

MMKG构建

MMKG构建的两种思路：

1 基于已有的KGs对图像进行[主体，属性，客体]三元组提取






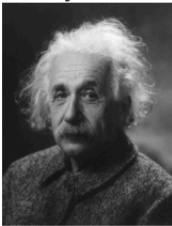
relation type	example	images	relation type	example	images
Concept-Concept	Keyboard is a part of Laptop.		Scene-Entity	Ferris wheel is found in Amusement park.	
					
Entity-Concept	BMW 320 is a kind of Car.		Scene-Attribute	Alleys are Narrow.	

TABLE 3: Examples of visual relations detected in NEIL [22]

2 用多模态数据描述已有KGs中的知识符号
对KGs中的实体、概念和关系进行多模态描述



concept type	visualizable concept	non-visualizable concept
example	Surgeon	Physicist
image		

MMKG实例：

Pub. Time	MMKGs	Types	Scale (#nodes / #images)	Data Sources	Sup. Tasks
2013-12	NEIL [42]	N-MMKG	1152 (classes) / 300K	WN / Image WSE	OD, etc.
2014-09	ImageNet [48]	A-MMKG	21K (classes) / 3.2M	WN / Image WSE	IMGC, OD, etc.
2016-02	VisualGenome [49]	A-MMKG	35 (classes) / 108K	WN / MS COCO / YFCC [50]	SGG, VQA, etc.
2016-09	WN9-IMG [51]	A-MMKG	6.5K (entities) / 14K	WN / ImageNet	MKG
2017-01	ImageGraph [52]	A-MMKG	15K (entities) / 837K	FB / Image WSE	CMR
2017-10	IMGpedia [53]	N-MMKG	2.6M (entities) / 15M	DBP / WM Commons	CMR
2019-03	MMKG [54]	A-MMKG	45K (entities) / 37K	FB / DBP / YG / Image WSE	MMEA, MKGC
2020-07	GAIA [41]	N-MMKG	457K (entities) / NA	FB / GeoNames / News Websites	MMIE
2020-08	VisualSem [44]	N-MMKG	90K (nodes) / 938K	WP / WN / ImageNet	CMR
2020-09	DBP-DWY-Vis [55]	A-MMKG	178K (entities) / 117K	WP / DBP15k [56] / DWY15K [57]	MMEA
2020-12	Richpedia [58]	N-MMKG	2.8M (entities) / 2.9M	WD / WM / Image WSE	MMKG Querying
2021-06	RESIN [59]	N-MMKG	51K (events) / NA	WD / News Websites	MMIE
2022-10	MKG-W&Y [60]	A-MMKG	30K (entities) / 29K	OpenEA [61] / Image WSE	MKG
2022-10	MarKG [62]	A-MMKG	11K (entities) / 76K	WD / Image WSE	MKG
2023-02	Multi-OpenEA [63]	A-MMKG	920K (entities) / 2.7M	OpenEA / Image WSE	MMEA
2023-03	UKnow [64]	N-MMKG	1.4M (entities) / 1.1M	WP / Image WSE	MKG, CMR
2023-07	UMVM [65]	A-MMKG	238K (entities) / 205K	DBP-DWY-Vis / Multi-OpenEA	MMEA
2023-08	AspectMMKG [45]	A-MMKG	2.3K (entities) / 645K	WP / Image WSE	MMEL
2023-10	TIVA-KG [66]	A-MMKG*	440K (entities) / 1.7M	CN / Image WSE	MKG
2023-11	MMpedia [67]	A-MMKG	2.7M (entities) / 19.5M	DBP / Image WSE	MKG
2023-12	VTKGs [68]	A-MMKG*	43K (entities) / 460K	CN / WN / UnRel [69] / VRD [70] HICO-DET [71] / VisKE [72]	MKG
2023-12	M ² ConceptBase [46]	A-MMKG	152K (concepts) / 951K	Wukong [73] / Baidu Encyclopedia	VQA, CU

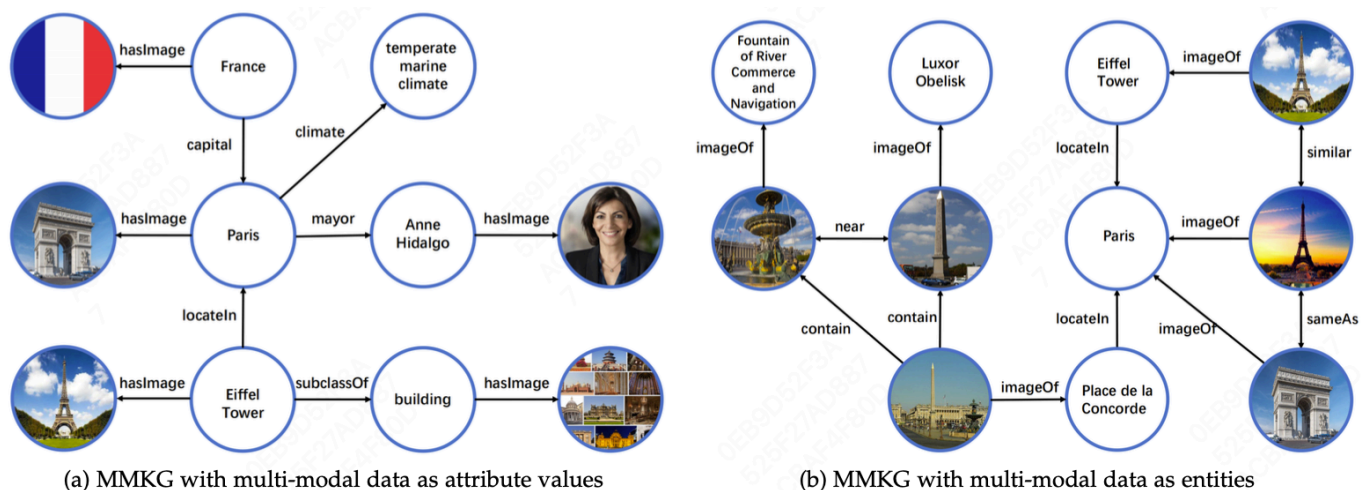
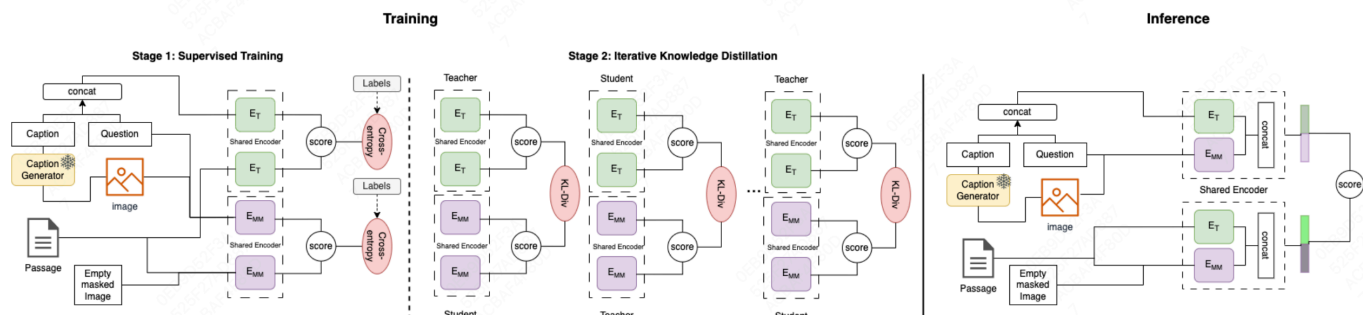


Fig. 1: Example MMKGs of two different types: A-MMKG and N-MMKG

KG中的知识检索

2023: A symmetric dual encoding dense retrieval framework for knowledge-intensive visual question answering



ET是文本编码器、EMM是多模态编码器（把图文对当作输入）

训练：1 文本编码器：拉进 图像描述+问题 和 对应检索内容 的距离 2 多模态编码器：拉进 图像+问题 和 空图 +对应检索内容 的距离

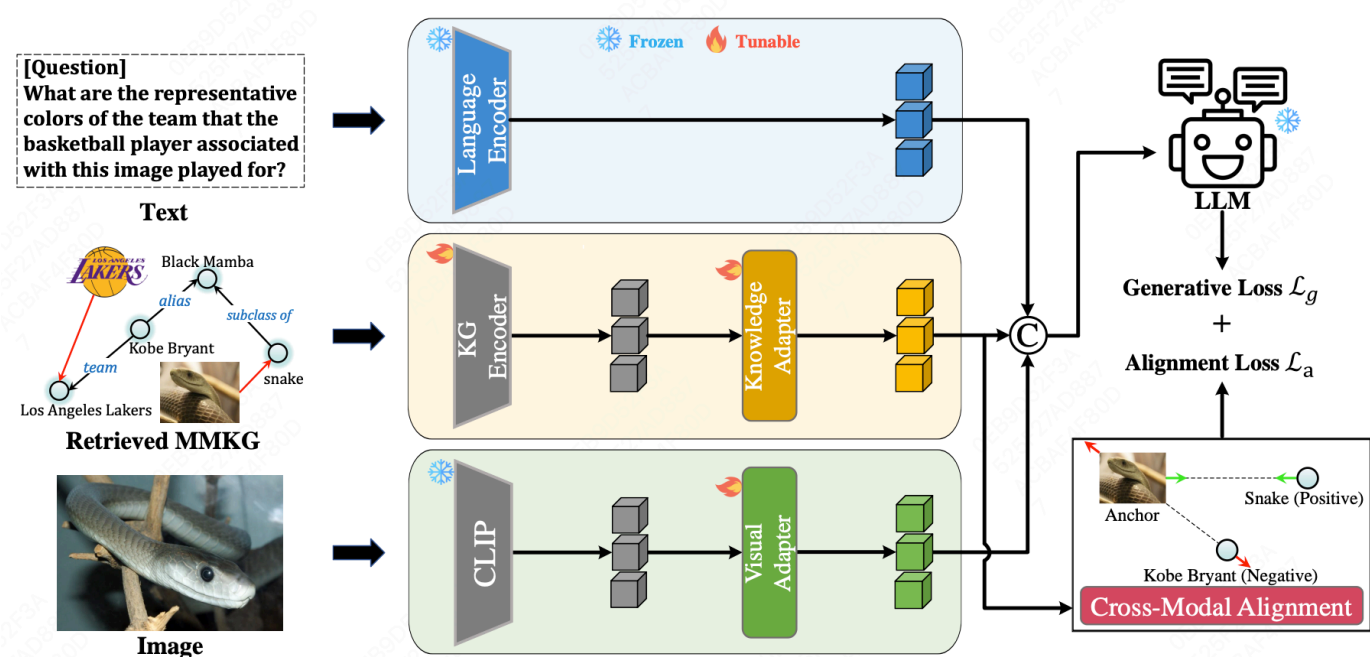


Figure 2: The overview of our MR-MKG approach. Text, multimodal knowledge graph, and image are independently embedded and then concatenated to form prompt embedding tokens. A cross-modal alignment module is designed to enhance the image-text alignment through a matching task within MMKGs.

文本编码器 + KG编码器 + 图像编码器

在检索出来的多模态知识图谱的子图上进行跨模态的对齐

Multi-modality RAG技术路线

Option 1:

- Use multimodal embeddings (such as CLIP) to embed images and text
- Retrieve both using similarity search
- Pass raw images and text chunks to a multimodal LLM for answer synthesis

Option 2:

- Use a multimodal LLM (such as GPT-4V, LLaVA, or FUYU-8b) to produce text summaries from images
- Embed and retrieve text
- Pass text chunks to an LLM for answer synthesis

Option 3

- Use a multimodal LLM (such as GPT-4V, LLaVA, or FUYU-8b) to produce text summaries from images
- Embed and retrieve image summaries with a reference to the raw image
- Pass raw images and text chunks to a multimodal LLM for answer synthesis

1 以图像/文本的嵌入检索图像/文本的嵌入，返回图像/文本

2 以图像描述的文本嵌入检索知识库中的文本嵌入，返回文本

3 以图像描述的文本嵌入检索知识库中的文本嵌入or图像描述嵌入，返回图像/文本