体、属性以及物体间的关系。场景图组成了具备可解释性以及完善结构的图片表征。研究人员还使用专有信息（privileged information）作为辅助特征来协助模型训练，并利用各种类型的专有信息来促进学习。例如，Yan 等人（2016）将文本作为专有信息，并同时利用视觉和文本特征进行主动样本选择。一种典型的多视觉线索整合方式是简单地应用后期融合（late fusion）。例如，自两分支动作识别模型提出以来，将光流模型和 RGB 模型融合的方法（Simonyan and Zisserman, 2014; Zhu et al., 2021）在各领域得到广泛应用。最近的研究中，Wang 等人（2020）考虑使用多分支框架来解决第一视角动作识别问题。这一工作采用注意力机制将多种视觉线索动态地整合在一起，并考虑它们之间的交互信息并挖掘其内在联系。在多模态分析（例如视觉对话）中，叙述文本的结构信息有待进一步探索。Fan 等人（2020）提出一种对话网络，用于学习上下文的叙述结构。该网络从句子层级的辨别器中学习知识，并利用该知识引导对生成模型的训练，减缓了词层面的过拟合问题，从而提升生成句子在语义上的连贯性。

2. 视觉知识辅助计算机图形学。计算机图形学研究的是数字化合成、操纵视觉内容的过程。近年出现的生成对抗网络（GAN）（Goodfellow et al., 2014）已成为视觉生成的一个非常好的替代品。随后的工作将 GAN 继续拓展到文本到图像、图像到文本、文本到视频以及视频到文本的生成中。作为深度学习方法之一，GAN 同样需要大量的训练数据，且其模型缺乏可解释性。为了弥补这一缺陷，一些近期的工作（Johnson et al., 2018; Gogoglou et al., 2019）考虑加入结构化知识，从而更好地控制生成过程。具体而言，Gogoglou 等人（2019）尝试控制生成物体的位置、属性以及类

别。Johnson 等人（2018）提出借助场景图来生成图像的方法。对场景图的使用显式地将物体之间的联系编码到生成过程中。

3. 多模态知识图谱。互联网在迅速发展过程中产生海量多模态数据。为了从这些海量数据中学习知识表达，DBpedia（Auer et al., 2007），Wikidata（Vrandečić and Krötzsch, 2014）以及 IMGpedia（Ferrada et al., 2017）分别建立了大规模知识图谱。然而，互联网数据通常包含许多噪声，并具有一定程度的数据偏见。高质量图片、视频以及音频数据比文本数据更为稀缺，因此利用多重知识表达挖掘跨模态的联系对提高多模态知识图谱的质量具有重要意义。

4. 神经符号网络。一些研究者提出将 DNN 与符号表达相结合形成混合网络，称为神经符号网络。早期工作（França et al., 2014）将符号化的知识编码到神经网络的权重中。具体做法是使用被编码为初始命题逻辑程序的背景知识来构建循环神经网络。同时，神经符号网络通过标准的反向传播来从数据中学习。这样一种结合方式兼具神经网络并行训练的好处以及命题逻辑的强大表征能力。近期工作（Serafini and d'Avila Garcez, 2016）提出统一框架——逻辑张量网络——来集成自动学习及推理。该网络在 DNN 中实现了“真正的”逻辑表达，因而得以同时从符号知识的演绎推理和数据驱动的机器学习中获益。

这些工作较早关注到符号知识和深度特征表征的结合。而多重知识表达通常有着更深远的研究目的，这一结合可视为多重知识表达的某种特殊情况。在表 1 中，我们对现有工作中使用到的表征进行了比较。正如第 1 节中提到的，多重知识表达不仅将多类型多层次的表达集成到一起，同时对各个组成部分进行增强，形成更完整、耦合性更强的表达。

表 1 近期工作中使用的几种表征

方法	S	H	V	D	K
场景图（Xu et al., 2017）	√	×	×	√	×
IMGpedia（Ferrada et al., 2017）	×	√	×	×	√
逻辑张量网络（Serafini and d'Avila, 2016）	√	×	×	√	×
多重知识表达（本文）	√	√	√	√	√

S：符号化知识表达；H：手工设计的特征表达；V：视觉知识；D：深度特征表达；K：知识图谱

3 大数据人工智能时代下从深度特征表达到多重知识表达

在过去十年里，大数据下深度学习的兴起推动 AI 迅猛发展，给学术界与工业界都带来深远影响。基于深度学习的表征使 AI 研究焕然一新，几乎在横跨语音识别、计算机视觉、自然语言处理以及机器翻译等多个领域的不同应用中占据主导地位。然而，深度神经网络由数据驱动且具有天然的黑盒属性，这为深度特征表达的发展带来一系列问题。多重知识表达增强了不同表达之间的优势，为解决上述问题提供了可行解。在本节中，我们将从泛化性和可解释性两个方面出发，探讨多重知识表达如何突破当下以深度学习为主导的人工智能研究的困境，同时我们将通过一些案例证明多重知识表达能够解决单一表达无法解决的问题。

3.1 多重知识表达提高泛化性

多重知识表达通过以下两种方式来提高泛化能力：首先，多重知识表达能够利用符号化知识（例如知识图谱）来消除数据的偏置；其次，多重知识表达提高了知识迁移的能力，促使从标注完善的数据中学到的知识能够更好地迁移到标注欠佳的数据甚至是完全没见过的新数据上。

1. 数据偏置指的是因数据集中的某类数据相较于其他数据而言被赋予的权重更高而导致的一类误差。作为一大主要挑战，数据偏置抑制了由数据驱动的 AI 算法泛化能力的提高。这一偏置的存在不仅会降低算法预测的准确性，同时会带来一些涉及道德和公平性的问题。一个著名例子是人脸识别算法的准确性会受到肤色偏见的影响。解决这一问题的其中一个方法是将符号知识的先验表达与深度学习学到的特征表达相结合，从而避免因肤色导致的数据偏置的影响。另一个例子是 Tang 等人（2020）的工作，他们将结构化的知识表达嵌入到深度表征中，以此来削减推理阶段的偏置。

2. 知识迁移是将先前学到的知识运用到新的问题中。在先前问题中学到的知识与现有新问题之间通常具有一定关联性，而值得注意的是，新旧问题常常属于不同领域。知识迁移是一种提高模型泛化能力非常有效的方式，其面对的最主要

问题是新旧任务之间存在所属领域的差异。多重知识表达能够增强符号知识表达和深度学习特征表达，将领域差异解耦为多个潜在因素，然后过滤掉不相关的干扰因素，最终提取出对新问题最有用的知识。例如，在行人检测任务中，基于深度学习的算法非常容易受服装样式变化干扰。如果结合“服装样式在行人检测中是不相关因素”这一符号知识，检测算法将丢弃服装样式信息，进而提高模型鲁棒性。近期研究结果表明，在深度学习中引入人体结构化信息可提高模型对行人识别的准确性（Miao et al., 2021）。

3.2 多重知识表达提高解释性

制约深度学习的另一大因素是其黑盒属性。即使是模型设计者也无法了解算法给出某个决策的缘由。然而，模型缺乏解释性意味着模型无法为其决策负责，这大大限制了 AI 的应用场景，尤其在那些安全性需要得到一定保证的领域中（Amodei et al., 2016）。

与黑盒 AI 系统相对的，白盒 AI 系统或者说可解释性 AI（XAI）系统指的是能够清晰展现决策过程的透明模型。因此，这样一个透明 AI 系统作出的决策不仅取决于输入数据，还必须受到可被人类理解的一些机制的限制，例如某种能够体现人类知识的正则化机制。多重知识表达正是提供了这样一种机制，将从数据中获取的知识与人类先验的符号知识相结合，因而成为构建 XAI 系统的首选项之一。

3.3 多重知识表达带来的变化

在本节中，我们将讨论近期一些与多重知识表达结合的相关研究，证实引入知识表达可带来更好的泛化性及可解释性。首先，我们展现了多重知识表达为现有人工智能研究带来新方向。然后，我们讨论了多重知识表达的出现为金融投资评估带来的变化。

3.3.1 符号知识增强合成图像多样性

在计算机视觉领域中，为满足模型对大数据训练的需求，研究者们将合成数据作为大量人工标注数据的补充（de Souza et al., 2017; Veeravasarapu et al., 2017; Singh and Zheng, 2020）。在合成新图像的过程中，多重知识表达的应用能够提高数据多样性并引入有用的符号知识，因而新合成的图像有助于 AI 模型的训练。下面我们将介绍 3 个例子：

1. 利用气候、地理等知识，图像生成模型能够更精细地生成不同天气、场景的新图像。通过在模型的训练过程中使用这些新生成的图像，AI 系统将获得更丰富的视觉知识，因而能消减因季节气候变化带来的识别图片场景变化过大的影响。

2. 运用动物的身体结构信息以及运动学知识，生成模型能够依据动物静止体态生成其在走、跑、跳等运动状态下的姿势。这些新合成的数据帮助 AI 系统学到不同动物之间体态、运动姿态的内在联系。

3. 结合光的折射、漫反射原理，图像生成引擎能够模拟在不同光照条件下各种材料的色泽形态。将这些新合成的数据加入训练数据中，有助于解决当前 AI 系统面临的因光照条件不同带来的数据域适应（domain adaptation）问题。

3.3.2 智能 AI 教育中的学生自动分组问题

在智能 AI 教育系统中，学生自动分组是重要问题之一。教育领域的数据具有异构化特性，阻碍了学生自动分组算法的发展。学生分组依赖于多种异构信息，通常以音频、文本、视频以及结构化表格等形式出现。这些信息包括学生在学习平台上的个人活动、组员之间的合作对话、学生与教师互动的历史记录以及其他与学生学习经历相关的信息。分组的质量极大地影响了学生的参与度，继而影响了后续小组作业的交流、教师的教学管理等等。多重知识表达通过构建图结构以及增强知识表达能力，从异构线索中自动发掘其因果关系、结构依赖，从而提高自动分组质量。

基于多重知识表达的自动分组机制首先将学生与学生、教师与学生之间的交互联系以及教育专家知识图谱嵌入同一个图空间中。然后，以符号知识图谱为引导，按照多重知识推断的流程提取出节点间的因果关系并抽象出多层知识表达。最终，我们将给出具备可解释性的学生与学生之间的关系图，其中图的边代表了分到同一组的权重。每对权重代表了统计意义上两个学生之间的关联性，这也可以从教育专家知识图谱上最近邻节点上得到相应的解释说明。将最终分组的推荐

结果可视化以供教师决策。在这一例子中，多重知识表达起到了将来自学生和学生、学生和教师之间产生的异质数据对齐的重要作用。更进一步，多重知识表达能够为在线教育系统实现自动分组提供支持，进而优化学习环境。

3.3.3 多重知识表达为金融投资评估提供更强的解释性

在金融领域，智能投资顾问是一款非常重要的 AI 应用。智能投资顾问主要负责利用投资组合理论，根据顾客的投资兴趣推荐合适的投资产品。对于个人投资建议而言，追求的结果通常是收益最大化。然而，对于国家投资而言，所要考虑的绝非止于此，相关因素包括地区发展的平衡性、缩小贫富差异、环境保护与可持续发展等。在这一问题上，多重知识表达的应用有助于为智能系统提供更合理的投资建议。例如，为平衡地区发展，智能投资顾问需要根据不同地区之间的地理环境、工业基础甚至居住人口的差异来动态调节策略。同样，智能投资顾问可以利用生物、地球科学等多重知识，尽可能减轻可能导致的环境危害，实现稳定的可持续发展。

4 结论

本文介绍了一种基于多重知识表达的框架级其相关应用和案例研究。多重知识表达作为新一代知识表达范式，从不同抽象层次、不同来源及不同方面的知识中学习有用的表达。这些知识表达互相增强，形成更完备、更强大的表征。大数据人工智能时代下的多重知识表达不仅能够提高传统任务（如检测、分类）的性能，更是赋予 AI 系统更完善、更丰富的功能及特性，如更好的泛化能力、更强的解释性和更强的推理能力。我们希望多重知识表达的出现能够助力新一代人工智能蓬勃发展，驱动人工智能技术登上新台阶。

贡献声明

潘云鹤提出主要思想并主导本研究。杨易和庄越挺查阅相关资料。杨易、庄越挺、潘云鹤进行了深入探讨，共同起草、修改并定稿。

致谢

作者对孙奕帆、朱霖潮、汪晓晗和武宇博士的建设性意见表示衷心感谢。

遵守伦理准则声明

作者声明发表这篇论文没有利益冲突。

参考文献

Amodei D, Olah C, Steinhardt J, et al., 2016. Concrete problems in AI safety. https://arxiv.org/abs/1606.06565v2

Arrieta AB, díaz-Rodríguez N, Del Ser J, et al., 2020. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inform Fus*, 58:82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Auer S, Bizer C, Kobilarov G, et al., 2007. DBpedia: a nucleus for a web of open data. Proc 6th Int Semantic Web Conf and 2nd Asian Semantic Web Conf the Semantic Web, p.722-735. https://doi.org/10.1007/978-3-540-76298-0_52

de Souza CR, Gaidon A, Cabon Y, et al., 2017. Procedural generation of videos to train deep action recognition networks. IEEE Conf on Computer Vision and Pattern Recognition, p.2594-2604. https://doi.org/10.1109/CVPR.2017.278

Fan HH, Zhu LC, Yang Y, et al., 2020. Recurrent attention network with reinforced generator for visual dialog. *ACM Trans Multim Comput Commun Appl*, 16(3):78. https://doi.org/10.1145/3390891

Ferrada S, Bustos B, Hogan A, 2017. IMGpedia: a linked dataset with content-based analysis of Wikimedia images. Proc 16th Int Semantic Web Conf on the Semantic Web, p.84-93. https://doi.org/10.1007/978-3-319-68204-4_8

França MVM, Zaverucha G, d'Avila Garcez AS, 2014. Fast relational learning using bottom clause propositionalization with artificial neural networks. *Mach Learn*, 94(1): 81-104. https://doi.org/10.1007/s10994-013-5392-1

Gogoglou A, Bruss CB, Hines KE, 2019. On the interpretability and evaluation of graph representation learning. https://arxiv.org/abs/1910.03081

Goodfellow IJ, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.

He KM, Zhang XY, Ren SQ, et al., 2016. Deep residual learning for image recognition. IEEE Conf on Computer Vision and Pattern Recognition, p.770-778. https://doi.org/10.1109/CVPR.2016.90

He KM, Fan HQ, Wu YX, et al., 2020. Momentum contrast for unsupervised visual representation learning. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.9726-9735. https://doi.org/10.1109/CVPR42600.2020.00975

Johnson J, Gupta A, Li FF, 2018. Image generation from scene graphs. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.1219-1228.

https://doi.org/10.1109/CVPR.2018.00133

Krizhevsky A, Sutskever I, Hinton GE, 2012. ImageNet classification with deep convolutional neural networks. Proc 25th Int Conf on Neural Information Processing Systems, p.1097-1105.

Miao JX, Wu Y, Yang Y, 2021. Identifying visible parts via pose estimation for occluded person re-identification. *IEEE Trans Neur Netw Learn Syst*, early access. https://doi.org/10.1109/TNNLS.2021.3059515

Pan YH, 2019. On visual knowledge. *Front Inform Technol Electron Eng*, 20(8):1021-1025. https://doi.org/10.1631/FITEE.1910001

Pan YH, 2020a. Miniaturized five fundamental issues about visual knowledge. *Front Inform Technol Electron Eng*, early access. https://doi.org/10.1631/FITEE.2040000

Pan YH, 2020b. Multiple knowledge representation of artificial intelligence. *Engineering*, 6(3):216-217. https://doi.org/10.1016/j.eng.2019.12.011

Pan YH, 2021. On visual understanding. *Front Inform Technol Electron Eng*, early access. https://doi.org/10.1631/FITEE.2130000

Serafini L, d'Avila Garcez A, 2016. Logic tensor networks: deep learning and logical reasoning from data and knowledge. Proc 11th Int Workshop on Neural Symbolic Learning and Reasoning Co-located with the Joint Multiconf on Human-Level Artificial Intelligence.

Simonyan K, Zisserman A, 2014. Two-stream convolutional networks for action recognition in videos. Proc 27th Int Conf on Neural Information Processing Systems, p.568-576.

Singh J, Zheng L, 2020. Combining semantic guidance and deep reinforcement learning for generating human level paintings. https://arxiv.org/abs/2011.12589

Sun YF, Cheng CM, Zhang YH, et al., 2020. Circle loss: a unified perspective of pair similarity optimization. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.6397-6406. https://doi.org/10.1109/CVPR42600.2020.00643

Tang KH, Niu YL, Huang JQ, et al., 2020. Unbiased scene graph generation from biased training. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.3716-3725. https://doi.org/10.1109/CVPR42600.2020.00377

Vaswani A, Shazeer N, Parmar N, et al., 2017. Attention is all you need. Proc 31st Int Conf on Neural Information Processing Systems, p.6000-6010.

Veeravasarapu V, Rothkopf C, Visvanathan R, 2017. Adversarially tuned scene generation. IEEE Conf on Computer Vision and Pattern Recognition, p.6441-6449. https://doi.org/10.1109/CVPR.2017.682

Vrandečić D, Krötzsch M, 2014. Wikidata: a free collaborative knowledge base. *Commun ACM*, 57(10):78-85. https://doi.org/10.1145/2629489

Wang XH, Zhu LC, Wu Y, et al., 2020. Symbiotic attention for egocentric action recognition with object-centric alignment. *IEEE Trans Patt Anal Mach Intell*, early access. https://doi.org/10.1109/TPAMI.2020.3015894

Xu DF, Zhu YK, Choy CB, et al., 2017. Scene graph generation by iterative message passing. IEEE Conf on Computer Vision and Pattern Recognition, p.3097-3106. https://doi.org/10.1109/CVPR.2017.330

Yan Y, Nie FP, Li W, et al., 2016. Image classification by cross-media active learning with privileged information. *IEEE Trans Multim*, 18(12):2494-2502. https://doi.org/10.1109/TMM.2016.2602938

Zhu LC, Fan HH, Luo YW, et al., 2021. Temporal cross-layer correlation mining for action recognition. *IEEE Trans Multim*, early access. https://doi.org/10.1109/TMM.2021.3057503